

# Implementasi Data Mining Dalam Menentukan Penjualan Paket Internet Dan Tv Kabel Dengan Metode Algoritma C4.5 Pada PT Link Net

*Implementation of Data Mining in Determining Sales of Internet and Cable Tv Packages*

*Using the C4.5 Algorithm Method at PT Link Net*

Fajar Arief Rachman<sup>1</sup>, Wahyu Hadikristanto<sup>2</sup>, Arvita Emarilis Intani<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

<sup>1</sup>fatzerst@gmail.com\*, <sup>2</sup>wahyu.hadikristanto@pelitabangsa.ac.id, <sup>3</sup>arvita@pelitabangsa.ac.id

## Abstract

*APT Link Net is a multimedia company that became the first provider in Indonesia that provides internet speeds of up to 100 Mbps and provides more than 90 local and international channels with digital picture quality and stereo sound, PT. Link Net offers several choices of internet and cable TV packages such as Dlite, Elite, Family, Family Plus, Maxima, School, Supreme, judging by the large number of sales for these packages and there is no predictor of determining the sales of the best-selling and not-selling packages. The research uses data mining techniques using the C4.5 algorithm to analyze or determine package sales so that they can prepare more supporting tools to be used for best-selling packages. This study uses sales data sourced directly from PT. Link Net in 2021. The stages in the C4.5 Algorithm use six KDD (Knowledge Discovery in Databases) steps which include activities, namely, selection, preprocessing, transformation, data mining, Interpretation Evaluation. Through testing using rapidminer with an accuracy rate of 99.5%, thus the C4.5 Algorithm is very well used to implement this sales data. From the explanation above, it can be concluded that the best-selling Internet and cable TV packages at PT. Link Net namely Family and Family Plus.*

**Keywords:** Data Mining, C4.5 Algorithm, Decision Tree, Prediction

## Abstrak

APT Link Net adalah perusahaan multimedia yang menjadi provider pertama di Indonesia yang menghadirkan kecepatan internet sampai dengan 100 Mbps dan menyajikan lebih dari 90 saluran lokal dan internasional dengan kualitas gambar digital dan suara stereo, PT. Link Net menawarkan beberapa pilihan paket internet dan TV kabel seperti *Dlite, Elite, Family, Family Plus, Maxima, School, Supreme*, dilihat dari banyaknya jumlah penjualan akan paket tersebut dan belum adanya prediksi penentu penjualan paket paling laris dan tidak laris maka dari itu dilakukan penelitian menggunakan teknik data mining menggunakan algoritma C4.5 untuk menganalisa atau menentukan penjualan paket sehingga dapat mempersiapkan alat pendukung yang lebih banyak untuk digunakan untuk paket terlaris. Penelitian ini menggunakan data penjualan yang bersumber langsung dari PT. Link Net pada tahun 2021. Tahapan dalam Algoritma C4.5 ini menggunakan enam Langkah KDD (*Knowledge Discovery in Databases*) yang mencakup aktivitas yaitu, *selection, preprocessing, transformation, data mining, Interpretation Evaluation*. Melalui pengujian menggunakan rapidminer dengan tingkat akurasi 99,5%, dengan demikian Algoritma C4.5 sangat baik digunakan mengimplementasikan data penjualan ini. Dari penjelasan diatas dapat disimpulkan paket Internet dan TV kabel terlaris pada PT. Link Net yaitu Family dan Family Plus.

**Kata kunci:** Data Mining, Algoritma C4.5, Decision Tree, Prediksi

## Pendahuluan

Jumlah pengguna Internet di dunia selalu mengalami pertumbuhan yang sangat pesat, pada tahun 2000 jumlah pengguna internet berjumlah sekitar 400 juta di dunia pengguna dan pada tahun 2015 berjumlah sekitar 3,2 Miliar pengguna [1]. Peningkatan tertinggi terjadi pada negara berkembang yaitu 100 juta pengguna pada tahun 2000 naik menjadi 3 Miliar pengguna pada tahun 2015 [2]. Kecenderungan masyarakat dalam menggunakan internet ikut serta berubah dan memberi pengaruh pada peningkatan jumlah pengguna internet di dunia dari tahun 2000 hingga tahun 2015, penetrasi pengguna *mobile-cellular* sebesar 97%, penetrasi para pengguna *mobile broadband* sebesar 47%, penetrasi pengguna *fixed broadband* sebesar 11%, dan penetrasi pengguna yang memiliki akses internet dari rumah meningkat dari 18% pada tahun 2005 dan naik menjadi 46% pada tahun 2015.[3]

Perkembangan teknologi juga akan mempengaruhi peningkatan jumlah pengguna jaringan internet [4], mulai dari perkembangan teknologi *mobile*, aplikasi, *device*, *tv cable* dan *smart*. First Media menawarkan beberapa paket Internet dan TV Kabel Seperti Family Plus, Elite, Dlite dll, setiap harinya produk itu ditawarkan dan dijual kepada masyarakat melalui peran para sales yang terbagi dalam 4 wilayah kantor cabang yaitu Fatmawati, Cideng, Karawaci, dan Bekasi, dalam setiap wilayah tentu memiliki penjualan yang berbeda - beda setiap harinya maupun bulan yang jumlahnya banyak apabila diakumulasikan dan datanya sulit untuk dianalisa. Analisa diperlukan untuk melihat pola dari data penjualan sehingga dapat menghasilkan prediksi penjualan laris dan tidak laris yang nantinya akan berguna untuk gambaran jenis apa saja yang akan memiliki peluang baik. Penggalian informasi dari kumpulan data yang berskala besar dan banyak dapat dilakukan dengan menggunakan teknik yaitu data mining. salah satu teknik dari Data Mining yang dapat digunakan yaitu adalah dengan menggunakan algoritma C4.5. Prediksi Data Mining dengan menggunakan algoritma C4.5 dapat menghasilkan knowledge yang dijadikan sebagai dasar pengambilan keputusan dalam memprediksi.

Algoritma C4.5 adalah bagian dari algoritma untuk klasifikasi dalam pembelajaran machine learning dan data mining [5]. C4.5 merupakan algoritma yang cocok digunakan untuk masalah klasifikasi pada machine learning dan data mining [6]. Dalam pembuatan pohon keputusan, setiap algoritma menerapkan ukuran pemilihan atribut yang berbeda-beda [7]. Ukuran pemilihan atribut merupakan ukuran yang digunakan dalam menentukan kriteria yang terbaik untuk mengelompokkan tuple. Ukuran pemilihan atribut ini juga disebut sebagai *splitting rules* karena menentukan bagaimana data akan dipisahkan kesetiap cabang. diciptakan oleh Claude Shannon dengan mempelajari nilai informasi dari data, dan menggunakan nilai tersebut sebagai acuan dalam menentukan atribut yang akan digunakan dalam menyusun pohon keputusan [8].

Algoritma C4.5 adalah salah satu metode untuk membuat *decision tree* berdasarkan training data yang telah disediakan. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah pilih atribut sebagai akar, buat cabang untuk masing - masing nilai, bagi kasus dalam cabang, ulangi proses untuk masing – masing cabang sampai semua kasus selesai [9].

## Metode Penelitian

Metode yang digunakan oleh penulis dalam mengklasifikasikan data penjualan paket internet dan TV kabel pada PT Link Net menggunakan algoritma C4.5 melalui proses lima langkah dalam KDD (*Knowledge Discovery in Database*) yang mencakup beberapa aktivitas *Selection*, *Pre Processing*, *Transformation*, *Data Mining*, dan *Evaluation* [10]. Pada tahap *selection* terdapat 500 data *record* akan diambil 5 atribut, yaitu: Nama Paket, Harga, Kecepatan, channel, dan Terjual. Data hasil seleksi ini akan digunakan dalam proses *data mining*.

Tabel 1. Atribut Yang Digunakan

No	Atribut	Tipe
1	Nama Paket	Text
2	Harga	Currency
3	Kecepatan	Text
4	Channel	Text
5	Terjual	Numeric

Pada proses pre-processing akan dilakukan pembersihan data untuk membuang data yang missing value yaitu data yang tidak konsisten dan juga memperbaiki data yang rusak

Tabel 2.Data Set.

No	Nama Paket	Harga	Kecepatan	Channel	Terjual
1	Dlite	Rp 451.000,00	25 Mbps	140	9
2	Elite	Rp 641.000,00	40 Mbps	157	2
3	Family	Rp 272.000,00	8 Mbps	124	239
4	Family Plus	Rp 361.000,00	15 Mbps	121	35
5	Maxima	Rp 1.799.000,00	150 Mbps	186	2
6	School	Rp 345.000,00	50 Mbps	26	0
7	Supreme	Rp 719.250,00	75 Mbps	170	1
8	Dlite	Rp 451.000,00	25 Mbps	140	7
9	Elite	Rp 641.000,00	40 Mbps	157	2
10	Family	Rp 272.000,00	8 Mbps	124	106
...	...	...	...	...	...
500	Family	Rp 272.000,00	8 Mbps	124	361

Setelah data sudah dipilih maka dilakukan tahapan untuk melakukan transformasi terhadap atribut, transformasi akan dilakukan untuk memodifikasi sumber data ke format yang berbeda yang dapat diterima oleh proses data mining pada tahap selanjutnya [11].

Table 3. Klasifikasi Berdasarkan Nama Paket

No	Nama Paket	Status	
		Laris ( $\geq 30$ )	Tidak Laris ( $< 30$ )
1	Dlite	1	57
2	Elite	0	57
3	Family	57	0
4	Family Plus	56	1
5	Maxima	0	57
6	School	0	57
7	Supreme	0	57

Tabel 4. Klasifikasi Berdasarkan Harga

No	Harga	Status	
		HR ( $< 500000$ )	HT ( $\geq 500000$ )
1	Rp 451.000,00		HR
2	Rp 641.000,00		HT
3	Rp 272.000,00		HR
4	Rp 361.000,00		HR
5	Rp 1.799.000,00		HT
6	Rp 345.000,00		HR
7	Rp 719.250,00		HT

Tabel 5. Klasifikasi Berdasarkan Kecepatan

No	Kecepatan	Status		
		KR ( $< 15$ )	KH ( $> 15$ )	KT ( $> 50$ )
1	25			KS
2	40			KS
3	8			KR
4	15			KR
5	150			KT
6	50			KS
7	75			KT

Tabel 6 Klasifikasi Berdasarkan Channel

No	Channel	Status		
		Sedikit ( $< 30$ )	Cukup Banyak ( $\geq 120$ )	Banyak ( $> 150$ )
1	140		Cukup Banyak	
2	157		Banyak	
3	124		Cukup Banyak	
4	121		Cukup Banyak	
5	186		Banyak	
6	26		Sedikit	
7	170		Banyak	

Tabel 7. Klasifikasi Berdasarkan Terjual

No	Terjual	Status	
		Laris ( $\geq 30$ )	Tidak
	400	114	

Pada penelitian ini akan dilakukan pengujian menggunakan metode Algoritma C4.5 untuk mengolah data penjualan pada PT Link Net. Metode Algoritma C4.5 dipilih karena satu keunggulannya adalah dapat mengatasi data numerik dan diskrit. Dengan menghitung konsep entropy dan gain untuk membentuk pohon keputusan. Data kemudian dihitung menggunakan metode algoritma sesuai dengan metodenya kemudian dicari hasil akurasi [12],[13],[14]. Beberapa tahap dalam membentuk decision tree (pohon keputusan) dengan Algoritma C4.5 antara lain:

Evaluasi dapat dilakukan dengan cara yaitu mengamati dan menganalisa data hasil dari algoritma C4.5 yang digunakan dalam pengujian data penjualan paket internet dan TV kabel dan memastikan bahwa hasil pengujian itu sesuai atau tidak sesuai dengan pembahasan yang telah dilakukan. Lalu untuk validasi dilakukan dengan mengukur hasil dari prediksi dari data yang digunakan untuk mengetahui tingkat akurasi, presisi dan recall.

Tabel 8. Akurasi, Presisi, Recall

No	Terjual	Status	
		Laris ( $\geq 30$ )	Tidak
	400	114	

## Hasil dan Pembahasan

Pembahasan dalam penelitian ini diantaranya membahas tentang Perhitungan pada performa, Perhitungan nilai Entropy, Information Gain,, Decision Tree dan Pengujian hasil Rapidminer.

Perhitungan nilai Entropy total merupakan Langkah awal dalam Algoritma C4.5 adalah mencari nilai entropy [15]. Pertama, tentukan terlebih dahulu nilai entropy total dalam kasus, rumusnya sebagai berikut: Diketahui:

$$\begin{aligned}
 \text{Jumlah kasus (total)} &= 400 \\
 \text{Jumlah hasil Laris} &= 114 \text{ Jumlah hasil Tidak Laris} = 286 \text{ Maka nilai entropy total adalah :} \\
 &= ((\text{Jumlah Laris}/\text{total}) * \text{IMLOG2}(\text{Jumlah Laris}/\text{total}) + (\text{Jumlah Tidak Laris}/\text{total}) * \text{IMLOG2}(\text{Jumlah Tidak Laris}/\text{total})) \\
 &= ((114/400) * \text{IMLOG2}(114/400) + (286/400) * \text{IMLOG2}(286/400)) \\
 &= 0,86217453
 \end{aligned}$$

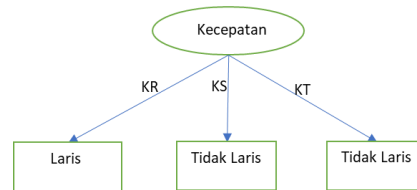
Perhitungan Nilai Gain

Entropy Total = 0,86217453 Tabel 4.1 Nilai Gain

Tabel 9. Nilai Gain

No	Attribute	Status		Jumlah Data	Entropy	Gain	
		Laris	Tidak Laris				
	Total	114	286	400	0,8621745		
1	Nama Paket	Dlite	1	57	58	0,1256581	
		Elite	0	57	57	0	
		Family	57	0	57	0	
		Family Plus	56	1	57	0,1274185	0,6
		Maxima	0	57	57	0	
		School	0	57	57	0	
		Supreme	0	57	57	0	
2	Harga	HR	114	115	229	0,9999862	0,2
		HT	0	171	171	0	
		KR	113	1	114	0,0725372	
3	Kecepatan	KS	1	171	172	0,0515393	0,8
		KT	0	114	114	0	
		Sedikit	0	57	57	0	
4	Channel	Cukup Banyak	114	58	172	0,9221231	0,4
		Banyak	0	171	171	0	

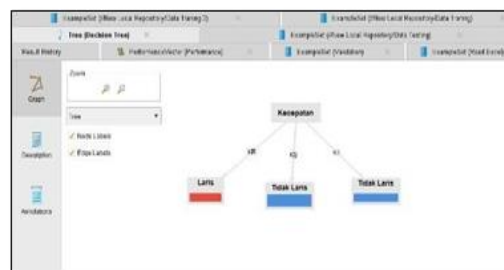
Decision Tree adalah alat pendukung keputusan yang menggunakan model keputusan seperti pohon dan konsekuensinya yang mungkin, termasuk hasil acara kebetulan, biaya sumber daya, dan utilitas [14]. Ini adalah salah satu cara untuk menampilkan algoritma yang hanya berisi pernyataan kontrol bersyarat. Dari tabel 4.4 diatas tentang hasil perhitungan entrophy dan gain maka didapat nilai gain tertinggi dijadikan node akar pertama. Pada tabel diatas node akar tertinggi adalah atribut “Kecepatan” dengan nilai gain 0,8193395, karena hasil laris dan tidak laris dari atribut “Kecepatan” sudah jelas perbandingannya antar KR dan KS maka pencarian node dinyatakan sudah selesai. Maka bisa dilihat dari pohon keputusan dengan root node seperti pada gambar dibawah ini :



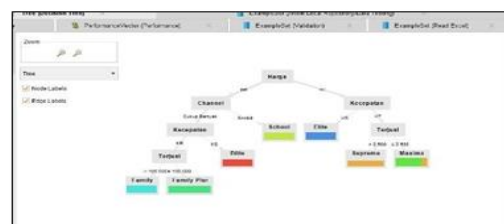
Gambar 1 Pohon Keputusan *RootNode*

Pada gambar diatas dapat disimpulkan bahwa root node adalah status yang memiliki 3 akar, yang terdiri dari KR ( Kecepatan Rendah ) laris, KS ( Kecepatan Sedang ) tidak laris dan KT ( Kecepatan Tinggi ) tidak laris.

Pengujian Hasil *RapidMiner*. Pada tahap ini metode *data mining* dan diterapkan untuk menemukan pengetahuan yang tersembunyi dan berharga dari data [16]. Metode yang digunakan adalah klasifikasi menggunakan algoritma C4.5. Berikut ini adalah langkah-langkah pada *tools RapidMiner* versi 9.10 untuk mengetahui bentuk dari perhitungan *decision Tree* pada data *training*.



Gambar 2 Tampilan Hasil Decision Tree Data Training



Gambar 3 Tampilan Hasil Decision Tree Data Testing

	True Tidak Laris	True Laris	Class precision
pred. Tidak Laris	205	1	99.65%
pred. Laris	1	132	99.12%
class recall	99.65%	99.12%	

Gambar 4 Tampilan Hasil Accuracy Data Training

## Kesimpulan

Setelah melakukan analisis, perancangan, implementasi beserta pengujian penerapan Algoritma C4.5 untuk menentukan Prediksi Paket Internet dan TV Kabel pada PT. Link Net dengan data yang digunakan 500 data yang selanjutnya data dibagi menjadi Data Training dan Data Testing menggunakan Split Validation dengan

perbandingan 80% digunakan untuk Data Training, 20% untuk Data Testing. Proses ini menghasilkan kombinasi item dengan menerapkan nilai Entrophy dan nilai gain, setelah didapat nilai gain tertinggi, maka dibentuk Decision Tree (Pohon Keputusan) dan dapat menunjukkan atribut mana yang dijadikan akar penentu produk di katakan Laris dan Tidak. Pengujian ini dilakukan dengan aplikasi RapidMiner .10, evaluasi hasilnya dengan Confusion Matrix dimana Data Training dengan tingkat accuracy sebesar 99,5% dan Data Testing dengan tingkat accuracy sebesar 97.00%. maka didapat kesimpulan bahwa dengan menggunakan Algoritma C4.5 sangat efektif dalam memprediksi paket paling Laris yaitu paket Family dan Family Plus dalam atribut HR (Harga Rendah) dan paket yang Tidak Laris yaitu Dlite, Elite, School, Maxima dan Supreme.

## Daftar Rujukan

- [1] Iqbal, Muhammad, and Muhammad Ashry Sallatu. (2022). "Dampak Ekspansi Alibaba Group Terhadap Perkembangan E-Commerce Di Indonesia." *Hasanuddin Journal of International Affairs*: 68-91.
- [2] Hasibuan, Lidya Rahmadani. (2019). "Implementasi Peraturan Daerah Kota Medan No. 3 Tahun 2014 Tentang Kawasan Tanpa Asap Rokok Pada Kota Medan." *Jurnal Hukum Responsif*: 96-101.
- [3] W. Suharso, (2016), "Penyelarasan Tujuan TI Dan Tujuan Bisnis pada Perusahaan Layanan Internet," *Semin. Nas. Teknol. dan Rekayasa*, vol. No.2, pp. 183–190.
- [4] Antoni, Darius, et al. (2021). "Pengembangan Infrastruktur Jaringan Untuk Meningkatkan Pelayanan Publik di Kota Palembang." *Jurnal Media Informatika Budidarma*: 1652-1659.
- [5] Fitriani, Eka. (2020). "Perbandingan Algoritma C4. 5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan." *SISTEMASI: Jurnal Sistem Informasi*: 103-115.
- [6] Fitriani, Eka. (2020). "Perbandingan Algoritma C4. 5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan." *SISTEMASI: Jurnal Sistem Informasi*: 103-115.
- [7] Asmoro, Achmad Shoddiq Bayu, Wahyu Sakti Gunawan Irianto, and Utomo Pujiyanto. (2019). "Perbandingan Kinerja Hasil Seleksi Fitur pada Prediksi Kinerja Akademik Siswa Berbasis Pohon Keputusan." *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*: 84-89.
- [8] Yahya, Yahya, and Reni Zuliana. (2018). "Prediksi Jumlah Penggunaan BBM Perbulan Menggunakan Algoritma Decision Tree (C4. 5) Pada Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur." *Infotek: Jurnal Informatika dan Teknologi*: 56-63.
- [9] Junaedi, Ifan, Ndaru Nuswantari, and Verdi Yasin. (2019). "Perancangan Dan Implementasi Algoritma C4. 5 Untuk Data Mining Analisis Tingkat Risiko Kematian Neonatum Pada Bayi." *Journal of Information System, Informatics and Computing*: 29-44.
- [10] Sugiyarto, Ipin, Rama Irawan, and Didi Rosiyadi. (2021). "Pengelompokan Dampak Gempa Bumi Dan Kerusakan Pada Wilayah Berpotensi Gempa Di Provinsi Sumatra Barat: Indonesia." *Journal of Students' Research in Computer Science*: 211-222.
- [11] Nurfajiah, Alfi, Tuti Hartati Mindara, and Dita Rizki Amalia. (2021). "Integrated Library System Untuk Meningkatkan Efektivitas Layanan Perpustakaan Dengan Menggunakan Metode Algoritma Apriori (Studi Kasus: Perpustakaan Kabupaten Cirebon): Metode Algoritma Apriori." *Journal of Information Technology*: 39-44.
- [12] Tulus, Tulus Harry Lamramot, Asep Id Hadiana, and Irma Santikarama. (2022). "Sistem Prediksi Awal Terhadap Atrisi Karyawan Menggunakan Algoritma C4. 5." *Informatics and Digital Expert (INDEX)*: 18-24.
- [13] Romli, I., & Zy, A. T. (2020). Penentuan jadwal overtime dengan klasifikasi data karyawan menggunakan algoritma C4. 5. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 4(2), 694-702.
- [14] Alfaridzi, S., Nugroho, A., & Sani, M. R. (2021). Penerapan data mining menggunakan algoritma C4. 5 dalam prediksi penyakit angin duduk. *JSI: Jurnal Sistem Informasi (E-Journal)*, 13(2).
- [15] Putri, Sanni Ucha, Eka Irawan, and Fitri Rizky. (2021). "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4. 5." *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*: 39-46.
- [16] Zein, Muhammad Helmi. (2021). "Implementasi Program Penanganan Prasarana Dan Sarana Umum (PPSU) Di Kelurahan Melawai Kecamatan Kebayoran Baru Kota Administrasi Jakarta Selatan." *Jurnal Inovasi Penelitian*: 1863-1874.
- [17] Dewi, PUTRI SUKMA, C. K. Sastradipraja, and D. Gustian. (2021). "Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Algoritma Naïve Bayes Classifier." *Jurnal Teknologi Dan Informasi*: 66-80.