Bit-Tech (Binary Digital - Technology)
Vol.7, No.2, Desember 2024
Available online at: http://jurnal.kdi.or.id/index.php/bt

Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Prediksi Loyalitas Pelanggan

Ardiane Rossi Kurniawan Maranto[1)*], Lily Damayanti[2)], Irvan Rahul Ramadika[3)]

[1)2)3)]*Universitas Buddhi Dharma*
Jl. Imam Bonjol No.41, Tangerang, Indonesia
[1)]ardiane.rossi@ubd.ac.id
[2)]lily.damayanti@ubd.ac.id
[3)]kalizaica@gmail.com

Article history:

Received 22 Nov 2024;
Revised 25 Nov 2024;
Accepted 29 Nov 2024;
Available online 27 Des 2024

Keywords:

Algoritma C4.5
Data Mining
Internet Service Provide
Loyalitas Pelanggan
Naïve Bayes

Abstrak

Industri *Internet Service Provider* (ISP) saat ini menghadapi tantangan besar dalam menjaga loyalitas pelanggan di tengah persaingan pasar yang semakin ketat. Loyalitas pelanggan menjadi aspek penting karena tidak hanya menjamin pendapatan yang stabil, tetapi juga meningkatkan reputasi perusahaan di mata publik. Penelitian ini berupaya mengevaluasi dua algoritma pembelajaran mesin, yaitu Algoritma C4.5 dan Naïve Bayes, dalam memprediksi loyalitas pelanggan ISP, memberikan wawasan tentang algoritma yang paling efektif untuk analisis loyalitas pelanggan. Dengan menggunakan pendekatan CRISP-DM, data pelanggan yang mengajukan penghentian layanan hingga akhir tahun 2020 menjadi data yang akan dianalisis. Data tersebut mencakup berbagai atribut, seperti alasan penghentian, jenis produk, metode penghentian, dan informasi demografis. Proses validasi penelitian ini menggunakan teknik 10-*fold cross-validation*, dengan hasil menunjukkan bahwa algoritma C4.5 memiliki performa lebih baik unggul dibandingkan Naïve Bayes. Algoritma C4.5 mencatat akurasi sebesar 80,67% dan nilai AUC 0,830, sementara Naïve Bayes mencatat akurasi 76,23% dan AUC 0,824. Dengan keunggulan ini, Algoritma C4.5 terbukti lebih akurat dalam membedakan pelanggan loyal dan tidak loyal. Hasil penelitian ini dapat memberikan rekomendasi strategis bagi ISP untuk meningkatkan pengelolaan pelanggan melalui analisis data yang optimal. Dengan algoritma yang tepat, perusahaan dapat mengembangkan strategi yang efektif untuk mempertahankan pelanggan setia, mengurangi tingkat kehilangan pelanggan, dan memperkuat daya saing mereka. Penelitian ini juga menekankan pentingnya adopsi teknologi prediktif untuk mendukung pengambilan keputusan strategis di industri ISP yang terus berubah.

I. PENDAHULUAN

Loyalitas pelanggan merupakan faktor krusial dalam *industry* saat ini. Memahami dan memprediksi loyalitas pelanggan dapat membantu perusahaan dalam merancang strategi retensi yang efektif untuk dapat memberikan pelayanan yang memuaskan [1]. Loyalitas pelanggan selain dapat memberikan pendapatan yang konsisten, hal ini menciptakan reputasi yang baik di pasar dan meminimalkan biaya pemasaran untuk mendapatkan pelanggan baru.

Dalam industri *Internet Service Provider* (ISP) yang sangat kompetitif saat ini, perusahaan perlu menerapkan strategi yang tepat dan memadai untuk menjaga loyalitas pelanggan. Dengan meningkatnya jumlah penyedia layanan internet, pelanggan memiliki banyak pilihan, sehingga perusahaan perlu lebih fokus pada peningkatan kualitas layanan dan pemahaman perilaku pelanggan agar tetap dapat mempertahankan mereka. Oleh karena itu, penting bagi perusahaan *internet service provider* untuk memanfaatkan data pelanggan guna meningkatkan keputusan strategis, terutama dalam hal retensi pelanggan.

Dalam konteks ini, penggunaan teknik pembelajaran mesin menjadi semakin relevan. Algoritma dapat digunakan untuk menganalisis data pelanggan dan memprediksi perilaku masa depan mereka, termasuk tingkat loyalitas. Teknik data mining seperti algoritma C4.5 dan Naïve Bayes sering digunakan dalam menganalisis perilaku pelanggan.

Hal ini didukung dengan beberapa penelitian sebelumnya telah melakukan uji coba metode Algoritma C4.5 dan Naïve Bayes seperti penelitian yang dilakukan [2] menunjukkan bahwa algoritma C4.5 lebih efektif

* Corresponding author

ISSN 2622-2728 (online) 2622-271X (print) © 2018 The Authors. Published by Komunitas Dosen Indonesia.
This is an open access article under the CC BY-SA 4.0 license (http://creativecommons.org/licenses/by/4.0/)
doi: 10.32877/bt.v7i2.1825

dibandingkan Naïve Bayes dalam menganalisis faktor-faktor yang dapat mempengaruhi ketepatan waktu pengiriman di PT. Rtrans Logistik Artamandiri, dengan akurasi 95% dibandingkan 91%. Penelitian yang dilakukan [3] menunjukkan bahwa algoritma C4.5 unggul dibandingkan dengan Naïve Bayes untuk topik mengklasifikasikan loyalitas pelanggan. Penelitian [4] dilakukan untuk mengatasi penurunan loyalitas pelanggan di PT MNC Play dengan menggunakan teknik data mining, didapatkan hasilnya algoritma C4.5 memberikan hasil akurasi 85,443%.

Dengan mengevaluasi dan membandingkan kinerja kedua algoritma ini, diharapkan penelitian ini dapat memberikan wawasan yang berharga kepada perusahaan ISP dalam memilih algoritma yang paling sesuai untuk menganalisis dan memprediksi perilaku loyalitas pelanggan ISP. Hal ini diharapkan akan membantu perusahaan untuk mengembangkan strategi yang lebih efektif dalam mempertahankan dan meningkatkan loyalitas pelanggan mereka, yang pada gilirannya akan berdampak positif pada pertumbuhan bisnis dan keberlanjutan jangka panjang perusahaan ISP.

II. TINJAUAN PUSTAKA

Loyalitas pelanggan merupakan aspek krusial dalam industri *Internet Service Provider* (ISP), terutama di tengah persaingan yang semakin ketat. Berbagai penelitian telah dilakukan untuk memahami dan memprediksi loyalitas pelanggan menggunakan metode data mining, khususnya algoritma C4.5 dan Naïve Bayes. Algoritma C4.5, yang dikenal sebagai metode pohon keputusan, telah banyak digunakan dalam klasifikasi data. Penelitian oleh Wati et al. membandingkan algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor (KNN) dalam mengklasifikasikan loyalitas pelanggan. Hasilnya menunjukkan bahwa baik Naïve Bayes maupun C4.5 mencapai akurasi sebesar 96,67%, dengan Naïve Bayes unggul dalam nilai AUC sebesar 0,997 [5]. Studi lain oleh Rahmayanti et al. membandingkan metode algoritma C4.5 dan Naïve Bayes untuk memprediksi kelulusan mahasiswa. Hasilnya menunjukkan bahwa algoritma C4.5 memiliki akurasi sebesar 90%, lebih baik dibandingkan Naïve Bayes yang memiliki akurasi 85% [6]. Selain itu, penelitian oleh Yunita & Ikasari membandingkan algoritma Naïve Bayes dan C4.5 dalam mengukur kepuasan pelanggan. Hasilnya menunjukkan bahwa algoritma C4.5 memiliki akurasi sebesar 94,17%, lebih tinggi dibandingkan Naïve Bayes dengan akurasi 85,83% [7].

A. Data Mining

Data mining merupakan proses menggali pola atau hubungan tersembunyi dalam kumpulan data yang sangat besar, yang terdiri dari ratusan hingga ribuan elemen terkait, dengan tujuan memperoleh informasi yang bernilai. Melalui serangkaian langkah analisis, data mining bertujuan mengungkap pengetahuan baru yang sebelumnya tidak terlihat, sehingga dapat memaksimalkan potensi nilai tambah dari data tersebut [8]. *Data mining* memiliki beberapa fungsi seperti berikut:

1. *Classification*: Proses pembuatan model untuk mengidentifikasi kelas suatu data yang kategorinya belum diketahui.
2. *Regression*: Teknik pemetaan data untuk menghasilkan nilai prediksi.
3. *Clustering*: Proses pengelompokan data berdasarkan karakteristik yang serupa.
4. *Association*: Teknik untuk menemukan aturan hubungan antara item-item dalam suatu kombinasi tertentu.
5. *Sequence*: Pendekatan untuk mengidentifikasi pola asosiasi antar item dalam kombinasi tertentu yang berlangsung selama beberapa periode waktu.
6. *Forecasting*: Proses memprediksi nilai tertentu berdasarkan pola yang teridentifikasi dalam sekelompok data.
7. *Solution*: Langkah menemukan akar masalah dan memberikan solusi untuk mendukung pengambilan keputusan dalam bisnis.

B. Algoritma C4.5

Algoritma C4.5 adalah metode yang digunakan untuk klasifikasi, segmentasi, atau pengelompokan dengan sifat prediktif. Algoritma ini memerlukan data input berupa sampel pelatihan dan labelnya [9]. Berikut adalah proses Algoritma C4.5 untuk menghitung *gain* [10][11]:

1. Mempersiapkan data training.
2. Menentukan akar pohon.
3. Rumus menghitung nilai entropy pada persamaan (1):

$$\text{Entropy (S)} = \sum_{j=1}^{n} - pi.log2\ pi \quad (1)$$

Penjelasan:
S : Kumpulan kasus
A : Fitur
n : Jumlah partisi dalam himpunan S

pi : Proporsi kasus dalam partisi Si terhadap S

4 Rumus perhitungan gain dapat dilihat pada persamaan (2):

$$\text{Gain(S,A)} = \text{Entropy (S)} - \sum_{j=1}^{n} \frac{|Si|}{|S|} * \text{Entropy}(S) \quad (2)$$

Keterangan:
S : Himpanun kasus
A : Atribut
n : Jumlah partisi berdasarkan atribut A
|Si| : Jumlah kasus pada partisi ke-i
|S| : Jumlah total kasus dalam S

C. Algoritma Naïve Bayes

Naïve Bayes *Classifier* adalah metode klasifikasi yang berasal dari Teorema Bayes. Karakteristik khas dari Naïve Bayes *Classifier* adalah asumsi yang sangat sederhana (naïf) mengenai independensi penuh antara setiap kondisi atau peristiwa [12]. Metode Naïve Bayes terkenal karena tingkat akurasinya yang tinggi saat diterapkan pada basis data yang besar [13]. Metode Naïve Bayes dijelaskan melalui persamaan 3 [14]:

$$\text{P}(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (3)$$

Keterangan:
X : Merupakan data dengan kelasnya belum diketahui
H : Menyatakan hipotesis bahwa data X termasuk kedalam suatu kelas tertentu
P(H|X) : Probabilitas hipotesis H yang diberikan berdasakan kondisi data X
P(H) : Probabilitas awal dari hipotesis H
P(X|H) : Probabilitas data X muncul dengan asumsi hipotesis H benar
P(X) : Probabilitas keseluruhan data X

D. *Confusion Matrix*

Confusion matrix adalah suatu *matriks* yang disusun dalam bentuk baris dan kolom, dimana baris mewakili kelas aktual dari *instance data*, sedangkan kolom mewakili kelas yang diprediksi [15]. *Confusion matrix* digunakan untuk menilai kinerja dan perilaku dari model klasifikasi. *Confusion matrix* mencakup empat pengukuran metrix utama yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) sebagaimana ditampulkan pada tabel 1 berikut:

TABEL 1
TABEL *CONFUSION MATRIX*

		Predicted	
		Positive (P)	***Negative (N)***
		+	-
Actual	***Positive (P)*** +	*True Positive (TP)*	*False Negative (FN)*
	Negative (N) -	*False Positive (FP)*	*True Negative (TN)*

1. *True Positive* (TP): Kondisi ketika model memprediksi hasil positif dan hasil sebenarnya juga positif.
2. *True Negative* (TN): Kondisi di mana model memprediksi hasil negatif dan hasil sebenarnya juga negatif.
3. *False Positive* (FP): Kondisi di mana model memprediksi hasil positif, namun hasil sebenamya negatif.
4. *False Negative* (FN): Kondisi di mana model memprediksi hasil negatif, namun hasil sebenamya positif.

Hasil evaluasi menggunakan metode *confusion matrix* memberikan nilai *Accuracy*, *Precision*, *Recall*, serta *Classification Error* dengan penjelasan nya sebagai berikut:

1. *Accuracy*
   Accuracy menggambarkan persentase keakuratan prediksi model, yaitu perbandingan antara jumlah prediksi yang akurat (baik memprediksi positif maupun memprediksi negatif) dengan keseluruhan data setelah proses klasifikasi diuji [16]. Rumus perhitungan *accuracy* dapat dilihat pada persamaan 4.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (4)$$

2. *Precision*
   Precision diukur untuk menilai seberapa akurat model dapat memprediksi suatu kelas. Ini dihitung dengan membandingkan jumlah data yang tepat benar diprediksi sebagai positif dibagi dengan hasil total prediksi positif [17]. Rumus perhitungan *precision* dapat dilihat pada persamaan 5.

$$Precision = \frac{TP}{(TP + FP)} \quad (5)$$

3. *Recall*
   Recall dihitung dengan membagi jumlah data positif yang berhasil diprediksi dengan benar oleh model dengan total keseluruhan data positif aktual dalam dataset. [17]. Rumus perhitungan *recall* dapat diliat pada persamaan 6.

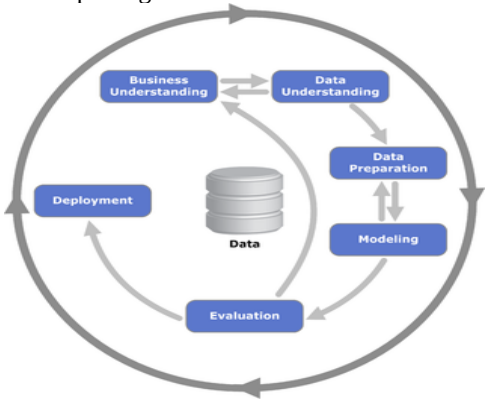
$$Recall = \frac{TP}{(TP + FN)} \quad (6)$$

4. *Clasification Error*
   Classification Error adalah estimasi tingkat kesalahan dalam klasifikasi yang memberikan gambaran tentang seberapa dekat perkiraan kesalahan dengan kesalahan sebenarnya. Rumus perhitungan *classification error* dapat dilihat pada persamaan 7.

$$Classification\ Error = \frac{(FP + FN)}{(TP + FP + FN + TN)} \quad (7)$$

III. METODE

Penelitian ini menerapkan pendekatan metode eksperimen dengan tujuan untuk menguji validitas hipotesis melalui analisis statistik dan mengaitkannya dengan permasalahan yang diangkat. Fokus utama penelitian ini adalah membandingkan dan mengevaluasi model prediksi dengan menggunakan algoritma C4.5 dan Naïve Bayes untuk menentukan algoritma mana yang lebih akurat dalam memprediksi loyalitas pelanggan. Proses analisis dan pengolahan data dilakukan dengan menerapkan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM), seperti yang ditunjukkan pada gambar 1.



Gambar 1 *CRISP-DM Framework*

Metodologi ini melibatkan 6 tahap yang dapat dijelaskan sebagai berikut:

a. *Business Understanding* (Pemahaman Bisnis)
   Tahap ini bertujuan untuk memahami secara mendalam tujuan dari analisis data yang berkaitan dengan identifikasi loyalitas pelanggan. Pemahaman ini mencakup analisis lingkungan layanan ISP yang dapat memengaruhi persepsi pelanggan. Selain itu, tahap ini melibatkan identifikasi kebutuhan spesifik dalam analisis data serta penentuan ruang lingkup dan *boundary* yang jelas. Langkah ini penting untuk memastikan fokus yang tepat dalam mengembangkan model yang efektif untuk mengidentifikasi tingkat loyalitas pelanggan, sehingga menjadi dasar yang kokoh dalam merancang strategi analisis data yang sesuai dengan tujuan penelitian.
b. *Data undestanding* (Pemahaman Data)

Data untuk penelitian ini diperoleh dari sebuah perusahaan yang menyediakan layanan yang bergerak di bidang *tv cable* dan internet di Indonesia. Untuk membatasi data yang akan diuji, maka hanya data pelanggan yang mengajukan permohonan berhenti berlanggan yang akan diuji. Data yang dikumpulkan hanya hingga akhir tahun 2020 dengan membatasi hingga akhir 2020, analisis dap atfokus pada pola yang muncul sebelum perubahan besar yang mungkin terjadi setelah tahun tersebut, sehinnga memberikan landasan yang stabil dan representatif untuk menduga tingkat loyalitas pelanggan ISP, *attribute – attribute* yang akan digunakan meliputi alasan berhenti (*disc_reason*), produk yang digunakan (*product*), cara pelanggan meminta berhenti (*transfer_call*) apakah pelanggan menghubungi langsung dengan transfer call, layanan yang ingin diberhentikan (*disc_service*), tagihan pelanggan (*rate*), bulan terhutang pelanggan (*aging*), saldo mengendap pelanggan (*balance*), umur pelanggan (*cust_age*), apakah pelanggan pernah berhenti sebelumnya (*ever_disc*), dan (*retain*) menyatakan pelanggan loyal atau tidak.

c. *Data Preparation* (Persiapan Data)

Pada tahap *Data Preparation*, penelitian ini menggunakan tahapan sebagai berikut:

1 *Data Collection* : mengumpulkan data yang relevan untuk menjadi penelitian dalam menduga loyalitas pelanggan seperti informasi pengguna, riwayat pengguna layanan internet, feedback pelanggan, data transaksi.

2 *Data Cleaning* : melakukan pemeriksaan untuk data dengan nilai yang kosong yang nantinya akan berdampak mengganggu analisis.

3 *Feature Engineering* : membuat fitur tambahan yang dapat digunakan untuk melakukan analisa dalam mengklasifikasi loyalitas pelanggan seperti tingkat penggunaan layanan (rendah,sedang,tinggi), tingkat kepuasan pelanggan (rendah,sedang,tinggi), durasi langganan yang akan dapat membantu lebih baik dalam menduga loyalitas pelanggan.

d. *Modeling* (Pemodelan)

Fase *Modeling* adalah tahap yang secara langsung terlibat dalam penerapan teknik *data mining* yang meliputi pemilihan algoritma dan penentuan parameter dengan nilai yang optimal. Pada tahap ini, proses pemilihan algoritma dilakukan untuk merancang model analisis yang sesuai dengan kebutuhan penelitian. Tahap ini melibatkan eksplorasi dan evaluasi berbagai teknik *data mining* guna menentukan algoritma yang paling cocok untuk mengatasi permasalahan identifikasi tingkat loyalitas pelanggan provider ISP. Dengan memperhatikan faktor-faktor yang relevan, tahap ini bertujuan untuk mengembangkan model analisis yang efektif dan dapat memberikan hasil yang akurat dan signifikan dalam konteks penelitian yang dilakukan.

e. *Evaluation* (Evaluasi)

Proses evaluasi sangat penting untuk menilai kinerja dan keefektivitas model identifikasi loyalitas pelanggan ISP. Berbagai metrik evaluasi yang relevan untuk penelitian ini mencakup perbandingan antara hasil prediksi model dan kondisi aktual dari data pengujian. Berikut ini adalah penjelasan tentang beberapa metrik evaluasi yang dapat diterapkan:

1 Akurasi (*Accuracy*): Melakukan perbandingan antara jumlah prediksi yang benar (*true positive* dan *true negative*) dengan total jumlah sampel. Akurasi memberikan gambaran umum tentang seberapa efektif model dalam mengklasifikasikan pelanggan loyal/tidak loyal.

2 Presisi (*Precision*): Menunjukkan seberapa banyak dari data pelanggan yang diklasifikasikan sebagai pelanggan loyalitas oleh model yang sesuai sebagai pelanggan yang loyal. Presisi digunakan untuk menilai sejauh mana keakuratan yang diberikan model dalam mengidentifikasi pelanggan loyal.

3 *Recall* (Sensitivitas): Mengukur seberapa banyak pelanggan loyal yang berhasil diidentifikasi dengan benar oleh model. *Recall* bermanfaat untuk menilai seberapa baik model dalam menangkap keseluruhan tingkat loyalitas pelanggan.

4 *Classification* Error: Menunjukkan seberapa banyak tingkat kesalahan klasifikasi dalam memprediksi loyalitas pelanggan dengan menghitung tingkat kesalahan klasifikasi. Semakin rendah nilai Classification Error, semakin baik model tersebut dalam memprediksi loyalitas pelanggan.

f. *Deployment* (Penerapan)

Dalam tahap penerapan penelitian ini melakukan pengujian kepada metode terpilih dan validasi untuk memastikan bahwa kinerjanya sesuai dengan harapan dan mampu memberikan hasil identifikasi loyalitas pelanggan dengan tingkat akurasi yang optimal.

IV. HASIL

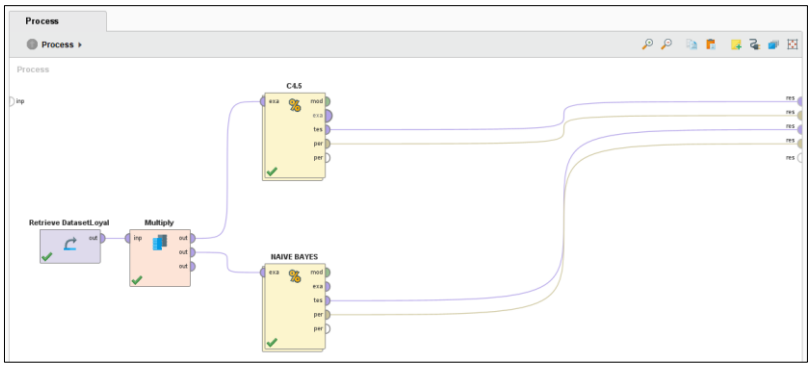
Penelitian ini menggunakan algoritma C4.5 dan Naïve Bayes untuk menganalisis pola loyalitas pelanggan pada perusahaan penyedia layanan ISP TV kabel dan internet di Indonesia. Dataset yang digunakan berasal dari data pelanggan yang mengajukan permintaan penghentian layanan. Proses penghentian ini biasanya dimulai dengan pelanggan menghubungi *Contact Center*, di mana mereka memberikan alasan penghentian layanan. Selama proses ini, pelanggan sering ditawarkan promosi atau alternatif untuk mempertahankan atau mengurangi penggunaan layanan. Dataset telah melalui proses *pre-processing* untuk meningkatkan kualitas data. Tahapan ini meliputi Menghapus *outliers*, *noise*, titik data kosong, serta data yang tidak konsisten.

Hasil pre-processing menghasilkan dataset berkualitas tinggi yang siap untuk analisis lebih lanjut. Pada tabel 2 dibawah, ditampilkan 20 sampel data hasil *pre-processing* sebagai representasi dari dataset keseluruhan.

TABEL 2
TRANSFORMASI DATA

Disc_reason	Product	Transfer_call	Disc_service	Rate	Aging	Balance	Cust_age	Ever_disc	Retain
Price	Combo	Transfer	Disconnect-Combo	365000	0	405000	3	2	Success
Price	Combo	Non Transfer	Disconnect-Combo	410000	0	3000	2	2	Success
Not interest	Combo	Transfer	Disconnect-Combo	60000	4	3025000	4	2	Success
Price	Combo	Transfer	Disconnect-Combo	290000	0	8000	5	2	Failed
Not interest	Inet	Non Transfer	Partial – CATV	335000	1	793468	2	2	Success
Move	Tv	Non Transfer	Partial - CNET	250000	0	289000	3	2	Failed
Service	Tv	Non Transfer	Disconnect-CATV	120000	4	3846736	3	2	Failed
Not interest	Inet	Transfer	Disconnect-CNET	440000	2	1477474	3	2	Success
Proce	Inet	Non Transfer	Partial – CNET	656000	0	12250	4	2	Failed
Switch	Combo	Transfer	Partial – CATV	435000	1	0	6	2	Success
Not interest	Combo	Non Transfer	Partial – CATV	265000	0	352000	2	2	Success
Not interest	Inet	Non Transfer	Partial – CNET	195000	0	437000	3	2	Success
Not interest	Tv	Transfer	Partial – CNET	90000	0	8000	17	1	Failed
Service	Tv	Transfer	Disconnect-Combo	295000	2	123250	3	1	Failed
Service	Inet	Non Transfer	Disconnect-CNET	60000	0	1580000	6	1	Failed
Switch	Tv	Non Transfer	Partial – CNET	100000	4	115000	2	2	Failed
Trial	Combo	Non Transfer	Partial - CNET	100000	0	5000	3	2	Failed
Temporary	Combo	Non Transfer	Disconnect-Combo	30000	0	1674000	49	2	Success
Trial	Inet	Transfer	Partial – CATV	40000	4	93000	22	2	Success
Price	Combo	Non Transfer	Partial – CATV	452000	1	105000	60	1	Success

Pada tahap pengujian ini, dilakukan desain uji algoritma C4.5 dan Naïve Bayes dengan validasi model dilakukan menggunakan metode 10-*fold cross-validation*, yang memastikan keandalan hasil dengan membagi data menjadi 10 subnet dan menguji model pada masing-masing subset secara bergantian. Desain pengujian dapat dilihat pada gambar 2.



Gambar 2 Desain uji *cross validation* algoritma C4.5 dan Naïve Bayes

Pada penelitian ini, evaluasi pengujian model dilakukan dengan menggunakan *confusion Matrix* dan AUC (*Area Under Curve*) pada ROC (*Receiver Operating Characteristic*). Pengujian *confusion matrix* untuk dataset yang diolah menggunakan algoritma C4.5 dijelaskan pada tabel 3.

TABEL 3
HASIL AKURASI ALGORITMA C4.5

	True NO	True LOYAL	Class Precision
Pred. NO	856	172	83.27%
Pred. LOYAL	408	1564	79.31%
Class Recall	67.72%	90.09%	

Perhitungan *accuracy* dari pengujian algoritma C4.5 dapat dilihat pada persamaan 8.

$$Accuracy = \frac{(TN+TL)}{(TN+FN+TL+FL)} \quad (8)$$
$$= \frac{(856 + 1564)}{(856+172+1564+408)}$$
$$= 0.8066 = \mathbf{80.67}\%$$

Perhitungan *precision* dari pengujian algoritma C4.5 dapat dilihat pada persamaan 9.

$$Precision = \frac{TL}{(TL+FL)} \quad (9)$$
$$= \frac{1564}{(1564+408)}$$
$$= 0.7931 = \mathbf{79.31}\%$$

Perhitungan *recall* dari pengujian algoritma C4.5 dapat dilihat pada persamaan 10.

$$Recall = \frac{TL}{(TL+FN)} \quad (10)$$
$$= \frac{1564}{(1564+172)}$$
$$= 0.9009 = \mathbf{90.09}\%$$

Pengujian *confusion matrix* untuk dataset yang diolah menggunakan naïve bayes dapat dilihat pada tabel 4.

TABEL 4
HASIL AKURASI NAÏVE BAYES

	True NO	True LOYAL	Class Precision
Pred. NO	747	196	79.22%
Pred. LOYAL	517	1540	74.87%
Class Recall	59.10%	88.71%	

Pengujian *accuracy* dari pengujian naïve bayes dapat dilihat pada persamaan 11.

$$Accuracy = \frac{(TN+TL)}{(TN+FN+TL+FL)} \quad (11)$$
$$= \frac{(747+1540)}{(747+196+1540+517)}$$
$$= 0.7623 = \mathbf{76.23}\%$$

Pengujian *precision* dari pengujian naïve bayes dapat dilihat pada persamaan 12.

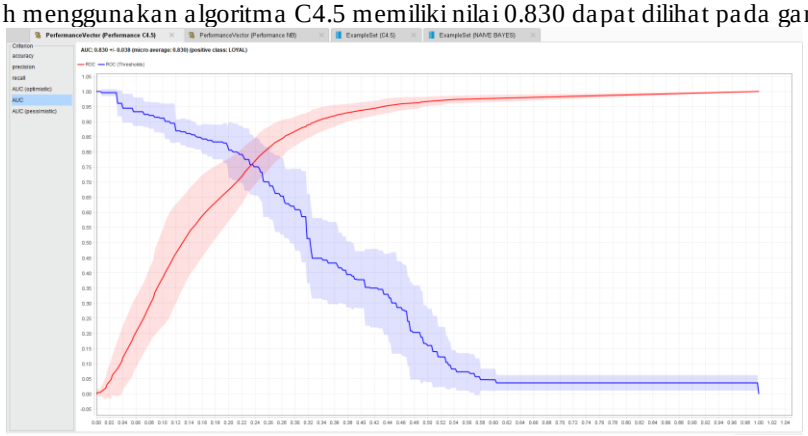
$$Precision = \frac{TL}{(TL+FL)} \quad (12)$$

$$= \frac{1540}{(1540+517)}$$
$$= 0.7486 = \mathbf{74.86}\%$$

Pengujian *recall* dari pengujian naïve bayes dapat dilihat pada persamaan 13.

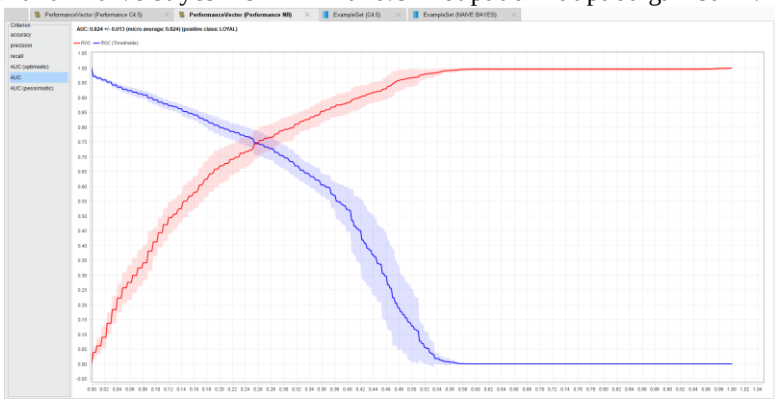
$$Recall = \frac{TL}{(TL+FN)} \qquad (13)$$
$$= \frac{1540}{(1540+196)}$$
$$= 0.8870 = \mathbf{88.70}\%$$

Pengujian selanjutnya dilakukan dengan AUC (*Area Under Curve*) dengan ROC (*Receiver Operating Characteristic*) curve yang akan dilakukan pada dataset yang diolah dengan menggunakan algoritma C4.5 dan naïve bayes. Pengujian AUC (*Area Under Curve*) dengan ROC (*Receiver Operating Characteristic*) *curve* untuk dataset yang diolah menggunakan algoritma C4.5 memiliki nilai 0.830 dapat dilihat pada gambar 3.



Gambar 3 AUC *Curve* Algoritma C4.5

Pengujian AUC (*Area Under Curve*) dengan ROC (*Receiver Operating Characteristic*) *curve* untuk dataset yang diolah menggunakan naïve bayes memiliki nilai 0.824 dapat dilihat pada gambar 4.



Gambar 4 AUC *Curve* naïve bayes

Hasil perbandingan antara algoritma C4.5 dan naïve bayes menghasilkan akurasi yang cukup baik dalam proses klasifikasi pelanggan. Dengan data pelanggan yang akan berhenti sebagai data uji, perbandingan antara data pelanggan loyal dan tidak loyal dapat tetap sebanding. Karena jumlah pelanggan tidak loyal sangat kecil dibandingkan dengan pelanggan loyal. Algoritma C4.5 dapat akurasi yang lebih tinggi yaitu sebesar 80.67%. Sedangkan naïve bayes mendapatkan akurasi cukup baik yaitu sebesar 76.23%. Untuk hasil perbandingan algoritma C4.5 dan naïve bayes ditunjukkan pada tabel 5.

TABEL 5
PERBANDINGAN ALGORITMA C4.5 DAN NAÏVE BAYES

	True NO	**True LOYAL**
Sukses prediksi pelanggan Loyal	1564	1540
Sukses prediksi pelanggan Tidak Loyal	856	747
Nilai Akurasi	80.67%	76.23%
AUC	0.830	0.824

V. PEMBAHASAN

Penelitian ini membandingkan algoritma C4.5 dan Naïve Bayes untuk menduga loyalitas pelanggan pada perusahaan *Internet Service Provider* (ISP). Loyalitas pelanggan sangat penting dalam industri ISP karena berkontribusi pada pendapatan stabil dan reputasi perusahaan.

Hasil penelitian menunjukkan bahwa Algoritma C4.5 dapat melakukan prediksi pelanggan loyal dengan benar sebanyak 1564, sedangkan naïve bayes dapat melakukan prediksi pelanggan loyal dengan benar sebanyak 1540. Algoritma C4.5 dapat melakukan prediksi pelanggan tidak loyal dengan benar sebanyak 856, sedangkan naïve bayes dapat melakukan prediksi pelanggan tidak loyal dengan benar sebanyak 747.

Algoritma C4.5 memiliki nilai akurasi 80.67%, sedangkan naïve bayes memiliki nilai akurasi 76.23%. algoritma C4.5 memiliki nilai AUC 0.830, sedangkan naïve bayes memiliki nilai AUC 0.824. Dengan melihat tabel perbandingan di atas pada tabel 16 dapat diketahui bahwa algoritma C4.5 memiliki nilai akurasi dan nilai AUC yang paling tinggi dengan nilai akurasi 80.67% dan nilai AUC 0.830.

Penelitian ini memberikan kontribusi praktis bagi perusahaan ISP dalam memilih algoritma yang tepat untuk analisis data pelanggan. Langkah selanjutnya dapat mencakup eksplorasi algoritma baru atau kombinasi metode untuk meningkatkan performa prediksi.

VI. KESIMPULAN

Penelitian ini memberikan kontribusi unik terhadap adopsi algoritma dalam industri *Internet Service Provider* (ISP) dengan menunjukkan bagaimana algoritma pembelajaran mesin, khususnya C4.5 dan Naïve Bayes, dapat digunakan untuk memprediksi loyalitas pelanggan secara efektif. Dengan membandingkan performa kedua algoritma, hasil penelitian ini menunjukkan keunggulan C4.5 dalam menangani atribut kompleks dan menghasilkan prediksi yang lebih akurat (akurasi 80,67%, AUC 0,830), dibandingkan dengan Naïve Bayes (akurasi 76,23%, AUC 0,824). Temuan ini memberikan wawasan berharga bagi perusahaan ISP dalam memilih algoritma yang tepat untuk mendukung strategi berbasis data.

Selain itu, penelitian ini menyoroti manfaat praktis dari adopsi algoritma dalam mendukung keputusan strategis, seperti peningkatan kualitas layanan, pengembangan strategi retensi pelanggan yang lebih terarah, dan optimalisasi sumber daya. Dengan menerapkan algoritma yang tepat, perusahaan ISP dapat memanfaatkan data pelanggan untuk mengidentifikasi pola loyalitas, mengurangi tingkat churn, dan meningkatkan kepuasan pelanggan secara keseluruhan.

Kontribusi unik lainnya terletak pada rekomendasi untuk penelitian lebih lanjut, seperti penggunaan dataset yang lebih besar dan beragam, eksplorasi algoritma tambahan seperti Random Forest, serta penerapan teknik optimasi untuk meningkatkan akurasi prediksi. Hal ini membuka peluang bagi industri ISP untuk terus mengembangkan sistem analitik berbasis data yang lebih canggih, sekaligus mendorong penerapan teknologi pembelajaran mesin yang lebih luas dalam mengelola loyalitas pelanggan di industri ISP yang sangat kompetitif.

DAFTAR PUSTAKA

[1] V. Apriana *et al*., "Penerapan Algoritma C4.5 Dalam Memprediksi Loyalitas Konsumen Pada Pt. Hiba Utama," *J. Students' Res. Comput. Sci.*, vol. 4, no. 1, pp. 145–156, 2023, doi: 10.31599/jsrcs.v4i1.2609.

[2] Y. A. Azzahra and Y. Akbar, "Ketepatan Waktu Pengiriman Barang Pada PT . Rtrans Logistik Artamandiri Abstrak," vol. 5, no. 3, pp. 2768–2780, 2024.

[3] M. G. Pradana and P. H. Saputro, "Komparasi Metode Naïve Bayes Dan C4.5 Dalam Klasifikasi Loyalitas Pelanggan Terhadap Layanan Perusahaan," *Indones. J. Bus. Intell.*, vol. 3, no. 1, p. 20, 2020, doi: 10.21927/ijubi.v3i1.1205.

[4] Y. T. Widayati, Y. Prihati, and S. Widjaja, "Pelanggan Mnc Play Kota Semarang," *Transformtika*, vol. 18, no. 2, pp. 161–172, 2021.

[5] E. F. Wati, E. S. Perangin-angin, and L. Indriyani, "Customer Loyalty Classification with Comparison of Naive Bayes , Universitas Bina Sarana Informatika , Indonesia," vol. 8, no. 158, pp. 177–185, 2024.

[6] A. Rahmayanti, L. Rusdiana, and S. Suratno, "Perbandingan Metode Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa," *Walisongo J. Inf. Technol.*, vol. 4, no. 1, pp. 11–22, 2022, doi: 10.21580/wjit.2022.4.1.9654.

[7] D. Yunita and I. H. Ikasari, "Perbandingan Metode Klasifikasi C4.5 dan Naïve Bayes untuk Mengukur Kepuasan Pelanggan," *J. Inform. Univ. Pamulang*, vol. 6, no. 3, pp. 2622–4615, 2021.

[8] B. Ferdiansyah and L. Geoirmanto, "Prediction of Loyalty in Employee Engagement to the Company Using the C4.5* Algorithm (Case Study of PT.XYZ)," *J. Inf. Syst. Technol.*, vol. 8, no. 1, pp. 1–11, 2020.

[9] S. Lestari and A. Suryadi, "Model Klasifikasi Kinerja Dan Seleksidosen Berprestasi Dengan," *Proseding Semin. Bisnis Teknol.*, pp. 15–16, 2014.

[10] E. Jannah, V. Sihombing, and M. Masrizal, "Penerapan Data Mining Klasifikasi Kepuasan Pelanggan Transportasi Online Menggunakan Algoritma C4.5," *MEANS (Media Inf. Anal. dan Sist.*, vol. 8, no. 1, pp.

1–7, 2023, doi: 10.54367/means.v8i1.2569.
[11] E. Prasetyaningrum and P. Susanti, "Analisa Tingkat Kepuasan Pelanggan Pada Percetakan Cv. Mega Media Menggunakan Algoritma C4.5," *Sisfotenika*, vol. 13, no. 1, pp. 65–75, 2023.
[12] R. Rahman and F. A. Sutanto, "Data Mining Untuk Memprediksi Tingkat Kepuasan Konsumen Gojek Menggunakan Algoritma Naive Bayes," *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 18, no. 1, pp. 8–18, 2023, doi: 10.35969/interkom.v18i1.280.
[13] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.
[14] D. Sartika and D. I. Sensuse, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *Jatisi*, vol. 1, no. 2, pp. 151–161, 2017.
[15] M. S. Amrullah, S. F. Pane, and M. N. Fauzan, *Analisis Sentimen Masyarakat Terhadap Kebijakan Polisi Tilang Manual Di Indonesia*. Buku Pedia, 2023.
[16] F. Rahutomo, I. Y. R. Pratiwi, and D. M. Ramadhani, "Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia," *J. Penelit. Komun. Dan Opini Publik*, vol. 23, no. 1, 2019, doi: 10.33299/jpkop.23.1.1805.
[17] T. Tasmalaila Hanifa and S. Al-Faraby, "Analisis Churn Prediction pada Data Pelanggan PT. Telekomunikasi dengan Logistic Regression dan Underbagging," *e-Proceeding Eng.*, vol. 4, no. 2, pp. 3210–3225, 2017.