



KOMPARASI ALGORITMA NAÏVE BAYES DAN ALGORITMA C4.5 DALAM KLASIFIKASI PELANGGAN PRODUK INDIHOME

Abdullah Syafii^{1)*}, Gifthera Dwilestari²⁾, Abdul Ajiz³⁾

^{1,2,3)} Teknik Informatika, STMIK IKMI Cirebon, Cirebon, Jawa Barat
email: abdullahsyafii47@gmail.com¹⁾, ggdwilestari@gmail.com²⁾,
abdulajiz2020@gmail.com³⁾

Abstrak

Teknologi yang semakin cepat menjadikan tiap perusahaan tak terkecuali PT. Telkom Indonesia pada produknya yaitu IndiHome untuk selalu mengikuti kemajuan teknologi dan terus meningkatkan sistemnya dalam pengelolaan data dan informasi yang lebih aman, terukur keakuratan, dan efisien dalam penggunaannya. Sedangkan permasalahan yang dialami terdapat kesulitan dalam mengelolah data secara baik untuk mengklasifikasi pelanggan potensi dan tidak berpotensi. Dalam mengkomparasi penelitian ini menggunakan sebuah model klasifikasi yang merupakan pengelompokan objek kedalam kelas-kelas tertentu berdasarkan kelompok, biasa disebut sebagai kelas. Salah satu metode klasifikasi yang paling umum digunakan adalah metode naive bayes dan C4.5 menggunakan tools Rapidminer. Tahapan penelitian ini menggunakan alur Cross Industry Standard Process for Data Mining (CRISP-DM) yang memiliki tahapan Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment. Tujuan penelitian ini mendapatkan pola pengetahuan terhadap klasifikasi pelanggan IndiHome yang memiliki pelanggan potensi dan tidak berpotensi, sehingga bisa dijadikan sebagai sumber pengambilan keputusan oleh PT telkom indonesia. Dataset yang digunakan sebanyak 1043 record dengan hasil accuracy tertinggi diperoleh pada klasifikasi Naive bayes mendapatkan nilai accuracy 99.81% dengan nilai classification error 0.19%, precision 100%, recall 99.57% dan nilai AUC 0.999 masuk ke dalam kategori Excellent Classification. Sedangkan klasifikasi algoritma C4.5 memperoleh nilai accuracy lebih rendah yaitu 94.63% dengan nilai classification error 5.37%, precision 93.23%, recall 94.84%, dan nilai AUC 0.967 masuk ke dalam kategori Excellent Classification.

Kata Kunci : Naive Bayes Classifier (NBC), C4.5, IndiHome, CRISP-DM.

Abstract

Technology that is getting faster makes every company, including PT. Telkom Indonesia in its product, namely IndiHome, to always keep abreast of technological advances and continue to improve its system in managing data and information that is safer, measurable in accuracy, and efficient in its use. While the problems experienced are difficulties in collecting data properly to classify potential and non-potential customers. This research uses a classification model which is a grouping of objects into specific classes based on groups, commonly referred to as classes. One of the most commonly used classification methods is the Naive Bayes method and C4.5 using Rapidminer tools. The stages of this research use the Cross Industry Standard Process for Data Mining (CRISP-DM) flow which has stages of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and



Deployment. The purpose of this study is to obtain a pattern of knowledge on the classification of IndiHome customers who have potential and non-potential customers so that they can be used as a source of decision making by PT Telkom Indonesia. The dataset used is 1043 records with the highest accuracy results obtained in the Naive Bayes classification getting an accuracy value of 99.81% with a value of classification error 0.19%, precision 100%, recall 99.57%, and an AUC value of 0.999 in the Excellent Classification category. While the classification algorithm C4.5 obtained a lower accuracy value of 94.63% with a value of classification error 5.37%, precision 93.23%, recall 94.84%, and an AUC value of 0.967 into the Good Classification category.

Keywords: Naive Bayes Classifier (NBC), C4.5, IndiHome, CRISP-DM.

PENDAHULUAN

Dengan perkembangan teknologi yang semakin pesat, mendorong dunia bisnis untuk senantiasa bersaing meningkatkan kualitas barang dan jasa. Salah satu perkembangan teknologi tersebut yakni pada industri telekomunikasi dan jaringan internet yang ditandai dengan terdapatnya bermacam produk yang diciptakan guna mempermudah sistem komunikasi untuk masyarakat [1]. Persaingan antara perusahaan yang menjalankan bisnis dibidang internet dan langganan siaran berbayar yang semakin pesat menjadikan PT. Telkom Indonesia dengan produknya yakni IndiHome (Indonesia Digital HOME) harus serius untuk menjaga kualitas dan pelayanan yang baik untuk dapat mempertahankan pelanggan dan memperoleh pelanggan baru yang potensial.

Data mining adalah suatu teknik manajemen data yang menemukan hubungan antara data yang tidak diketahui pengguna dan menyajikannya dalam format yang mudah dipahami, dan hubungan data ini dapat digunakan sebagai dasar pengambilan keputusan. Data mining dapat dibagi menjadi beberapa kelompok, seperti Deskripsi, estimasi, prediksi, klasifikasi, clustering, dan asosiasi terbagi atas dasar tugas yang dapat dilakukan. [2]. Klasifikasi

merupakan suatu metode digunakan untuk memastikan keanggotaan kelompok bersumber pada data-data yang telah ada. Konsep dasar dari klasifikasi merupakan sebagian data yang mempunyai struktur data yang mirip akan mempunyai klasifikasi yang mirip pula [3].

Menurut penelitian terdahulu mengenai Klasifikasi Pelanggan Produk IndiHome pernah dilakukan oleh Indah Purnamasari, permasalahan yang diangkat pada penelitian ini adalah banyaknya penawaran produk baru yang ditolak oleh pelanggan eksisting. Pendekatan yang digunakan metode algoritma naïve bayes classifier dengan algoritma genetik untuk mengklasifikasikan pelanggan. Hasil dari penelitian menjelaskan bahwa dengan pemilihan fungsi Algoritma Genetika telah membuat model Naive Bayes Classifier menjadi model klasifikasi yang lebih baik untuk diterapkan pada klasifikasi pelanggan yang ada. PT. TELKOM dalam memasarkan produk baru Indihome. Akurasi model Naive Bayes Classifier adalah 85,08% dan 0.841 untuk AUC. Akurasi yang dihasilkan model Naive Bayes Classifier menggunakan pemilihan fitur Algoritma Genetika meningkat menjadi 89,31% untuk accuracy dan 0.843 untuk AUC [4].



Permasalahan yang dialami di PT. Telkom Cirebon adalah terdapat kesulitan dalam mengelolah data secara baik untuk mengklasifikasi pelanggan potensi dan tidak berpotensi. Usulan yang akan dilakukan dalam penelitian ini yaitu menggunakan algoritma naïve bayes classifier dan C4.5 dalam klasifikasi data pelanggan. Kedua algoritma ini akan dikomparasikan untuk mengklasifikasi berdasarkan label atribut dan karakteristik data agar dapat diperoleh tingkat akurasi terbaik dalam proses data mining.

Kajian yang akan diusulkan penulis ialah mengenai pengklasifikasian data terhadap pelanggan telepon dan pelanggan internet fiber eksisting berdasarkan tipe layanan yang ada di PT. Telkom Cirebon. Teknik yang digunakan adalah Supervised learning untuk menemukan hubungan antara atribut input dan atribut target atau label.

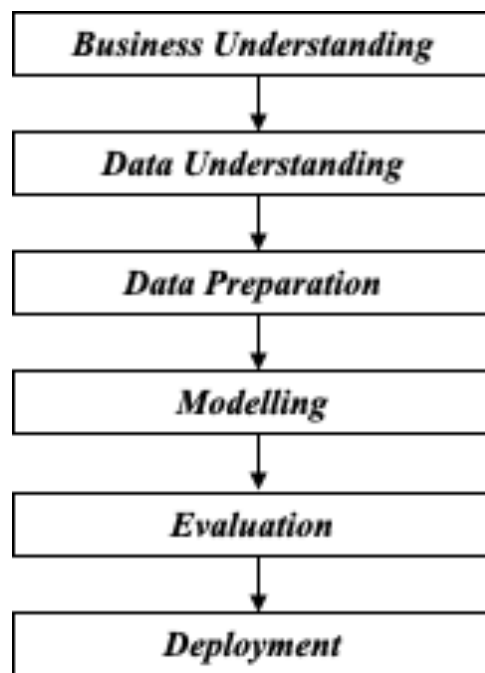
Tujuan utama dari penelitian ini adalah untuk komparasi klasifikasi data pelanggan menggunakan pendekatan CRISP-DM dengan Algoritma naïve bayes dan C4.5 untuk memperoleh tingkat akurasi terbaik agar dapat membantu dalam proses pengambilan keputusan guna memberikan informasi produk baru kepada pelanggan yang memiliki potensi.

METODE PENELITIAN

Jenis penelitian yang peneliti gunakan dalam penelitian ini merupakan jenis penelitian kuantitatif dan penelitian deskriptif. Penelitian kuantitatif yakni penelitian dengan data berupa nilai numerik dan menganalisisnya menggunakan statistik [5]. Penelitian deskriptif ialah studi yang berusaha terutama untuk memberikan refleksi atau penjelasan objektif tentang kondisi tersebut. Penelitian deskriptif dapat digunakan untuk memaparkan tentang

pelaksanaan pengendalian internal dalam sesuatu institusi [6].

Penelitian ini akan menggunakan metode Cross Industry Standard Process for Data Mining (CRISP-DM) untuk analisis dan pengolahan data. CRISP-DM adalah metode yang diprakarsai oleh beberapa perusahaan terkemuka Daimler-Benz, dengan tujuan memformalkan dan menstandarisasi pendekatan data mining. Fase CRISP-DM dimulai dengan Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment [7]. Adapun proses yang dilakukan adalah sebagai berikut:



Gambar 1. Proses CRISP-DM

HASIL DAN PEMBAHASAN

1. Business Understanding

Dalam penelitian ini pelanggan dibagi menjadi dua kelas: Pelanggan potensial dan tidak potensial. Adapun karakteristik dari kategori pelanggan potensial dan tidak potensial didefinisikan pada Tabel 1.



Tabel 1. Kategori Pelanggan

POTENSIAL	TIPE LAYANAN	KECEPATAN INTERNET	TIPE TRANSAKSI
		3P	≥ 30 Mbps
TIDAK POTENSIAL	2P	≤ 30 Mbps	INTERNET+IPTV Dan TLP+INTERNET

Pada kategori pelanggan Potensial akan dimasukkan ke dalam kategori HVC (High Value Customer) IndiHome dengan selalu menjaga kualitas pelayanan dan akan mendapatkan treatment khusus berupa reward dan prioritas layanan maintenance lebih baik kepada pelanggan. Kategori pelanggan Tidak Potensial akan ditindak lanjuti dengan melakukan penawaran kepada pelanggan untuk melakukan peningkatan layanan IndiHome.

2. Data Understanding

Pada penelitian ini menggunakan data yang berasal dari proses order pelanggan pemasangan indihome dan kemudian data tersebut disimpan untuk dataset menggunakan Microsoft Excel. Tabel yang digunakan untuk mengklasifikasikan pelanggan adalah tabel transaksi yang diperoleh dari proses order pelanggan pemasangan indihome. Atribut tabel terdiri atas Tanggal Order, Nama Customer, ND INTERNET, ND VOICE, Nama Alpro, Tipe Transaksi, STO, KECEPATAN, Tipe Layanan, DATEL, Kategori.

1. Tanggal Order: Tanggal input order pelanggan
2. Nama Customer: Nama pelanggan pemasangan indihome
3. ND INTERNET: Nomor internet pelanggan
4. ND VOICE: Nomor telpon kabel pelanggan
5. Nama Alpro: Nama perangkat jaringan optical distribution point yang terpasang di pelanggan

6. Tipe Transaksi: Jenis order pelanggan (TLP+INET+IPTV,INTERNET+IPTV, TLP+INETERNET)
7. STO: Sentral Telepon Otomat merupakan kode wilayah layanan control station pusat (AWN, BON, CBN, CKC, CKI, CKY, CLI, HAR, IMY, JBN, JTB, JTW, KAD, KNG, KRM, LOS, LSR, MJL, PAB, PRD, PTR, RGA, SDU)
8. KECEPATAN: Kecepatan koneksi internet (20 Mbps, 30 Mbps, 40 Mbps, 50 Mbps, 100 Mbps, 200 Mbps, 300 Mbps)
9. Tipe Layanan: Jenis layanan pelanggan saat ini (2P, 3P)
10. DATEL: Daerah Telkom (Cirebon, Indramayu, Kuningan, Majalengka)
11. Kategori : Kategori pelanggan (POTENSIAL, TIDAK POTENSIAL)

Tabel 2. Sumber Data

NO	Tanggal Order	Nama Customer	ND INTERNET	ND VOICE	Nama Alpro	Tipe Transaksi	STO	KECEPATAN	Tipe Layanan	DATEL	Kategori
1	20/01/2022	MOH MUS LIM	131266106467	2318862xxx	ODP-AW-N-FAA/052	TL-P+INE-T+I	AWN	30 Mbps	3P	CIREBON	POTENSIAL
2	20/01/2022	DIDI TURMU DI	131241107872	2338318xxx	ODP-CKL-FAG/02	TL-P+INE-T+I	CKI	30 Mbps	3P	MAJALENGKA	POTENSIAL
3	20/01/2022	YOHANIS	131232157264	2318250xxx	ODP-CBN-FDV/01	TL-P+INE-T+I	CBN	30 Mbps	3P	CIREBON	POTENSIAL
4	20/01/2022	YOHANIS	131240121570	2338296xxx	ODP-MIL-FAM/04	TL-P+INE-T+I	MJL	30 Mbps	3P	MAJALENGKA	POTENSIAL
10	01/2022	SRIWAHYU NI	131239102866	2328912xxx	ODP-KNG-FCY/004	TL-P+INE-T+I	KNG	30 Mbps	3P	KUNINGAN	POTENSIAL

3. Data Preparation

Dalam tahap data selection dilakukan untuk memilih atribut yang akan digunakan dalam proses klasifikasi pelanggan. Atribut

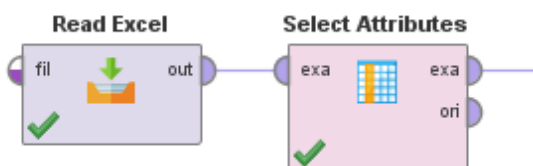


ini disebut predictor attribute, dan 1 atribut class label yang merupakan hasil dari akhir prediksi. Diketahui jumlah data yang di uji sebanyak 1043 record terdiri atas : Tipe Transaksi, KECEPATAN, Tipe Layanan, DATEL, Kategori.

Tabel 3. Pembagian Variable Prediktor dan Variable Target

Nama	Tipe	Jenis
Tipe Layanan	Polynomial	Atribut
Tipe Transaksi	Polynomial	Atribut
KECEPATAN	Polynomial	Atribut
DATEL	Polynomial	Atribut
Kategori	Binominal	Label

Alat bantu untuk proses pengolahan data mining pada penelitian ini menggunakan Rapidminer Studio 9.5.001 untuk proses klasifikasi pelanggan.



Gambar 2. Data Preparation

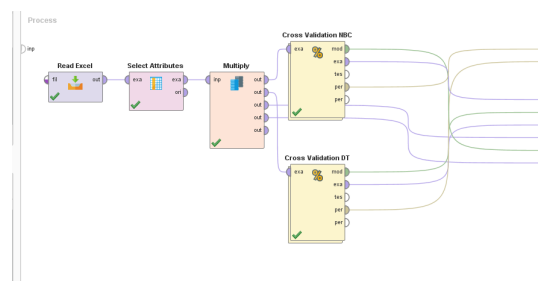
Berdasarkan gambar diatas adalah operator pada rapidminer, operator Read Excel digunakan untuk input dataset dengan format file excel atau .xlsx dan operator Select Attributes digunakan untuk proses seleksi data memilih atribut yang akan digunakan dalam proses pengklasifikasian pelanggan.

Row No.	Kategori	Tipe Transaksi	DATEL	KECEPATAN	Tipe Layanan
1	POTENSIAL	TLP+INET+IPTV	CIREBON	30 Mbps	3P
2	POTENSIAL	TLP+INET+IPTV	MAJALENGKA	30 Mbps	3P
3	POTENSIAL	TLP+INET+IPTV	CIREBON	30 Mbps	3P
4	POTENSIAL	TLP+INET+IPTV	MAJALENGKA	30 Mbps	3P
5	POTENSIAL	TLP+INET+IPTV	CIREBON	30 Mbps	3P
6	TIDAK POTE...	INTERNET+IPTV	CIREBON	30 Mbps	2P
7	TIDAK POTE...	TLP+INTERNET	CIREBON	20 Mbps	2P
8	POTENSIAL	TLP+INET+IPTV	KUNINGAN	30 Mbps	3P
9	TIDAK POTE...	TLP+INTERNET	CIREBON	30 Mbps	2P
10	TIDAK POTE...	TLP+INTERNET	CIREBON	20 Mbps	2P
11	POTENSIAL	TLP+INET+IPTV	MAJALENGKA	30 Mbps	3P
12	POTENSIAL	TLP+INET+IPTV	MAJALENGKA	30 Mbps	3P
13	TIDAK POTE...	TLP+INTERNET	MAJALENGKA	20 Mbps	2P
14	TIDAK POTE...	INTERNET+IPTV	CIREBON	20 Mbps	2P

Gambar 3. Hasil Select Attributes

4. Modeling

Tahap Modeling dalam penelitian ini dilakukan dengan teknik data mining, yakni klasifikasi menggunakan algoritma Naive Bayes dan C4.5 atau Decision Tree. Dilakukan pemrosesan data training sesuai tujuan untuk menghasilkan keputusan dari proses klasifikasi untuk menentukan pelanggan potensial dan tidak potensial.



Gambar 4. Proses Modeling Klasifikasi Pelanggan

Pada proses modeling klasifikasi pelanggan menggunakan beberapa operator pada rapidminer, berikut keterangan dan fungsi dari operator-operator tersebut :

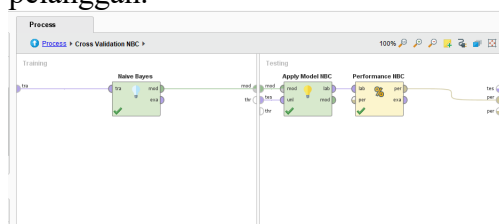
1. Read Excel: untuk input dataset dengan format file excel atau .xlsx.
2. Select Attributes: digunakan untuk proses seleksi data memilih atribut yang akan digunakan dalam proses pengklasifikasian pelanggan.



3. Multiply: digunakan untuk membuat salinan objek dari port input dan mengirimkan salinan tersebut ke port output.
4. Cross Validation: operator ini mengerjakan validasi silang untuk menghitung kinerja model. Operator cross validation dibagi menjadi dua subproses, yaitu subproses training untuk proses melatih model dan subproses testing untuk proses pengujian dan pengukuran kinerja model.

a. Algoritma Naive Bayes Classifier

Setelah memasukkan data ke dalam tools Rapid Miner, dapat menggunakan algoritma Naive Bayes untuk melakukan proses klasifikasi pelanggan.

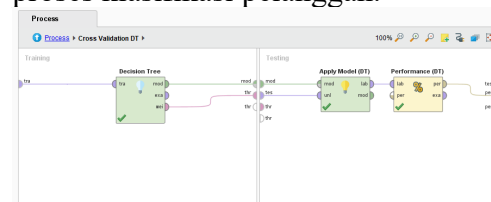


Gambar 5. Proses Cross Validation Naive Bayes

Berdasarkan gambar diatas dilakukan proses pembagian data menjadi data training dan data testing didalam operator Cross Validation menggunakan Algoritma Naive Bayes. Pada subproses training menggunakan operator naive bayes sebagai penerapan model algoritma. Sedangkan subproses testing digunakan operator Apply Model untuk menguji data hasil operator naive bayes, selanjutnya operator performance digunakan untuk menampilkan hasil evaluasi algoritma yang diukur kinerjanya.

b. Algoritma C4.5

Setelah memasukkan data ke dalam tools Rapid Miner, dapat menggunakan algoritma C4.5 untuk melakukan proses klasifikasi pelanggan.



Gambar 6. Proses Cross Validation C4.5

Berdasarkan gambar diatas dilakukan proses pembagian data menjadi data training dan data testing pada operator Cross Validation menggunakan Algoritma C4.5 atau Decision Tree. Pada subproses training menggunakan operator Decision Tree sebagai penerapan model algoritma. Sedangkan subproses testing digunakan operator Apply Model untuk menguji data hasil operator naive bayes, selanjutnya operator performance digunakan untuk menampilkan hasil evaluasi algoritma yang diukur kinerjanya.

5. Evaluation

Setelah pola klasifikasi sudah didapatkan pada Algoritma Naive Bayes dan Algoritma C4.5 selanjutnya akan dilakukan komparasi atau perbandingan algoritma tersebut dengan parameter-parameter dari Confusion Matrix untuk mendapatkan informasi tentang perbandingan hasil yang telah dilakukan. Pada tahap evaluasi akurasi klasifikasi secara visual, akan memodelkan hasil klasifikasi dengan aturan nilai akurasi, presisi dan recall serta ROC Curve (Receiver Operating Characteristic).



a. Evaluation Algoritma Naive Bayes

Parameter Confusion Matrix membentuk sebuah matrix terdiri dari true positive, true negative, false positive, dan false negative

	true POTENSIAL	true TIDAK POTENSIAL	class precision
pred. POTENSIAL	463	0	100.00%
pred. TIDAK POTENSIAL	2	578	99.66%
class recall	99.57%	100.00%	

Gambar 7. Hasil Klasifikasi Naive Bayes

Dari hasil klasifikasi algoritma naive bayes diperoleh hasil akurasi sebesar 99.81%. Pada detail jumlah True Positive (TP) 463, True Negative (TN) 578, False Positive (FP) 0, dan False Negative (FN) 2.

True POTENSIAL dan True TIDAK POTENSIAL adalah kelas asli atau nilai aktual. Pred. POTENSIAL dan pred. TIDAK POTENSIAL adalah kelas prediksi atau nilai prediksi. Class precision adalah kelas yang mengukur tingkat akurasi antara informasi yang diinginkan pengguna dan hasil prediksi yang diberikan dari sistem. Class recall adalah kelas yang menilai hasil yang diberikan dari sistem dalam memprediksi sebuah informasi. Adapun nilai pengukuran klasifikasi yang diperoleh pada Algoritma Naive Bayes, dapat dihitung dengan menggunakan rumus sebagai berikut:

1. Accuracy

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\%$$

$$\text{Accuracy} = \frac{463 + 578}{463 + 578 + 0 + 2} \times 100\%$$

$$\text{Accuracy} = 0.9981 = 99.81\%$$

Dari perhitungan tersebut menunjukkan hasil pengujian klasifikasi menggunakan Algoritma Naive Bayes memiliki nilai akurasi 99.81%. Artinya sistem dapat mengklasifikasikan pelanggan berdasarkan pelanggan potensial dan tidak potensial sejauh 99.81% dengan tingkat classification_error sebesar 0.19%. Dengan perhitungan sebagai berikut :

$$\text{Classification_error} = 1 - \text{Accuracy} \times 100\% = 1 - 0.9981 \times 100\% = 0.19\%$$

2. Precision

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\%$$

$$\text{Precision} = \frac{463}{463 + 0} \times 100\%$$

$$\text{Precision} = 100 = 100\%$$

Precision pada kelas Potensial sebesar 100% menunjukkan bahwa persentase pelanggan Potensial yang benar merupakan pelanggan potensial dari keseluruhan data pelanggan yang diprediksi Potensial.

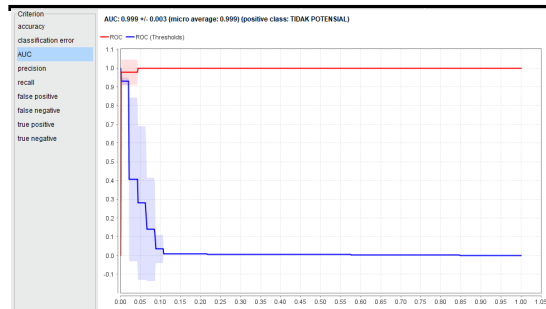
3. Recall

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%$$

$$\text{Recall} = \frac{463}{463 + 2} \times 100\%$$

$$\text{Recall} = 0.9957 = 99.57\%$$

4. ROC Curve (Receiver Operating Characteristic)



Gambar 8. ROC Curve Naive Bayes

Dengan hasil AUC sebesar 0.999, model Naive Bayes ini dapat ditunjukkan sebagai model yang sangat baik untuk digunakan sebagai klasifikasi pelanggan dengan akurasi $0.90 - 1.00 =$ Excellent classification.

b. Evaluation Algoritma C4.5

Parameter confusion matrix akan membentuk sebuah matrix terdiri dari true positive, true negative, false positive, dan false negative.

	true POTENSIAL	true TIDAK POTENSIAL	class precision
pred. POTENSIAL	441	32	93.23%
pred. TIDAK POTENSIAL	24	546	95.79%
class recall	94.84%	94.46%	

Gambar 9. Hasil Klasifikasi C4.5

Dari hasil klasifikasi algoritma C4.5 diperoleh hasil akurasi sebesar 94.63%. Pada detail jumlah True Positive (TP) 441, True Negative (TN) 546, False Positive (FP) 32, dan False Negative (FN) 24.

True POTENSIAL dan True TIDAK POTENSIAL adalah kelas asli atau nilai aktual. Pred. POTENSIAL dan pred. TIDAK POTENSIAL adalah kelas prediksi atau nilai prediksi. Class precision adalah kelas yang mengukur tingkat akurasi antara informasi yang diinginkan pengguna dan hasil prediksi yang diberikan dari sistem. Class recall adalah kelas yang menilai hasil

yang diberikan dari sistem dalam memprediksi sebuah informasi.



Gambar 10. Grafik Tree Hasil Klasifikasi C4.5

Adapun nilai pengukuran klasifikasi yang diperoleh pada Algoritma C4.5, dapat dihitung dengan menggunakan rumus sebagai berikut :

1. Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$\text{Accuracy} = \frac{441 + 546}{441 + 546 + 32 + 24} \times 100\%$$

$$\text{Accuracy} = 0.9463 = 94.63\%$$

Dari perhitungan tersebut disimpulkan hasil pengujian klasifikasi menggunakan Algoritma C4.5 atau Decision Tree memiliki nilai akurasi 94.63%. Artinya sistem dapat mengklasifikasikan pelanggan berdasarkan pelanggan potensial dan tidak potensial sejauh 94.63% dengan tingkat classification_error sebesar 5.37%. Dengan perhitungan sebagai berikut :

$$\text{Classification_error} = 1 - \text{Accuracy} \times 100\% = 1 - 0.9463 \times 100\% = 5.37\%$$

2. Precision

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%$$

$$\text{Precision} = \frac{441}{441 + 32} \times 100\%$$

$$\text{Precision} = 0.9323 = 93.23\%$$



Precision pada kelas Potensial sebesar 93.23% menunjukkan bahwa persentase pelanggan Potensial yang benar merupakan pelanggan potensial dari keseluruhan data pelanggan yang diprediksi Potensial.

3. Recall

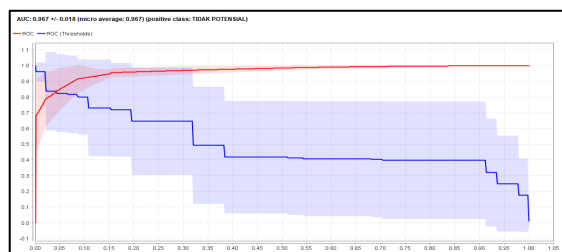
$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%$$

$$\text{Recall} = \frac{441}{441 + 24} \times 100\%$$

$$\text{Recall} = 0.9484 = 94.84\%$$

Recall untuk kelas Potensial sebesar 94.84% menunjukkan bahwa persentase pelanggan Potensial yang diprediksi merupakan benar pelanggan Potensial dari keseluruhan data pelanggan.

4. ROC Curve (Receiver Operating Characteristic)



Gambar 11. ROC Curve C4.5

Hasil AUC adalah 0.967 sehingga dapat menunjukkan bahwa model C4.5 ini merupakan model yang Baik digunakan sebagai klasifikasi pelanggan bahwa akurasi $0.90 - 1.00 =$ Excellent classification.

6. Deployment

Perbandingan algoritma klasifikasi pada penelitian ini berdasarkan hasil output pada kedua teknik klasifikasi yaitu Naive

Bayes dan C4.5 menggunakan tools Rapidminer dengan mode *Cross Validation* dengan nilai $k \text{ fold}=10$. Tabel 2 menunjukkan hasil perbandingan kinerja dari masing-masing Algoritma klasifikasi pada proses pengklasifikasian pelanggan potensial dan tidak potensial.

Pengukuran	HASIL	
	Naive Bayes	C4.5
Accuracy	99.81%	94.63%
Classification Error	0.19%	5.37%
Precision	100%	93.23%
Recall	99.57%	94.84%
AUC	0.999	0.967

Tabel 4. Komparasi algoritma Naive Bayes dan C4.5

Pada penelitian ini dilakukan untuk menganalisa perbandingan algoritma klasifikasi dengan data mining untuk pengklasifikasian pelanggan potensial dan tidak potensial. Berdasarkan hasil pengujian terhadap kedua model algoritma klasifikasi yaitu Naive Bayes dan C4.5 dengan menggunakan dataset yang sama maka nilai accuracy tertinggi diperoleh pada klasifikasi Algoritma Naive bayes yang mendapat nilai 99.81% dengan Classification Error 0.19%, Precision 100%, dan Recall 99.57%. Sedangkan klasifikasi Algoritma C4.5 memiliki accuracy lebih rendah yaitu 94.63% dengan Classification Error 5.37%, Precision 93.23%, dan Recall 94.84%.

Perbandingan hasil pengujian Area Under Curve (AUC), pengukuran Naive Bayes tertinggi yaitu 0.999 memiliki nilai AUC diantara 0.90 – 1.00 masuk ke dalam kategori Excellent Classification.



Sedangkan algoritma C4.5 nilai AUC lebih rendah yaitu 0.967 memiliki nilai AUC diantara 0.90 – 1.00 masuk ke dalam kategori Excellent Classification. Semakin tinggi nilai Area Under Curve (AUC) maka performansi model klasifikasi semakin baik.

Berdasarkan hasil pengukuran klasifikasi menggunakan kedua algoritma yaitu Naive Bayes dan C4.5, Diperoleh hasil bahwa nilai akurasi algoritma Naive Bayes lebih akurat dibandingkan dengan C4.5. Dengan perhitungan jumlah data yang sama bisa dikatakan penggunaan algoritma Naive Bayes lebih baik daripada algoritma C4.5 atau Decision Tree. Berikut akan dijabarkan hasil klasifikasi kedua algoritma tersebut.

KESIMPULAN

Berdasarkan penjelasan yang diberikan, maka dapat diambil kesimpulan sebagai berikut :

1. Nilai akurasi yang diperoleh model Naive Bayes sebesar 99.81% dan nilai akurasi model C4.5 memiliki nilai 94.63%.
2. Hasil komparasi klasifikasi menggunakan algoritma naïve bayes dan C4.5 menyatakan bahwa algoritma naïve bayes menjadi model klasifikasi terbaik yang dapat digunakan pada klasifikasi pelanggan IndiHome untuk menentukan pelanggan potensial dan tidak potensial.

SARAN

Saran untuk penelitian yang akan dilakukan selanjutnya agar menambahkan atribut seperti jumlah pembayaran dan tanggal bayar untuk memperoleh hasil dan pola yang paling baik bisa dioptimalkan menggunakan metode lain sebagai

perbandingan pada metode Naive Bayes Classifier dan C4.5.

DAFTAR PUSTAKA

- [1] M. Chaidir, “Analisis Data Pelanggan Indihome Dengan Algoritma Rough Set Untuk Memprediksi Calon Pelanggan Baru Yang Potensial (Studi Kasus : Pt . Telekomunikasi Indonesia , Tbk),” *J. Pelita Inform.*, vol. 8, no. 2, pp. 198–203, 2019.
- [2] H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, “Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan,” *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.
- [3] Y. I. Kurniawan, “Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 455, 2018, doi: 10.25126/jtiik.201854803.
- [4] I. Purnamasari, “Klasifikasi Pelanggan Produk IndiHome Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Algoritma Genetik,” *J. Tek. Komput.*, vol. 4, no. 1, pp. 8–16, 2018.
- [5] I. Imron, “Analisa Pengaruh Kualitas Produk Terhadap Kepuasan Konsumen Menggunakan Metode Kuantitatif Pada CV. Meubele Berkah Tangerang,” *Indones. J. Softw. Eng.*, vol. 5, no. 1, pp. 19–28, 2019, doi: 10.31294/ijse.v5i1.5861.
- [6] D. S. Purnia, M. F. Adiwisatra, H. Muhajir, and D. Supriadi, “Pengukuran Kesenjangan Digital Menggunakan Metode Deskriptif Berbasis Website,” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 2, 2020, doi: 10.31294/evolusi.v8i2.8942.
- [7] A. Rohanah, D. L. Rianti, and B. N. Sari, “Perbandingan Naive Bayes dan



-
- Support Vector Machine untuk Klasifikasi Ulasan Pelanggan Indihome,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 1, p. 23, 2021, doi: 10.30998/string.v6i1.9232.
- [8] Ahyar, H. dkk. (2020). *Buku Metode Penelitian Kualitatif & Kuantitatif*. Yogyakarta: CV. Pustaka Ilmu. (Issue March).
- [9] Arfandi, Windarto, A. P., & Saragih, I. S. (2021). *Penerapan Data Mining Klasifikasi Pada Calon Pelanggan Baru Indihome dengan C.45*. 1(1), 31–38.
- [10] Bastian, Y., Tambunan, H. S., & Saputra, W. (2021). Analisis Penerapan Algoritma C4 . 5 Dalam Mengukur Tingkat Kepuasan Pelanggan Indihome Pada Kota Pematangsiantar. *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, 2(1), 62–69.
- [11] Umam, K., Puspitasari, D., & Nurhadi, A. (2020). Penerapan Algoritma C4.5 Untuk Prediksi Loyalitas Nasabah PT Erdika Elit Jakarta. *Jurnal Media Informatika Budidarma*, 4(1), 65. <https://doi.org/10.30865/mib.v4i1.1652>.
- [12] Widiawati, W. Y., & Atok, R. M. (2018). Analisis Klasifikasi Pelanggan Listrik Rumah Tangga Bersubsidi Kota Surabaya Menggunakan Support Vector Machine dan Naïve Bayes Classifier. *Jurnal Sains Dan Seni ITS*, 7(2). <https://doi.org/10.12962/j23373520.v7i2.35366>.