

Perbandingan Metode Klasifikasi *Data Mining* untuk Nasabah *Bank Telemarketing*

Pungkas Subarkah, Enggar Pri Pambudi, Septi Oktaviani Nur Hidayah
Universitas Amikom Purwokerto, Indonesia

Article Info

Article history:

Received, 23 July 2020

Revised, 7 August 2020

Accepted, 8 September 2020

Kata kunci:

Bank telemarketing

Klasifikasi

CART

Naive bayes

Data Mining

ABSTRAK

Bank merupakan perusahaan yang memiliki data yang besar yang tersimpan di dalam *database* dan diolah menghasilkan sebuah informasi yang saling berkaitan tentang nasabah. *Bank*, harus memiliki ide dan terobosan baru guna mengetahui kendala pada nasabah telemarketing yang ingin melakukan deposito pada *Bank* tersebut, agar *Bank* terhindar dari ancaman krisis keuangan. Penelitian ini menguji keberhasilan *Bank* telemarketing dengan cara melakukan klasifikasi keputusan nasabah dengan menerapkan *data mining*. Metode yang di gunakan algoritma *Classification and Regression Trees* (CART) dan *naive bayes* menggunakan *dataset* diambil dari *University of California Irvine* (UCI) *Repository Learning*. Adapun metode validasi dan evaluasi yang digunakan yaitu *10-cross validation* dan *confusion matrix*. Hasil akurasi pada algoritma CART yaitu 89.51% dengan nilai *precision* 87%, *Recall* 89% dan *F-Measure* 88% dan pada algoritma *naive bayes* mendapatkan nilai akurasi sebesar 86.88% dengan nilai *precision* 87%, *Recall* 86% dan *F-Measure* 87%. Dari hasil tersebut dapat disimpulkan bahwa algoritma CART lebih baik dalam memprediksi keputusan nasabah telemarketing tepat dalam penawaran deposito.

ABSTRACT

Banks are companies that have large data stored in a *database* and processed to produce interrelated information about customers. *Banks* must have new ideas and breakthroughs in order to find out the obstacles to telemarketing customers who want to make deposits with the *Bank*, so that the *Bank* can avoid the threat of financial crisis. This study examines the success of telemarketing banks by classifying customer decisions by implementing *data mining*. The method used is the *Classification and Regression Trees* (CART) algorithm and *Naive Bayes* using a *dataset* taken from *University of California Irvine* (UCI) *Repository Learning*. The validation and evaluation methods used are *10-cross validation* and *confusion matrix*. The accuracy of the CART algorithm is 89.51% with a *precision* value of 87%, *Recall* 89% and *F-Measure* 88% and the *Naive Bayes* algorithm gets an accuracy value of 86.88% with a *precision* value of 87%, *Recall* 86% and *F-Measure* 87%. From these results it can be concluded that the CART algorithm is better at predicting the correct telemarketing customer decisions in the deposit offer

Keywords:

Telemarketing bank

Classification

CART

Naive bayes

Data Mining

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Korespondensi:

Pungkas Subarkah,

Program Studi Informatika,

Universitas Amikom Purwokerto.

Email: subarkah@amikompurwokerto.ac.id

1. PENDAHULUAN

Bank merupakan perusahaan yang memiliki data yang besar yang tersimpan di dalam database dan diolah menghasilkan sebuah informasi yang saling berkaitan tentang nasabah, data tersebut serta dapat digunakan untuk menjaga hubungan antar bank dengan nasabah yang valid, sehingga berguna untuk menentukan secara individual tentang penawaran produk bank [1].

Masalah utama yang dihadapi dalam *bank telemarketing* ini merupakan bagaimana untuk mencapai sebuah akurasi yang tinggi dalam proses klasifikasi berdasarkan kecepatan informasi tertentu yang diperoleh dari nasabah dan dianggap penting oleh pihak bank [2]. Hal ini mempengaruhi pada pihak *bank telemarketing* dalam menentukan pengambilan keputusan, karena akurasi yang tinggi pada proses klasifikasi. Pihak bank harus memiliki cara bagaimana untuk mengetahui kendala pada nasabah *bank telemarketing* yang ingin melakukan deposito di bank [3]. Target utama pada sebuah perusahaan *bank telemarketing* yaitu mencoba memprediksi harapan terhadap nasabah yang memiliki kemungkinan tertinggi dalam pelayanan dengan menggunakan teknik *data mining* [4]. Pada bank harus selektif dalam memilih nasabah yang menerima pinjaman kredit [5].

Perusahaan Bank, harus memiliki ide dan terobosan baru guna mengetahui kendala pada nasabah telemarketing yang ingin melakukan deposito pada bank tersebut, yang bertujuan agar bank terhindar dari ancaman krisis keuangan. Dengan kata lain Bank harus berupaya meningkatkan modal perbankan, salah satunya dengan cara menawarkan produk, yaitu deposito baik jangka pendek ataupun jangka panjang.

Bank telemarketing harus memiliki dan membuat sebuah rencana untuk target nasabah mana yang berpotensi, agar dapat ditawarkan dan melakukan deposito dengan menganalisis data klien bank yang ada [3]. *Data mining* adalah sebuah proses yang digunakan untuk memanipulasi data dengan mengekstraksi informasi yang sebelumnya tidak diketahui dari dataset yang berukuran besar [6]. Saat ini, peran penggunaan data mining sering digunakan pada beberapa industri termasuk asuransi dan perbankan. Pada penggunaan teknik data mining dalam *bank telemarketing* bertujuan untuk menganalisa data pelanggan dan mengembangkan data pelanggan secara statistik berdasarkan produk dan pelayanan yang lebih diminati oleh pelanggan. Dalam menyelesaikan permasalahan klasifikasi, pada penggunaan metode atau teknik bertujuan untuk mempermudah pada proses klasifikasi. Dengan semakin pesat perkembangan ilmu *data mining* banyak penelitian yang digunakan untuk memprediksi suatu kasus khususnya perbankan, seperti klasifikasi persetujuan kredit [7], nasabah kredit potensial [8] dan nasabah deposito sesuai target [9].

Penelitian [10] melakukan komparasi metode *data mining* dalam memprediksi nasabah bank yang akan memilih tabungan deposito menggunakan algoritma klasifikasi. Penelitian tersebut menggunakan algoritma klasifikasi yaitu algoritma *Decision Tree*, *K-Nearest Neighbour*, dan *naive bayes*. Hasil yang didapatkan yaitu algoritma *decision tree* mendapatkan akurasi yang paling baik yaitu sebesar 88.48% , sedangkan algoritma *naive bayes* menghasilkan tingkat akurasi sebesar 86.84%, dan algoritma *K-Nearest Neighbour* tingkat akurasinya sebesar 84.96%.

Penelitian [9] membahas tentang kendala yang dialami nasabah dalam pengambilan keputusan untuk melakukan deposito terhadap bank. Penelitian tersebut menggunakan teknik *data mining* yaitu algoritma *naive bayes* dan genetika. Dengan menggunakan pengujian *cross validation* hasil eksperimen menggunakan algoritma *naive bayes* akurasinya sebesar 86.71%, sedangkan pengujian *cross validation* menggunakan algoritma genetika menghasilkan tingkat akurasi 90.27% *Root* membuktikan prediksi *data time series* metode *naive bayes* dan genetika menghasilkan akurasi sebesar 90,27%, Sehingga dapat di simpulkan bahwa menggunakan algoritma *naive bayes* dan genetika dapat mengoptimalkan dalam memprediksi keputusan klien Telemarketing tepat dalam penawaran deposito.

Penelitian [11] membahas tentang prediksi status kelulusan mahasiswa. Penulis menggunakan algoritma CART dalam memprediksi status kelulusan mahasiswa. Untuk itu perlu diketahui secara dini pola dari mahasiswa yang cenderung tidak lulus tepat waktu. Pada penelitian ini diusulkan algoritma CART untuk memprediksi kelulusan mahasiswa. Algoritma ini akan mengolah *dataset* profil mahasiswa yang berjumlah 238 data. Dataset ini memiliki label status kelulusan. Nilai dari label bernilai kategorikal, yaitu tepat dan terlambat. Adapun fitur atau atribut yang digunakan, yaitu jenis kelamin, asal sekolah, nilai uan, nilai sekolah, dan umur. Hasil dari algoritma CART ini berupa model pohon keputusan dan hasil eksperimen, akurasi dari model algoritma ini sebesar 64.37%.

Penelitian [12] melakukan penelitian di SMK TIK dan membahas tentang klasifikasi sumber belajar mata pelajaran produktif dari media internet, penelitian tersebut melakukan klasifikasi berdasarkan kriteria ciri esensial setiap mata pelajaran produktif di SMK TIK, pengelompokkan dengan algoritma *naive bayes* yang digunakan untuk dokumen teks dan menganggap semua atribut tidak berhubungan. Tujuan penelitian ini adalah

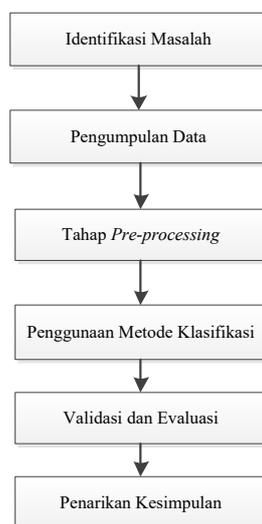
untuk mendeskripsikan hasil klasifikasi dan evaluasi kualitas klasifikasi sumber belajar berbasis teks dengan menggunakan algoritma *naive bayes* dan pengujian menghasilkan nilai akurasi tertinggi 81,48%, sedangkan nilai akurasi terendah sebesar 79,63%.

Dari kajian penelitian yang relevan, sehingga penulis menarik kesimpulan bahwa dari keterbaruan penelitian ini yaitu pada penggunaan algoritma CART. Dalam penelitian ini menggunakan algoritma CART dan algoritma naive bayes dalam klasifikasi nasabah *bank telemarketing*. CART merupakan metodologi statistik non-parametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk variabel respon kategorik maupun kontinu. Metode ini merupakan metode yang biasa diterapkan suatu himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan biner [13]. Kelebihan dari algoritma CART ialah metode yang bersifat non-parametrik/cocok yang digunakan untuk data yang berjenis *numeric*. CART tidak memerlukan variabel yang akan dipilih terlebih dahulu dan CART menghasilkan *invariant* untuk transformasi variabel yang independen. Pada klasifikasi algoritma CART, sebuah *record* akan diklasifikasikan ke dalam salah satu dari sekian klasifikasi yang tersedia pada variabel tujuan berdasarkan nilai-nilai variabel prediktornya [14].

Sedangkan pada Algoritma *naive bayes* merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan bayes dengan asumsi ketidaktergantungan yang kuat atau bersifat bebas (*independence*). Selain itu naive bayes juga dapat menganalisa variabel – variabel yang paling mempengaruhinya dalam bentuk peluang [15]. Kelebihan algoritma naive bayes antara lain algoritma naive bayes mudah digunakan untuk data *machine learning*, algoritma naive bayes hanya membutuhkan satu kali scan data training, dan digunakan untuk penanganan nilai atribut yang hilang dan data kontinu [16]. Penelitian ini bertujuan untuk menguji keberhasilan *bank telemarketing* dengan cara melakukan klasifikasi keputusan nasabah dengan menerapkan *data mining* sehingga luarannya mendapatkan sebuah informasi dari pola yang ada dan mengekstraksi informasi, algoritma klasifikasi yang digunakan yaitu *Classification and Regression Trees* dan algoritma naive bayes dengan metode algoritma sebagai yang menghasilkan klasifikasi nasabah *bank telemarketing*.

2. METODE PENELITIAN

Tahapan penelitian dilakukan guna mempermudah dalam melakukan penelitian. Desain penelitian perbandingan metode algoritma klasifikasi *Classification And Regression Trees* [14] dan algoritma *naive bayes* [16] untuk klasifikasi *bank telemarketing* dilihat pada gambar 1.



Gambar 1. Tahap Penelitian

1. Identifikasi Masalah.

Pada tahapan ini dilakukan untuk mengetahui permasalahan serta metode yang sesuai sehingga dapat ditentukan poin-poin untuk target nasabah *bank telemarketing*.

2. Pengumpulan Data.

Dalam tahapan ini pengumpulan data menggunakan data sekunder yang digunakan, diambil dari *database UCI Repository*. Dataset terdiri dari 4521 *record*, memiliki 17 *attribute* (16 *attribute* dan 1 *attribute target*),

yang akan digunakan sebagai *output* data *bank telemarketing* [17]. Kedua *output* pada *class attribute* ini akan digunakan sebagai *output* data nasabah *bank telemarketing*. Berikut ini merupakan penjelasan pada atribut *bank telemarketing*, dapat dilihat pada Tabel 1 dan Tabel 2.

Tabel 1. *Attribute Dataset* Nasabah *Bank Telemarketing*

Nama	Informasi
<i>Attribute 1</i>	Umur nasabah.
<i>Attribute 2</i>	Pekerjaan nasabah.
<i>Attribute 3</i>	Status perkawinan nasabah.
<i>Attribute 4</i>	Status pendidikan nasabah.
<i>Attribute 5</i>	Jenis pinjaman nasabah.
<i>Attribute 6</i>	Saldo tahunan rata-rata, dalam kurs mata uang Euro.
<i>Attribute 7</i>	Rekam jejak nasabah, khususnya mempunyai pinjaman perusahaan
<i>Attribute 8</i>	Rekam jejak pinjaman pribadi nasabah.
<i>Attribute 9</i>	Jejak kontak nasabah tipe komunikasi (seluler, tidak diketahui dan telepon).
<i>Attribute 10</i>	Hari kontak terakhir dengan nasabah.
<i>Attribute 11</i>	Bulan terakhir kontak dengan nasabah.
<i>Attribute 12</i>	Durasi kontak terakhir dengan nasabah (dalam detik).
<i>Attribute 13</i>	Jumlah hari selama kampanye yang dilakukan oleh nasabah.
<i>Attribute 14</i>	Jumlah hari yang dilewati setelah nasabah telah dihubungi.
<i>Attribute 15</i>	Jumlah kontak yang telah dilakukan sebelum kampanye ini dan untuk nasabah
<i>Attribute 16</i>	Hasil dari kampanye pemasaran sebelumnya
<i>Class_Target_Attribute 17</i>	Label kelas. Yes = Ya jika sudah berlangganan deposito berjangka. Label kelas. No = Tidak berlangganan deposito berjangka.

Tabel 2. Jumlah *Dataset* Berdasarkan *Class*

Jenis klasifikasi	Jumlah <i>record dataset</i>
<i>Yes</i>	521
<i>No</i>	4000
Jumlah	4521

3. Tahap *pre-processing*

Pada tahapan ini dilakukan suatu proses seleksi data dengan tujuan untuk mendapatkan data yang bersih dan siap digunakan sebagai bahan penelitian. Tahap *pre-processing* data meliputi identifikasi dan pemilihan atribut (*attribute identification and selection*), penanganan nilai atribut yang hilang atau atribut yang tidak lengkap (*handling missing values*) dan proses diskretisasi nilai.

4. Penggunaan metode klasifikasi

Metode yang peneliti gunakan yaitu algoritma CART dan naive bayes, dari hasil *confusion matrix* dapat dihitung dari nilai *precision*, *recall*, dan *F-Measure*. Berikut merupakan rincian *confusion matrix* dapat dilihat pada Tabel 3 sebagai berikut [16]:

Tabel 3. *Confusion Matrix*

<i>Correct classification</i>	<i>Classification as</i>	
	Yes	No
Yes	<i>True Positives</i>	<i>False Negatives</i>
No	<i>False Positives</i>	<i>True Negatives</i>

Adapun rumus perhitungan dari sebuah *precision*, *recall*, dan *F-Measure* dari sebuah *accuracy* [16], dapat dilihat pada persamaan dibawah ini :

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F-Measure = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

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

5. Validasi dan Evaluasi

Dalam tahap ini dilakukan validasi dan evaluasi serta pengukuran keakuratan hasil yang dicapai oleh model menggunakan teknik yang terdapat dalam aplikasi weka yaitu *confusion matrix* dan *10-fold cross validation*.

6. Penarikan Kesimpulan

Pada tahapan ini yaitu menyimpulkan menyimpulkan hasil yang diperoleh dari penelitian dengan menggunakan algoritma CART dan algoritma naive bayes yang memberikan hasil akurasi terbaik untuk mengklasifikasi nasabah *bank telemarketing* berdasarkan nilai *precision*, *recall*, dan *F-Measure* dari tiap masing-masing algoritma dengan tingkat klasifikasi [18] sebagai berikut:

1. *Excellent classification* = 0.90 - 1.00
2. *Good classification* = 0.80 - 0.90
3. *Fair classification* = 0.70 - 0.80
4. *Poor classification* = 0.60 – 0.70
5. *Failure* = 0.50 – 0.60

3. HASIL DAN ANALISIS

3.1. Identifikasi Masalah

Dalam tahapan ini penulis melakukan beberapa studi pendahuluan dengan mempelajari literatur-literatur yang berkaitan dengan penelitian penulis yaitu untuk mengklasifikasi dataset nasabah *bank telemarketing* serta pemilihan algoritma yang sesuai untuk penelitian penulis ini. Dari hasil tersebut penulis melakukan penelitian dengan menggunakan algoritma CART dan algoritma naive bayes untuk menganalisa *dataset*. Tahapan algoritma CART pada metode klasifikasi, langkah-langkah pada algoritma CART sebagai berikut [14]

1. Susunlah calon cabang (*candidate split*). Penyusunan ini dilakukan terhadap seluruh variabel prediktor atau prediksi secara lengkap (*exhaustive*). Daftar yang berisi calon cabang disebut calon cabang mutakhir.
2. Berikan penilaian keseluruhan calon cabang mutakhir dengan menghitung nilai besaran \emptyset (s|t).
3. Terakhir, menentukan calon manakah yang akan benar – benar dijadikan cabang dengan memilih calon cabang yang dimiliki nilai kebaikan \emptyset (s|t), terbesar. Setelah itu, gambarkanlah percabangan. Jika tidak ada lagi noktah keputusan, pelaksanaan algoritma CART akan dihentikan.

Sedangkan pada algoritma naive bayes, dalam penyelesaian klasifikasi algoritma naive bayes dengan cara mengkombinasikan suatu peluang dengan syarat kemungkinan terjadi sebuah formula yang dapat digunakan untuk menghitung peluang dari tiap kemungkinan yang telah terjadi.

3.2. Pengumpulan Data

Dalam penelitian penulis ini, data yang akan digunakan dalam penelitian yaitu dengan mengambil dari *repository* diambil dari *database UCI repository*. *Dataset* terdiri dari 4521 *record*, memiliki 17 *attribute* (16 *attribute* dan 1 *attribute* target), yang akan digunakan sebagai *output* data *bank telemarketing*. Pada tahap ini dibutuhkan waktu yang relatif lama untuk melakukan eksperimen agar mendapatkan hasil yang valid. Data tersebut harus diolah terlebih dahulu yaitu melalui tahap *pre-processing*, dimana tahapan ini bertujuan untuk menyesuaikan atribut – atribut yang akan digunakan dalam mengolah *dataset* tersebut karena setiap atribut pada *dataset* ini akan memberikan hasil yang berbeda – beda pada hasil akurasi dalam mengklasifikasi data *bank telemarketing*, dan hal ini berkaitan ada atau tidaknya kelengkapan nilai pada setiap atribut. Untuk bahan penelitian dan terlebih dahulu disesuaikan dengan tipe data yang dapat dibaca oleh aplikasi weka.

3.3. Tahap *Pre-processing*

Tahapan ini, dilakukan identifikasi dan penyesuaian atribut, serta penyeleksian dari *dataset* nasabah *bank telemarketing* agar data yang diperoleh ialah data yang benar-benar siap untuk digunakan. Tabel 4., yang merupakan *dataset* nasabah *bank telemarketing* yang telah dilakukan penyesuaian *attribute* untuk aplikasi weka.

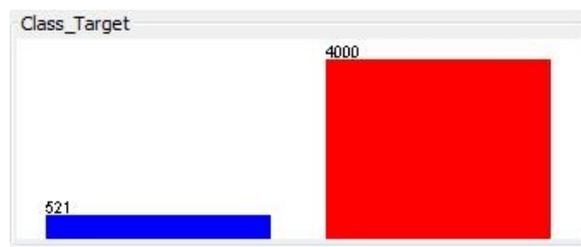
Tabel 4. Data *Pre-Processing*

Data Asli	Data Hasil <i>Pre-processing</i>	Keterangan
35	35	Mendeskripsikan umur nasabah
<i>Management</i>	<i>Management</i>	Pekerjaan nasabah
<i>Single</i>	<i>Single</i>	Status perkawinan nasabah
<i>Tertiary</i>	<i>Tertiary</i>	Status pendidikan nasabah
<i>No</i>	<i>No</i>	Jenis pinjaman nasabah
1350	1350	Saldo tahunan rata-rata, dalam kurs mata uang Euro
<i>Yes</i>	<i>Yes</i>	Mendeskripsikan rekam jejak khususnya mempunyai pinjaman perumahan
<i>No</i>	<i>No</i>	Mendeskripsikan rekam jejak pinjaman pribadi nasabah
<i>Cellular</i>	<i>Cellular</i>	Jenis kontak nasabah, tipe komunikasi (seluler, telepon, tidak dikenal)
16	16	Mendeskripsikan hari koontak terakhir dengan nasabah
Apr	Apr	Mendeskripsikan bulan kontak terakhir dengan nasabah
185	185	Mendeskripsikan durasi kontak terakhir dengan nasabah (dalam detik)
1	1	Mendeskripsikan jumlah kontak selama kampanye yang dilakukan oleh nasabah
130	130	Mendeskripsikan jumlah hari yang dilewati setelah nasabah terakhir dihubungi
1	1	Mendeskripsikan jumlah hari yang dilakukan sebelum kampanye ini dan untuk nasabah
<i>Failure</i>	<i>Failure</i>	Hasil dari kampanye pemasaran sebelumnya
<i>No</i>	<i>No</i>	Label kelas. <i>Yes</i> = Ya jika sudah berlangganan deposito berjangka, Label kelas. <i>No</i> = Tidak berlangganan deposito berjangka

Dari hasil analisis pada Tabel 4 tentang hasil data *pre-processing dataset bank telemarketing* yang telah dilakukan, bahwa semua atribut nilainya lengkap atau tidak *missing value*.

3.4. Penggunaan Metode Klasifikasi

Setelah melalui tahapan *pre-pocessing* selanjutnya *dataset* mulai diolah dengan menggunakan aplikasi Weka. Tahapan ini bertujuan untuk menghasilkan *confusion matrix* berdasarkan metode evaluasi *10-fold cross validation*, dimana *dataset* tersebut dibagi menjadi 10 *subsets* (9 *subsets* sebagai *training sets* dan 1 *subsets* sebagai *testing set*) dengan jumlah 10 kali iterasi. Adapun *classifier* yang digunakan untuk uji coba pada dataset adalah SimpleCart algoritma CART dan Naive Bayes (algoritma naive bayes). Gambar 3 menunjukkan visualisasi data, yang dijabarkan sesuai dengan atribut setelah melewati tahapan *pre-processing*. Selanjutnya data-data tersebut diklasifikasi menggunakan algoritma CART (SimpleCart dalam penggunaan pada aplikasi Weka) dan algoritma naive bayes (NaiveBayes dalam penggunaan pada aplikasi weka) yang menghasilkan pola klasifikasi. Berikut gambar 2 tentang visualisasi data keseluruhan dari *class_target*, sebagai berikut :



Gambar 2. Visualisasi data keseluruhan

Keterangan dari Gambar 2 diatas sebagai berikut:

1. Gambar grafik dengan warna biru yaitu (*Yes*) berlangganan deposito berjangka.
2. Gambar grafik dengan warna merah yaitu (*No*) tidak berlangganan deposito berjangka.

Pada algoritma CART menghasilkan nilai akurasi sebesar 89.51%. Nilai akurasi tersebut didapatkan dari hasil perhitungan dari *precision*, *recall* dan *F-Measure*. Proses perhitungannya dapat dilihat pada tabel 6. *Confusion* Kelas Yes

Tabel 6. *Confusion* Kelas Yes

178 (<i>True Positive</i>)	343 (<i>False Negative</i>)
131 (<i>False Positive</i>)	3869 (<i>True Negative</i>)

Pada tabel 6. *Confusion* kelas yes, merupakan penjabaran dari *confusion matrix* yang dapat dilihat pada tabel 6. *Confusion* kelas yes.

Kelas Yes

$$Precision = \frac{TP}{TP+FP} = \frac{178}{178+131} = 0.576$$

$$Recall = \frac{TP}{TP+FN} = \frac{178}{178+343} = 0.342$$

$$F-Measure = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times 0.576 \times 0.342}{0.576 + 0.342} = \frac{2 \times 0.576 \times 0.342}{0.576 + 0.342} = 0.429$$

Tabel 7. *Confusion* Kelas No

3869 (<i>True Positive</i>)	131 (<i>False Negative</i>)
343 (<i>False Positive</i>)	178 (<i>True Negative</i>)

Kelas No

$$Precision = \frac{TP}{TP+FP} = \frac{3869}{3869+343} = 0.919$$

$$Recall = \frac{TP}{TP+FN} = \frac{3869}{3869+131} = 0.967$$

$$F-Measure = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times 0.919 \times 0.967}{0.919 + 0.967} = 0.942$$

Dari perhitungan perhitungan diatas antara lain *precision*, *recall* dan *F-Measure* yang dihasilkan dari kelas Yes dan kelas No dapat dihitung nilai rata-rata dari kelas-kelas yang ada (*Weighted Avg*) dengan terlebih dahulu menjumlahkan nilai A = (178 + 343) = 521 dan nilai B = (131 + 3869) = 4000. Rumus untuk menghitung *Weighted Avg*, sebagai berikut:

$$Weighted\ Avg\ (precision) = \frac{0,576 \times A + 0,919 \times B}{4521} = \frac{0,576 \times 521 + 0,919 \times 4000}{4521} = 0.879$$

$$Weighted\ Avg\ (recall) = \frac{0,342 \times A + 0,967 \times B}{4521} = \frac{0,342 \times 521 + 0,967 \times 4000}{4521} = 0.895$$

$$Weighted\ Avg\ (F-Measure) = \frac{0,429 \times A + 0,942 \times B}{4521} = \frac{0,429 \times 521 + 0,942 \times 4000}{4521} = 0.883$$

Hasil diatas merupakan perhitungan nilai akurasi berdasarkan *confusion matrix* hasil algoritma CART yang disajikan pada tabel 8. Nilai akurasi berdasarkan *confusion matrix*, sebagai berikut:

Tabel 8. Nilai Akurasi Berdasarkan *Confusion Matrix* Algoritma CART

Class	Precision	Recall	F-Measure
Yes	0.576	0.342	0.429
No	0.919	0.967	0.942
Weighted Avg	0.879	0.895	0.883

Pengklasifikasian untuk menguji *dataset bank telemarketing* dengan algoritma CART menggunakan aplikasi weka 3.6 dan waktu yang dibutuhkan adalah 4.75 second.

Sedangkan pada algoritma naive bayes menghasilkan nilai akurasi sebesar 86.88%. Nilai akurasi tersebut didapatkan dari hasil perhitungan dari *precision*, *recall* dan *F-Measure*. Proses perhitungannya dapat dilihat pada tabel 9. *Confusion* Kelas Yes.

Tabel 9. *Confusion* Kelas Yes

265 (<i>True Positive</i>)	256 (<i>False Negative</i>)
337 (<i>False Positive</i>)	3663 (<i>True Negative</i>)

Pada tabel 9. *Confusion* kelas yes, merupakan penjabaran dari *confusion matrix* yang dilihat pada tabel 9. *Confusion* kelas yes.

Kelas yes

$$Precision = \frac{TP}{TP+FP} = \frac{265}{265+337} = 0.44$$

$$Recall = \frac{TP}{TP+FN} = \frac{265}{265+256} = 0.509$$

$$F-Measure = \frac{2x \text{ precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{2x0,44 \times 0,509}{0,44 + 0,509} = 0.472$$

Tabel 10. *Confusion* Kelas No

3663 (<i>True Positive</i>)	337 (<i>False Negative</i>)
256 (<i>False Positive</i>)	265 (<i>True Negative</i>)

Kelas No

$$Precision = \frac{TP}{TP+FP} = \frac{3663}{3663+256} = 0.935$$

$$Recall = \frac{TP}{TP+FN} = \frac{3663}{3663+337} = 0.916$$

$$F-Measure = \frac{2x \text{ precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{2x0,935 \times 0,916}{0,935 + 0,916} = 0.925$$

Dari hasil perhitungan diatas antara lain *precision*, *recall* dan *F-Measure* yang dihasilkan dari kelas *yes* dan kelas *no* dapat dihitung nilai rata-rata dari kelas-kelas yang ada (*Weighted Avg*) dengan terlebih dahulu menjumlahkan nilai A = (265+ 256) = 521 dan nilai B = (337 + 3663) = 4000. Rumus untuk menghitung *Weighted Avg*, sebagai berikut :

$$Weighted \text{ Avg (precision)} = \frac{0,44x A + 0,935 \times B}{4521} = \frac{0,44 x 521 + 0,935 \times 4000}{4521} = 0.878$$

$$Weighted \text{ Avg (recall)} = \frac{0,509x A + 0,916 \times B}{4521} = \frac{0,509 x 521 + 0,916 \times 4000}{4521} = 0.869$$

$$Weighted \text{ Avg (F-Measure)} = \frac{0,472x A + 0,925 \times B}{4521} = \frac{0,472 x 521 + 0,925 \times 4000}{4521} = 0.873$$

Hasil diatas merupakan perhitungan nilai akurasi berdasarkan *confusion matrix* hasil algoritma naive bayes yang disajikan pada Tabel 11.

Tabel 11. Nilai Akurasi Berdasarkan *Confusion Matrix* Algoritma Naive Bayes

Class	Precision	Recall	F-Measure
Yes	0.44	0.509	0.472
No	0.935	0.916	0.925
Weighted Avg	0.878	0.869	0.873

Pengklasifikasian untuk menguji *dataset bank telemarketing* dengan algoritma naive bayes menggunakan aplikasi Weka 3.6 dan waktu yang dibutuhkan adalah 0.14 second. Dari hasil perhitungan pada algoritma CART dan algoritma *naive bayes*. Berikut perbedaan hasil akurasi dari kedua algoritma tersebut, yang disajikan pada tabel 12. Perbandingan hasil akurasi CART dan *naive bayes*.

Tabel 12. Perbandingan Hasil Akurasi Algoritma CART dan Algoritma *Naive Bayes*

Algoritma	Hasil akurasi	Precision	Recall	F-Measure	Waktu
CART	89.51%	0.87	0.89	0.88	4.75 second
Naive bayes	86.88%	0.87	0.86	0.78	0.14 second

Dari penjabaran Tabel 12 bahwa nilai akurasi algoritma CART lebih baik dibandingkan dengan algoritma naive bayes pada algoritma CART dalam klasifikasi nasabah *bank telemarketing*, karena algoritma cocok dengan *dataset bank telemarketing* terdiri dari *dataset* yang terdiri jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran [13].

3.5. Validasi dan Evaluasi

Untuk mengukur tingkat akurasi dari algoritma klasifikasi yang digunakan yaitu menggunakan *confusion matrix* yang disajikan pada tabel 13. dan tabel 14. Tabel 13 merupakan tabel yang dihasilkan oleh *confusion matrix* dari pengujian *dataset bank telemarketing* menggunakan algoritma CART dengan metode *10-fold cross validation*.

Tabel 13. *Confusion Matrix* Algoritma CART

	Yes	No
Yes	178	343
No	131	3869
4521	521	4000

Dari Tabel 13 terlihat bahwa jumlah data hasil bentukan *rule* yang *Yes* berlangganan deposito berjangka *bank telemarketing* yang sama dengan data testing yang juga *Yes* berlangganan deposito berjangka *bank telemarketing* sebanyak 178. Kemudian, jumlah data hasil bentukan *rule* yang *No* berlangganan deposito berjangka *bank telemarketing* dengan data testing yang yang *Yes* berlangganan deposito berjangka *bank telemarketing* sebanyak 343. Selanjutnya, jumlah data hasil bentukan *rule* yang *Yes* berlangganan deposito berjangka *bank telemarketing* dan data testing yang *No* berlangganan deposito berjangka *bank telemarketing* sebanyak 131. Terakhir, jumlah data hasil bentukan *rule* yang *No* berlangganan deposito berjangka *bank telemarketing* yang sama dengan data testing yang juga *No* berlangganan deposito berjangka *bank telemarketing* sebanyak 3869. Sedangkan pada Tabel 14 merupakan tabel yang dihasilkan oleh *confusion matrix* dari pengujian *dataset bank telemarketing* menggunakan algoritma naive bayes dengan metode *10-fold cross validation*.

Tabel 14. *Confusion Matrix* Algoritma *Naive Bayes*

	Yes	No
Yes	265	256
No	337	3663
4521	602	3919

Dari Tabel 14 terlihat bahwa jumlah data hasil bentukan *rule* yang *Yes* berlangganan deposito berjangka *bank telemarketing* yang sama dengan data testing yang juga *Yes* berlangganan deposito berjangka *bank telemarketing* sebanyak 265. Kemudian, jumlah data hasil bentukan *rule* yang *No* berlangganan deposito berjangka *bank telemarketing* dengan data testing yang yang *Yes* berlangganan deposito berjangka *bank telemarketing* sebanyak 256. Selanjutnya, jumlah data hasil bentukan *rule* yang *Yes* berlangganan deposito berjangka *bank telemarketing* dan data testing yang *No* berlangganan deposito berjangka *bank telemarketing* sebanyak 337. Terakhir, jumlah data hasil bentukan *rule* yang *No* berlangganan deposito berjangka *bank telemarketing* yang sama dengan data testing yang juga *No* berlangganan deposito berjangka *bank telemarketing* sebanyak 3663.

3.6. Penarikan Kesimpulan

Dari hasil perhitungan yang telah dilakukan pada kedua algoritma, didapatkan hasil akurasi dari masing – masing algoritma yaitu 89.51% dengan nilai *precision* 87%, *Recall* 89% dan *F-Measure* 88% pada algoritma CART dan 86.88%, pada algoritma naive bayes dengan nilai *precision* 87%, *Recall* 86%, dan *F-Measure* 87% disajikan pada Tabel 15.

Tabel 15. Perbandingan Algoritma CART dan *Naive Bayes*

Algoritma	Hasil akurasi	Precision	Recall	F-Measure	Waktu
CART	89.51%	0.87	0.89	0.88	4.75 second
Naive bayes	86.88%	0.87	0.86	0.78	0.14 second

Pada Tabel 15 merupakan hasil dari perhitungan kedua algoritma yaitu algoritma CART dan algoritma naive bayes, sedangkan pada penelitian sebelumnya yang relevan, yaitu penelitian oleh [9] mendapatkan hasil

akurasi algoritma naive bayes sebesar 86.71%, penelitian oleh [10] mendapatkan hasil akurasi algoritma naive bayes sebesar 86.84%, dan penelitian yang dilakukan oleh [11] mendapatkan hasil akurasi algoritma CART sebesar 64.37%.

4. KESIMPULAN

Dari hasil penelitian ini, peneliti mengambil kesimpulan bahwa hasil klasifikasi algoritma CART memiliki nilai akurasi sebesar 89.51% lebih baik daripada algoritma naive bayes yaitu dengan nilai akurasi sebesar 86.88%, pada dataset *bank telemarketing*, dengan nilai selisih 2.63%. Hal ini menjadi rujukan pihak *bank telemarketing* dalam proses penentuan keputusan nasabah *bank*, khususnya dalam klasifikasi keputusan nasabah *bank telemarketing* sehingga mendapatkan luaran informasi yang sesuai target.

UCAPAN TERIMA KASIH

Pertama, peneliti ucapkan terima kasih kepada Tuhan Yang Maha Esa, karena dengan Rahmat dan Karunia-Nya, penelitian dan jurnal ini dapat terselesaikan, sehingga peneliti ini dapat menyumbangkan berupa ilmu pengetahuan dari hasil penelitian. Kemudian terima kasih juga untuk Universitas Amikom Purwokerto yang telah membantu peneliti baik materi dan non materi., sehingga penelitian ini terselesaikan dengan baik.

REFERENSI

- [1] P. R. Patil, P. Revankar, and P. Joshi, "The Application of Data Mining for Direct Marketing," in *Second International Conference on Emerging Trends in Engineering & Technology*, 2009, pp. 312–316.
- [2] H. A. Elsalamony and A. M. Elsayad, "Bank Direct Marketing Based on Neural Network," *International Journal of Engineering and Advanced Technology*, vol. 2, no. 6, pp. 392–400, 2013.
- [3] D. T. Larose, *Data Mining Methods and Models*. New Jersey, Canada: Simultaneously, 2011.
- [4] R. Vaidehi, "Predictive Modelling to Improve Success Rate of Bank Direct Marketing Campaign," *International Journal of Management and Business Study*, vol. 6, no. 1, pp. 22–24, 2016.
- [5] Y. P. Anggodo, W. Cahyaningrum, A. N. Fauziyah, I. L. Khoiriyah, O. Kartikasari, and I. Cholissodin, "Hybrid K-Means dan Particle Swarm Optimization untuk Clustering Nasabah Kredit," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 2, pp. 104–110, 2017.
- [6] V. Vijayakumar and R. Nedunchezian, "A study on video data mining," *International Journal of Multimedia Information Retrieval*, vol. 1, no. 3, pp. 153–172, Oct. 2012.
- [7] G. J. Lobo, "Accounting research in banking – A review," *China Journal of Accounting Research*, vol. 10, no. 1, pp. 1–7, 2017.
- [8] Q. Chi and L. Wenjing, "Economic Policy Uncertainty, Credit Risks and Banks' Lending Decisions: Evidence from Chinese Commercial Banks," *China Journal of Accounting Research*, vol. 10, no. 1, pp. 33–50, 2017.
- [9] A. A. Aqham and K. D. Hartomo, "Data Mining untuk Nasabah Bank Telemarketing Menggunakan Kombinasi Algoritma Naive Bayes Dan Algoritma Genetik," *Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 4, no. 1, pp. 48–56, 2019.
- [10] W. Darmawan, "Komparasi Metode Data Mining dalam Memprediksi Nasabah Bank yang akan Memilih Tabungan Deposito Menggunakan Algoritma Klasifikasi," *Jurnal IC-Tech :Majalah Ilmiah*, vol. 13, no. 1, pp. 49–55, 2018.
- [11] M. F. Amin, "Penerapan Algoritma Cart Untuk Memprediksi Status Kelulusan Mahasiswa," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 5, no. 3, pp. 1215–1222, 2016.
- [12] A. D. Herlambang and S. H. Wijoyo, "Algoritma Naive Bayes untuk Klasifikasi Sumber Belajar Berbasis Teks Pada Mata Pelajaran Produktif di SMK Rumpun Teknologi Informasi Dan Komunikasi," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 4, pp. 431–436, 2019.
- [13] Nuriyah, "Perbandingan Metode Chi-square Automatic Interaction Detection (CHAID) dan Classification and Regression Tree (CART) dalam Menentukan Klasifikasi Alumni UIN Sunan Kalijaga Berdasarkan Masa Studi," Universitas Islam Negeri Sunan Kalijaga, 2013.
- [14] R. Timofeev, *Classification and Regression Trees (CART) Theory and Applications*. Berlin: Humboldt University, 2004.
- [15] E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset, 2012.
- [16] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed. San Fransisco: Elsevier Science & Technology, 2006.
- [17] S. Moro, P. Cortez, and P. Rita, "A Data-driven Approach to Predict the Success of Bank Telemarketing," *Decision Support Systems*, vol. 62, pp. 22–31, Jun. 2014.
- [18] F. Gorunescu, *Data mining Concepts, Models and Techniques*. Verlen Berlin: Springer, 2011.