



Perbandingan Algoritma Naïve Bayes dan C4.5 dalam Klasifikasi Pencapaian Target Penjualan Tabungan Bank XYZ

Mohammad Farras Daffauzan¹, Roy Mubarak^{*2}

^{1,2}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Jakarta, Indonesia
Email: ¹farrasdaffauzan@gmail.com; ^{*2}roy.mubarak@mercubuana.ac.id;

Daffauzan, M. F., & Mubarak, R. (2025). Perbandingan Algoritma Naïve Bayes dan C4.5 dalam Klasifikasi Pencapaian Target Penjualan Tabungan Bank XYZ. *Journal Cerita: Creative Education of Research in Information Technology and Artificial Informatics*, 11(1), 93-102

DOI: <https://doi.org/10.33050/cerita.v11i1.3337>

ABSTRAK

Penelitian ini mengkaji klasifikasi pencapaian target penjualan produk tabungan pada Kantor Cabang Pembantu (KCP) Bank XYZ menggunakan algoritma Naïve Bayes dan C4.5. Dengan perhatian terhadap pertumbuhan pesat dalam industri perbankan, peningkatan target penjualan produk menjadi tantangan utama yang dihadapi Bank XYZ. Penggunaan data mining, khususnya algoritma klasifikasi, menjadi strategi untuk meningkatkan efektivitas penjualan produk tabungan. Data historis penjualan produk tabungan dari Bank XYZ digunakan dalam studi kasus ini, dengan data yang telah dipreproses dan divisualisasikan untuk analisis yang lebih baik. Hasil evaluasi menunjukkan bahwa algoritma C4.5 memberikan akurasi yang lebih tinggi (95,1%) dibandingkan dengan Naïve Bayes (78,32%). Selain itu, evaluasi menggunakan metrik precision, recall, dan F1-score menunjukkan kinerja yang lebih baik dari algoritma C4.5 dalam mengklasifikasikan kedua kelas (0 dan 1) dengan precision, recall, dan F1-score yang lebih tinggi. Dengan demikian, penggunaan algoritma C4.5 lebih disarankan untuk mengklasifikasikan pencapaian target penjualan produk tabungan pada Kantor Cabang Pembantu Bank XYZ, karena memberikan hasil yang lebih baik dibandingkan dengan Naïve Bayes. Penelitian ini memberikan wawasan yang berharga bagi Bank XYZ dalam merancang strategi penjualan yang lebih efektif untuk meningkatkan pencapaian target penjualan produk tabungannya.

Kata kunci: data mining, klasifikasi, algoritma C4.5, algoritma naive bayes, perbandingan algoritme

ABSTRACT

This research examines the classification of savings product sales target achievement at Bank XYZ Sub-Branch Offices (KCP) using the Naïve Bayes and C4.5 algorithms. With attention to the rapid growth in the banking industry, increasing product sales targets is the main challenge faced by Bank XYZ. The use of data mining, especially classification algorithms, is a strategy to increase the effectiveness of savings product sales. Historical data on sales of savings products from Bank XYZ is used in this case study, with the data preprocessed and visualized for better analysis. The evaluation results show that the C4.5 algorithm provides higher accuracy (95.1%) compared to Naïve Bayes (78.32%). In addition, evaluation using precision, recall and F1-score metrics shows better performance of the C4.5 algorithm in classifying both classes (0 and 1) with higher precision, recall and F1-score. Thus, the use of the C4.5 algorithm is more recommended for classifying the achievement of savings product sales targets at XYZ Bank Sub-Branch Offices, because it provides better results compared to Naïve Bayes. This research provides valuable insight for Bank XYZ in designing more effective sales strategies to increase the achievement of sales targets for its savings products.

Keywords: data mining, classification, C4.5 algorithm, naive Bayes algorithm, algorithm comparison

I. PENDAHULUAN

Perkembangan pada dunia perbankan semakin meningkat dengan banyak tantangan, selama revolusi industri perbankan 4.0 mengalami pertumbuhan yang menjadi penggerak utama negara. Sektor perbankan di Indonesia pada saat ini telah mengalami perkembangan yang pesat, baik munculnya bank – bank baru yang berasal dari bank pemerintah ataupun bank swasta yang membuat setiap bank saling bersaing untuk menjadi pilihan Masyarakat. Oleh karena itu, setiap bank berkompetisi untuk memberikan fasilitas yang baik, berkualitas, dan berorientasi pada kebutuhan nasabahnya. Salah satu upaya bank dalam menarik nasabah yaitu dengan memberikan produk yang berkualitas. Produk-produk yang ditawarkan oleh bank harus sesuai dengan kebutuhan dan keinginan nasabahnya.

Bank XYZ merupakan salah satu bank yang berperan sebagai bank yang diberi mandat untuk memperbaiki ekonomi rakyat dan berpartisipasi dalam pembangunan nasional. Bank XYZ menyediakan beberapa produk salah satunya yaitu Produk Tabungan yang berguna untuk memenuhi kebutuhan nasabah. Bank XYZ memberikan layanan terbaik yang disesuaikan dengan kebutuhan nasabah sejak kecil, remaja, dewasa, hingga pensiun.

Dalam menghadapi persaingan yang semakin ketat, pencapaian target penjualan produk menjadi salah satu tantangan utama yang dihadapi oleh Bank XYZ. Demi mempertahankan dan meningkatkan pangsa pasar, bank perlu memastikan bahwa produk-

produk yang ditawarkan tidak hanya berkualitas, tetapi juga berhasil menjangkau dan memenuhi kebutuhan nasabah secara efektif. Kualitas produk merupakan faktor penting yang mempengaruhi keputusan setiap customer dalam pembelian produk, jika kualitas produk semakin baik maka akan semakin meningkat juga minat customer (Ernawati, 2019). Oleh karena itu, penting bagi Bank XYZ untuk memiliki pemahaman yang mendalam tentang faktor-faktor yang memengaruhi pencapaian target penjualan produk di berbagai lokasi, termasuk Kantor Cabang Pembantu (KCP).

Dengan memanfaatkan perkembangan teknologi saat ini, penggunaan data mining menjadi salah satu cara dalam upaya untuk meningkatkan penjualan produk tabungan. Data mining adalah Teknik yang digunakan untuk menemukan informasi guna mengungkapkan hubungan antara data yang akan dikelompokkan dalam satu maupun lebih sehingga nilai dalam Kelompok secara substansi memiliki kemiripan satu sama lain sebagai pendekatan untuk mengambil pola dan data (Tarigan et al, 2022).

Data mining dapat mengidentifikasi sebuah basis data untuk mencari informasi berguna dari data yang banyak (Big Data). Data mining dapat menggambarkan beberapa kumpulan teknik yang memiliki tujuan untuk penentuan pola yang telah ada dan tidak diketahui pada akumulasi sebuah data, maka kemungkinan besar akan memperoleh informasi baru (Zai, 2022).

Data mining adalah sebuah metode yang menerapkan serangkaian mekanisme untuk memperoleh pola, hubungan, dan nilai dari data

banyak (big data) yang bertujuan untuk menggali wawasan yang bermanfaat. Dalam proses yang dilakukan melibatkan penggunaan ilmu statistic dan matematika untuk menyederhanakan sebuah data dan menghasilkan informasi yang dapat dipahami (Purwanto et al, 2020). Hasil penggunaan data mining tersebut dapat digunakan untuk pengambilan sebuah keputusan dalam Analisa data.

Data penjualan produk Tabungan pada KCP Bank XYZ dilakukan proses klasifikasi yang merupakan salah satu metode dalam data mining. Klasifikasi adalah cara dalam melakukan sebuah pengelompokan data berdasarkan Tingkat kesamaan dan karakteristik dari data ke dalam sebuah kelas (Putro et al, 2020). Klasifikasi dapat mengidentifikasi pola – pola tertentu yang memiliki korelasi dengan keberhasilan mencapai target. Klasifikasi merupakan sebuah proses dalam menemukan sebuah model atau membedakan antar kelas pada data. Klasifikasi melakukan sebuah analisis berdasarkan satu set data pelatihan dan digunakan untuk melakukan prediksi label kelas yang belum diketahui. Beberapa algoritma yang menggunakan metode klasifikasi yaitu Naïve Bayes, Support Vector Machine, AdaBoost, Decision Tree dan lainnya (Hidayati & Hermawan, 2021). Algoritma klasifikasi yang digunakan untuk membantu Bank XYZ dalam menganalisis pola data penjualan produk Tabungan yaitu algoritma Naïve Bayes dan C4.5.

Dalam memprediksi probabilitas atribut dapat menggunakan algoritma naïve bayes yang memiliki keunggulan mempunyai nilai performa yang baik (Siddik et al, 2020). Algoritma naïve bayes merupakan algoritma supervised learning dimana data akan dilatih dengan data input untuk melakukan klasifikasi dengan akurat (Linawati et al, 2020).

Algoritma C4.5 merupakan algoritma yang digunakan untuk memprediksi dengan membuat sebuah pohon Keputusan (Sihaan et al, 2022). Algoritma C4.5 merupakan salah satu algoritma yang digunakan dalam melakukan segmentasi maupun klasifikasi yang bersifat prediktif (Fikrul Ilmi Zer et al, 2019).

Penelitian yang dilakukan Susanti melakukan klasifikasi kelayakan pemberian kredit nasabah bank xyz menggunakan metode algoritma C4.5 dan Naïve Bayes. Pada model

yang dilakukan hasil akurasi yang didapatkan dari algoritma C4.5 dengan tiga kali pengujian yaitu 65,75%, 67,70%, dan 64,95%, sedangkan algoritma naïve bayes hasil akurasi yang didapatkan yaitu 64,72%, 66,67%, dan 63,40% (Susanti et al, 2022).

Penelitian yang dilakukan Rian Pratama melakukan perbandingan algoritma c4.5 dan naïve bayes dalam menentukan persediaan stok menggunakan tools RapidMiner. Pada model yang dilakukan hasil akurasi yang didapatkan dari algoritma C4.5 sebesar 96,80% sedangkan algoritma Naïve Bayes sebesar 91,20%. Dalam penelitian Rian Pratama, memprediksi item yang perlu ditambah atau dikurangi agar stok persediaan tetap stabil dan juga menghindari kerugian yang disebabkan oleh kadaluarsa (Pratama et al, 2022).

Pada penelitian Kohsasih melakukan analisis perbandingan algoritma c4.5 dan Naïve Bayes dalam memprediksi penyakit Cerebrovascular atau stroke. Pada model yang dilakukan hasil analisis akurasi algoritma C4.5 sebesar 95% sedangkan algoritma Naïve Bayes mendapatkan akurasi sebesar 91% (Kohsasih et al, 2022).

Penelitian yang dilakukan Hendra melakukan perbandingan algoritma C4.5 dan Naïve Bayes untuk klasifikasi mustahik. Hasil penelitian yang dilakukan bahwa penggunaan algoritma C4.5 akurasi yang didapatkan sebesar 75% sedangkan algoritma naïve bayes mendapatkan akurasi sebesar 50% (Bayu Suseno et al, 2020).

Astrid melakukan analisis kinerja algoritma C4.5 dan naïve bayes untuk memprediksi prestasi siswa sekolah menengah kejuruan. Hasil penelitian menunjukkan algoritma naïve bayes memiliki nilai akurasi sebesar 95,67% sedangkan algoritma C4.5 memiliki nilai akurasi sebesar 90,91% (Noviriandini & Nurajijah, 2019).

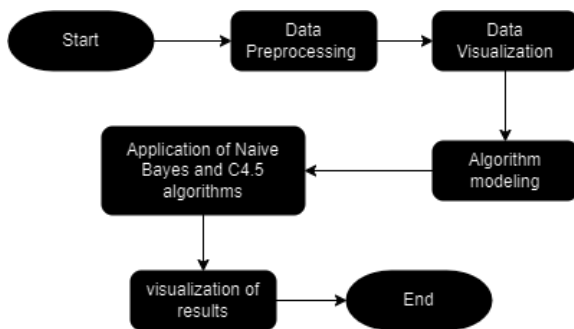
Dataset penjualan produk Tabungan yang telah berlangsung pada Bank XYZ penelitian ini memfokuskan pada klasifikasi pencapaian target penjualan produk Tabungan pada kantor cabang pembantu agar Bank XYZ dapat menentukan langkah – langkah yang tepat dalam merancang strategi penjualan untuk meningkatkan target penjualan produk Tabungan pada kantor cabang pembantunya. Penelitian ini juga bertujuan untuk membandingkan hasil dari algoritma Naïve

Bayes dan C4.5 guna mencari hasil akurasi yang terbaik.

Tabungan L	Int64
Jumlah Penjualan	Int64
Target Tercapi	Object

II. METODOLOGI PENELITIAN

Desain penelitian ini menerapkan pendekatan studi kasus yang melibatkan penggunaan data historis penjualan produk tabungan Bank XYZ. Data yang digunakan yaitu dari bulan Desember 2019 hingga Desember 2021. Adapun alur penelitian yang dilakukan Dapat dilihat pada gambar 1.



Gambar. 1 Alur Penelitian

A. Data Pre - processing

Penelitian ini dibagi menjadi tiga tahap Tahapan preprocessing merupakan tahapan yang dilakukan untuk pembersihan data, hal ini mengatasi masalah seperti data hilang maupun tidak valid. Sehingga pada pengolahan data dimana awal data mentah atau tidak tersusun diubah menjadi data yang siap digunakan untuk dilakukan analisis lebih lanjut. Adapun data yang telah dilakukan proses preprocessing memiliki 17 atribut.

TABEL I DATA PREPROCESSING

Atribut	Tipe Data
Bulan	Object
Tahun	Int64
Nama	Object
Tabungan A	Int64
Tabungan B	Int64
Tabungan C	Int64
Tabungan D	Int64
Tabungan E	Int64
Tabungan F	Int64
Tabungan G	Int64
Tabungan H	Int64
Tabungan I	Int64
Tabungan J	Int64
Tabungan K	Int64

B. Data Visualization

Tahapan Data Visualisasi atau disebut dengan Exploratory Data analysis (EDA) merupakan proses menganalisis dan menampilkan untuk mendapatkan pemahaman dari data yang lebih baik (Sahoo et al, 2019). Visualisasi Data adalah tahapan dari analisis data, proses yang dilakukan yaitu menyakikan sebuah visualisasi berbentuk table, grafik, maupun diagram seperti melihat visualisasi dari total penjualan tabungan dari tipe target tercapai.

C. Permodelan Algoritma

Tahapan permodelan algoritma merupakan tahapan yang dilakukan sebelum melakukan penerapan klasifikasi algoritma Naïve Bayes dan C4.5 dimana data yang dimiliki hanya terdapat nilai numerik supaya model algoritma dalam melakukan penghitungan dan prediksi menjadi lebih akurat. Selanjutnya memisahkan data untuk kategori fitur berdasarkan variabel yang akan digunakan untuk membuat sebuah prediksi dan kategori label berdasarkan variabel yang ingin dilakukan prediksi. Pada kategori fitur dilakukan pengskalaan nilai supaya nilai antar atribut tidak memiliki rentan yang cukup jauh lalu data dibagi menjadi dalam rasio 30:70 dan memiliki random state yang bernilai 101 guna untuk mereplikasi hasil secara konsisten dan memungkinkan perbandingan yang adil.

D. Klasifikasi Algoritma Naive Bayes

Algoritma Naïve bayes merupakan sebuah klasifikasi yang dilakukan berdasarkan probabilitas yang akan memprediksi di masa depan. Algoritma Naïve Bayes merupakan metode klasifikasi yang populer atau Teknik pembelajaran mesin yang digunakan untuk melakukan sebuah klasifikasi data. Naïve Bayes dalam pemilihan fitur sangat sensitive apabila terlalu banyak fitur maka akan terjadi tingkatan waktu perhitungan dan kurangi akurasi dalam klasifikasi (Shaizadi Meraj et al, 2019).

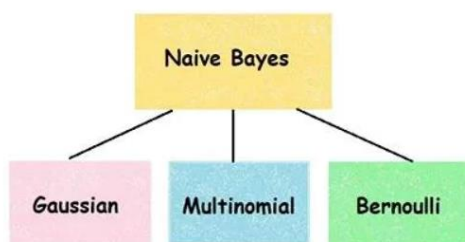
Naïve bayes merupakan metode klasifikasi kumpulan algoritma disusun berdasarkan dengan Teorema Bayes yang menggunakan metode probabilitas dan statistic dikemukakan oleh Thomas Bayes merupakan seorang ilmuwan yang

berasal dari inggris. Algoritma naive bayes melakukan prediksi peluang di masa depan berdasarkan pengalaman sebelumnya yang dikenal dengan teorema bayes (Pujianto et al, 2019). Naïve bayes merupakan metode yang cocok untuk melakukan klasifikasi biner dan multiclass yang menerapkan Teknik supervised klasifikasi objek dengan menetapkan kelas ke instance menggunakan probabilitas bersyarat. Probabilitas bersyarat merupakan ukuran peluang dalam sebuah peristiwa yang terjadi berdasarkan peristiwa lain yang telah terjadi.

Salah satu algoritma machine learning yang digunakan untuk melakukan prediksi dan klasifikasi (Pengelompokan data) adalah Algoritma Naïve Bayes, dasar dari algoritma ini yaitu Teorema Probabilitas dikemukakan oleh Thomas Bayes. Algoritma Naïve Bayes digunakan untuk melakukan prediksi probabilitas masa depan berdasarkan masa lalu dan hasil yang didapat digunakan untuk pengambilan sebuah Keputusan (Damanik et al, 2021). Algoritma Naïve Bayes merupakan algoritma machine learning probabilistic yang digunakan untuk melakukan klasifikasi. Rumus dasar dari Teorema Bayes dijabarkan pada rumus 1.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Rumus 1 menjelaskan nilai $P(H | X) =$ Probabilitas hipotesis H sesuai kondisi X (Posteriori Probability). $P(X | H) =$ Probabilitas X sesuai kondisi terhadap hipotesis H. $P(H) =$ Probabilitas hipotesis H (Prior Probability). $P(X) =$ Probabilitas X.



Gambar. 2 Tipe Naïve Bayes

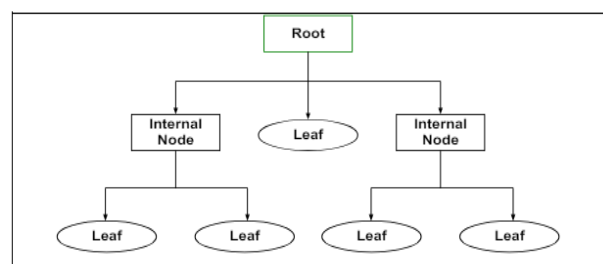
Berdasarkan Gambar 2, Algoritma Naïve Bayes Memiliki 3 tipe klasifikasi dalam melakukan pelatihan machine learning yaitu Gaussian, Multinomial, dan Bernoulli. Gaussian Naïve Bayes cocok untuk fitur yang kontinu, dengan diasumsikan distribusi Gaussian. Dalam representasi data, Gaussian Naïve Bayes

menampilkan kurva lonceng simetris yang menggambarkan rata-rata fitur tersebut. Multinomial Naïve Bayes sering dipakai dalam mengklasifikasikan dokumen, membantu mengenali kategori dokumen yang signifikan, yang dapat diabaikan, atau dokumen yang tidak diinginkan. Bernoulli Naïve Bayes, seperti Multinomial, memanfaatkan variabel Boolean untuk memprediksi kelas dengan hanya mengambil nilai ya atau tidak.

E. Klasifikasi Algoritma C4.5

Algoritma C4.5 merupakan algoritma klasifikasi yang digunakan untuk membuat model prediksi dengan membangun model pohon keputusan atau decision tree. Algoritma C4.5 dikembangkan oleh Ross Quinlan tahun 1993 dari algoritma sebelumnya yaitu ID3 (Iterative Dichotomiser 3). Prinsip kerja algoritma ID3 memiliki kesamaan dengan algoritma C4.5 namun memiliki perbedaan sehingga algoritma C4.5 mendapatkan hasil yang lebih baik dibandingkan algoritma ID3 (Pujianto et al, 2019).

Algoritma C4.5 merupakan model klasifikasi yang menrepresentasikan struktur pohon, setiap node menampilkan atribut, cabangnya tersebut menampilkan nilai dari atribut dan daun menampilkan kelas atau kategori dari data yang diolah (Aulia et al, 2023). Adapun



bagan dari pohon keputusan dapat dilihat pada gambar 3.

Gambar. 3 Pohon Keputusan C4.5

Berdasarkan Gambar 3, Pohon keputusan dimulai dengan root node, yang merupakan titik awal untuk pengambilan keputusan. Di antara root node dan leaf node, terdapat internal node yang membagi dataset menjadi subset lebih kecil berdasarkan atribut, memperjelas pola. Leaf node adalah simpul terakhir yang tidak memiliki cabang tambahan dan mewakili prediksi akhir atau keputusan untuk data yang diberikan. Algoritma C4.5 membentuk pohon keputusan

dari atas (akar) ke bawah (daun), di mana atribut teratas menjadi akar atau node yang mewakili atribut, dan yang terbawah disebut daun yang mewakili kelas (Aldino & Sulistiani, 2020).

Pemilihan atribut sebagai akar adalah yang memiliki nilai Gain tertinggi dari atribut- atribut lainnya. Akan tetapi sebelum mencari nilai Gain didahulukan mencari nilai Entropy. Informasi gain atau entropy digunakan untuk memilih pembagian akar yang optimal pada algoritma C4.5 (Ulfah & Jepara, 2022).

Entropy digunakan untuk menentukan informatif masukan atribut untuk menghasilkan sebuah atribut. Adapun untuk menentukan Gain dan Entropy dijabarkan pada rumus 2 dan 3.

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

$$Entropy(S_i) \quad (2)$$

Keterangan :

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (3)$$

Keterangan :

S : Himpunan Kasus

n : Jumlah partisi S

p_i : Proporsi dari S_i

F. Evaluasi Hasil

Tahap evaluasi hasil dilakukan sebuah pengujian dari algoritma Naive Bayes dan C4.5 yang dihasilkan menggunakan confusion matrix. Confusion Matrix merupakan alat ukur performa untuk menyelesaikan masalah klasifikasi dalam machine learning dimana hasil yang didapat berupa dua kelas atau lebih. Confusion matrix merupakan hasil evaluasi yang menggunakan tabel matrik, evaluasi confusion matrix akan menghasilkan nilai Akurasi, precision, recall dan f1-score (Putra & Wibowo, 2020).

TABEL III
DATA PREPROCESSING

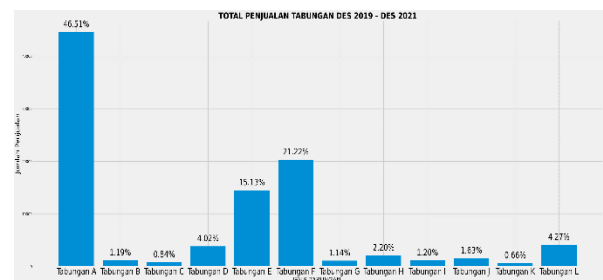
Nilai Prediksi		Positive	Negatif
	Positive	TP	FP
	Negatif	FN	TN

Berdasarkan Tabel 2. True Positive (TP) menggambarkan jumlah data yang benar - benar positif dan diprediksi dengan benar sebagai positif. False Positive (FP) mencakup data yang sebenarnya negatif tetapi keliru diprediksi sebagai positif. False Negative (FN) adalah data yang sebenarnya positif namun salah diprediksi sebagai negatif. Sedangkan True Negative (TN) mewakili jumlah data yang benar-benar negatif dan diprediksi dengan benar sebagai negatif. Dengan menggunakan metrik-metrik ini, kita dapat mengukur seberapa baik model klasifikasi dalam memprediksi kelas yang benar.

III. HASIL DAN PEMBAHASAN

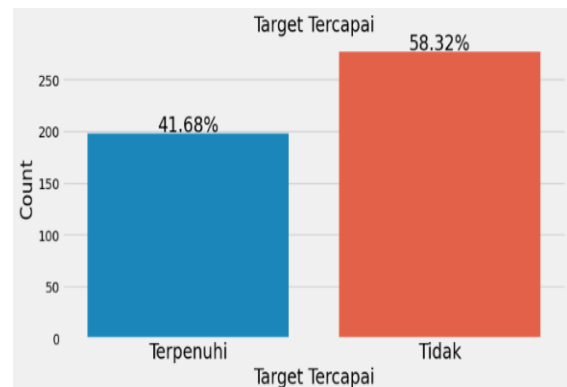
A. Visualisasi Data

Analisis jumlah penjualan produk tabungan untuk jenis tabungan dapat dilihat pada Gambar 4. Terdapat 12 jenis produk tabungan dan jumlah penjualan tertinggi terdapat pada produk Tabungan A dengan nilai persentase 46,52%. Sedangkan untuk jumlah penjualan terendah terdapat pada produk Tabungan K dengan nilai persentase 0,66%.



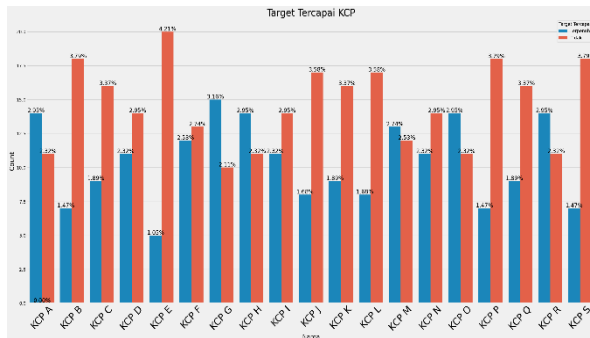
Gambar. 4 Jumlah Jenis Produk Tabungan

Persentase target tercapai dari jumlah penjualan produk dapat dilihat pada Gambar 5. target tercapai kategori tidak memiliki persentase sebesar 58,32% sedangkan kategori ya memiliki persentase sebesar 41,68%



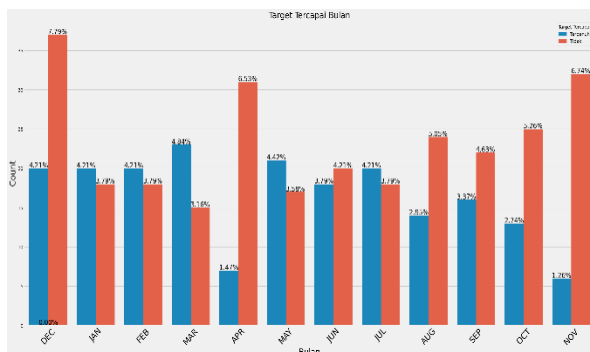
Gambar. 5 Jumlah Target Tercapai

Untuk target tercapai berdasarkan KCP dapat dilihat pada Gambar 6. Dimana persentase terbesar kategori tidak terdapat pada KCP E mencapai 4,21% sedangkan untuk kategori ya terdapat pada KCP G mencapai 3,16%.



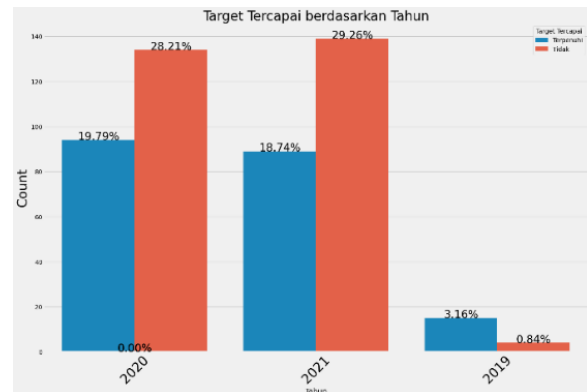
Gambar. 6 Jumlah Target Tercapai Berdasarkan KCP

Target tercapai berdasarkan bulan dapat dilihat pada Gambar 7. Dimana persentase terbesar kategori tidak terdapat pada Bulan Desember mencapai 7,79% sedangkan untuk kategori ya terdapat pada Bulan Maret mencapai 4,84%.



Gambar. 7 Jumlah Target Tercapai Berdasarkan Bulan

Target tercapai berdasarkan tahun dapat dilihat pada Gambar 8. Dimana persentase terbesar kategori tidak terdapat pada Tahun 2021 mencapai 29.26% sedangkan untuk kategori ya terdapat pada Tahun 2020 mencapai 19,79%.



Gambar. 8 Jumlah Target Tercapai Berdasarkan Tahun

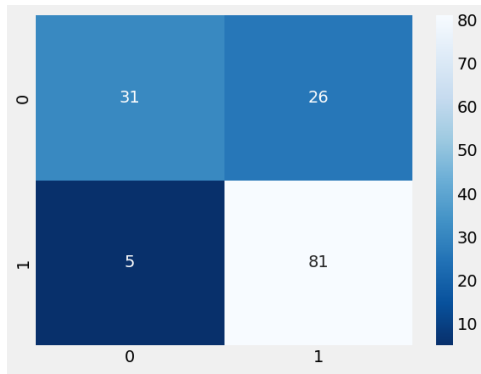
B. Hasil Klasifikasi Algoritma Naive Bayes dan C4.5

Data penjualan produk tabungan yang telah dilakukan preprocessing dan permodelan algoritma kemudian data dilakukan pembagian dengan tujuan untuk memisahkan data dengan presentasi data testing sebesar 30% dan data training sebesar 70%. Diberikan nilai random state sebesar 101 untuk hasil pemisah data akan konsisten setiap kode dijalankan dengan nilai yang sama untuk mereplikasikan hasil secara konsisten dan memungkinkan perbandingan yang adil. Kemudian dilakukan pengujian menggunakan 2 algoritma yaitu Naïve Bayes dan C4.5. Hasil pengujian klasifikasi terhadap data penjualan produk tabungan bank XYZ menggunakan algoritma naïve bayes dan algoritma c4.5 dapat dilihat pada tabel 3.

TABEL IIIII
HASIL AKURASI PENGUJIAN

Algoritma	Akurasi
Naïve Bayes	78,31%
C4.5	95,1%

Berdasarkan tabel 3 dapat dilihat bahwa hasil nilai akurasi pada algoritma naïve bayes sebesar 78,32% sedangkan menggunakan algoritma C4.5 sebesar 95,1%. Hasil pengujian yang diperoleh menunjukkan bahwa hasil yang terbaik adalah menggunakan algoritma C4.5. Confusion matrix untuk kinerja algoritma naïve bayes dapat dilihat pada Gambar 9.



Gambar. 9 Confusion Matrix Naive Bayes

Berdasarkan Gambar 9 menunjukan bahwa model algoritma Naïve Bayes menghasilkan hasil prediksi True Positive sebanyak 31 sampel yang diklasifikasikan dengan benar sebagai kelas pertama, Kemudian True Negative sebanyak 81 sampel yang diklasifikasikan benar sebagai kelas kedua, kemudian False Positive sebanyak 26 sampel yang salah diklasifikasikan sebagai kelas kedua, dan False Negative sebanyak 5 sampel yang salah diklasifikasikan sebagai kelas pertama. Confusion matrix untuk kinerja algoritma naïve bayes dapat dilihat pada Gambar 10.



Gambar. 10 Confusion Matrix C4.5

Berdasarkan Gambar 10 menunjukan bahwa model algoritma C4.5 menghasilkan hasil prediksi True Positive sebanyak 52 sampel yang diklasifikasikan dengan benar sebagai kelas pertama, Kemudian True Negative sebanyak 84 sampel yang diklasifikasikan benar sebagai kelas kedua, kemudian False Positive sebanyak 5 sampel yang salah diklasifikasikan sebagai kelas kedua, dan False Negative sebanyak 2 sampel yang salah diklasifikasikan sebagai kelas pertama.

Hasil evaluasi kinerja model klasifikasi algoritma naïve bayes dapat dilihat pada Tabel 4.

TABEL IVV
 EVALUASI KLASIFIKASI NAIVE BAYES

	Precision	recall	F1-Score
0	0,86	0,54	0,67
1	0,76	0,94	0,84

Berdasarkan tabel 4 menyajikan Evaluasi Model klasifikasi algoritma Naïve Bayes pada setiap kelas 0 dan 1. Model Naïve bayes pada kelas 0 memiliki Tingkat precision 86%, recall 54%, dan f1-score 67%. Sedangkan pada kelas 1 memiliki Tingkat precision 76%, recall 94%, dan f1-score 84%. Hasil evaluasi kinerja model klasifikasi algoritma C4.5 dapat dilihat pada Tabel 5.

TABEL V
 EVALUASI KLASIFIKASI C4.5

	Precision	recall	F1-Score
0	0,96	0,91	0,94
1	0,94	0,98	0,96

Berdasarkan tabel 5 menyajikan Evaluasi Model klasifikasi algoritma C4.5 pada setiap kelas 0 dan 1. Model C4.5 pada kelas 0 memiliki Tingkat precision 96%, recall 91%, dan f1-score 94%. Sedangkan pada kelas 1 memiliki Tingkat precision 94%, recall 98%, dan f1-score 96%.

IV. KESIMPULAN

Dari penelitian yang telah dilakukan menggunakan algoritma Naïve bayes dan C4.5 pada dataset penjualan produk tabungan Bank XYZ menghasilkan nilai akurasi yang berbeda. Nilai akurasi berdasarkan pengujian yang dilakukan algoritma naïve bayes sebesar 78,32%, sedangkan pengujian yang dilakukan algoritma C4.5 mendapatkan nilai akurasi sebesar 95,1%. Sehingga dapat disimpulkan bahwa penggunaan algoritma C4.5 dalam mengklasifikasikan pencapaian target penjualan produk tabungan lebih baik dibandingkan menggunakan algoritma naïve bayes.

DAFTAR PUSTAKA

- [1]. D. Poespita Ernawati, "Development of the Tourism Industry as the Motor of Economic Growth in Indonesia," 2019. doi: 10.54783/ijsoc.v1i4.300.

- [2]. P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang," *Jurnal Janitra Informatika dan Sistem Informasi*, vol. 2, no. 1, pp. 9–19, Apr. 2022, doi: 10.25008/janitra.v2i1.142.
- [3]. C. Zai, "Implementasi Data Mining Sebagai Pengolahan Data," 2022.
- [4]. A. Purwanto, A. Primajaya, and A. Voutama, "Penerapan Algoritma C4.5 Dalam Prediksi Potensi Tingkat Kasus Pneumonia Di Kabupaten Karawang," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 4, p. 390, Oct. 2020, doi: 10.26418/justin.v8i4.41959.
- [5]. H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *Jurnal Teknologi Informasi dan Komunikasi (TIKomsin)*, vol. 8, no. 2, Oct. 2020, doi: 10.30646/tikomsin.v8i2.500.
- [6]. N. Hidayati and A. Hermawan, "K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation," *Journal of Engineering and Applied Technology*, vol. 2, no. 2, Aug. 2021, doi: 10.21831/jeatech.v2i2.42777.
- [7]. M. Siddik, H. Hendri, R. N. Putri, Y. Desnelita, and G. Gustientiedina, "Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naïve Bayes," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 3, no. 2, pp. 162–166, Nov. 2020, doi: 10.31539/INTECOMS.V3I2.1654.
- [8]. S. Linawati, R. Ade Safitri, A. Rifqy Alfian, W. Endah Pangesti, M. Nur Winarto, and S. Nusa Mandiri, "PERBANDINGAN ALGORITMA KLASIFIKASI NAIVE BAYES DAN SVM PADA STUDI KASUS PEMBERIAN PENERIMA BEASISWA PPA," *JURNAL SWABUMI*, vol. 8, no. 1, p. 2020, 2020, doi: 10.31294/swabumi.v8i1.7708.
- [9]. N. Viktor Boyke Siahaan, D. Suhendro, D. Hartama, S. Tunas Bangsa, S. Utara, and A. Tunas Bangsa, "Penerapan Data Mining Algoritma C4.5 Terhadap Prediksi Faktor Menurunnya Hasil Panen Padi," *Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 7, pp. 27–35, 2022, doi: <http://dx.doi.org/10.30645/jurasik.v7i1.412>
- [10]. P. R. Fikrul Ilmi Zer, D. Hartama, and S. Retno Andani, "Analisa Faktor Dominan Mahasiswa Kesulitan Memahami Bahasa Pemrograman Menggunakan Metode C4.5," *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2019, doi: 10.30645/senaris.v1i0.55.
- [11]. Y. Susanti, R. Igo, A. Aziz, and M. Ahsan, "KLASIFIKASI KELAYAKAN PEMBERIAN KREDIT NASABAH BANK XYZ MENGGUNAKAN METODE ALGORITMA C4.5 DAN NAIVE BAYES," 2022. doi: 10.33479/kurawal.v5i1.551.
- [12]. Rian Pratama, B. Huda, E. Novalia, and H. Kabir, "Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Menentukan Persediaan Stok," *METIK JURNAL*, vol. 6, no. 2, pp. 115–122, Dec. 2022, doi: 10.47002/metik.v6i2.379.
- [13]. K. L. Kohsasih, Z. Situmorang, and I. Artikel, "Analisis Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular," *JURNAL INFORMATIKA*, vol. 9, no. 1, pp. 13–17, 2022, doi: 10.31294/inf.v9i1.11931.
- [14]. H. Bayu Suseno, A. Wanhari, and S. Ummi Masruroh, "Comparison of C4.5 and Naïve Bayes Algorithm for Mustahik Classification," *EAI*, 2020, doi: <http://dx.doi.org/10.4108/eai.7-11-2019.2294560>.
- [15]. A. Noviriandini and ; Nurajijah, "ANALISIS KINERJA ALGORITMA C4.5 DAN NAÏVE BAYES UNTUK MEMPREDIKSI PRESTASI SISWA SEKOLAH MENENGAH KEJURUAN," *JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER (JITK)*, vol. 5, pp. 23–28, 2019, doi: <https://doi.org/10.33480/jitk.v5i1.607>.
- [16]. K. Sahoo, A. K. Samal, J. Pramanik, and S. K. Pani, "Exploratory data analysis using python," *International Journal of Innovative*

- Technology and Exploring Engineering*, vol. 8, no. 12, pp. 4727–4735, Oct. 2019, doi: 10.35940/ijitee.L3591.1081219.
- [17]. S. Shaizadi Meraj, R. Yaakob, A. Azman, S. Nuralain Mohd Rum, and A. Shahrel Ahmad Nazri, “Artificial Intelligence in Diagnosing Tuberculosis: A Review,” vol. 9, no. 1, 2019, doi: 10.18517/ijaseit.9.1.7567.
- [18]. U. Pujiyanto, A. L. Setiawan, H. A. Rosyid, and A. M. M. Salah, “Comparison of Naïve Bayes Algorithm and Decision Tree C4.5 for Hospital Readmission Diabetes Patients using HbA1c Measurement,” *Knowledge Engineering and Data Science*, vol. 2, no. 2, p. 58, Dec. 2019, doi: 10.17977/um018v2i22019p58-71.
- [19]. A. R. Damanik, S. Sumijan, and G. W. Nurcahyo, “Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma Naïve Bayes,” *Jurnal Sistim Informasi dan Teknologi*, pp. 88–94, Aug. 2021, doi: 10.37034/jsisfotek.v3i3.49.
- [20]. A. C. Aulia, M. Fatekurohman, and I. M. Tirta, “Comparison of Online and Offline Learning During The COVID-19 Pandemic using Naïve Bayes Method and C4.5,” *BERKALA SAINSTEK*, vol. 11, no. 3, p. 153, Sep. 2023, doi: 10.19184/bst.v11i3.31737.
- [21]. A. A. Aldino and H. Sulistiani, “Decision Tree C4.5 Algorithm For Tuition Aid Grant Program Classification (Case Study: Department Of Information System, Universitas Teknokrat Indonesia),” 2020. doi: 10.21107/edutic.v7i1.8849.
- [22]. S. Ulfah and P. B. Jepara, “Komparasi Algoritma C4.5 Dengan Algoritma Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Balita Comparison Of C4.5 Algorithm With The Nave Bayes Classifier Algorithm For Classification Of Nutritional Status Of Toodler,” *JURNAL DISPROTEK*, vol. 13, no. 1, pp. 2548–4168, 2022, doi: 10.34001/jdpt.v12i2.
- [23]. D. Putra and A. Wibowo, “Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes,” *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, vol. 2, pp. 84–92, 2020, doi: 10.30645/senaris.v2i0.147.