JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Dalam Klasifikasi Data Penjualan Buku PT. XYZ

Erfina Rianty*, Kurnia Budi, Effiyaldi

Fakultas Ilmu Komputer, Program Studi Magister Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia
Email: [1,*]erfinarianty028@gmail.com, [2]kbudiz@yahoo.com, [3]effiyaldi@unama.ac.id
Email Penulis Korespondensi: erfinarianty028@gmail.com
Submitted **14-10-2025;** Accepted **01-12-2025;** Published **15-12-2025**

Abstrak

Data penjualan buku merupakan komponen penting dalam mendukung strategi pemasaran dan pengambilan keputusan manajerial. Penelitian ini bertujuan untuk mengevaluasi sekaligus membandingkan performa algoritma klasifikasi C4.5 dan Naive Bayes dalam mengolah data penjualan buku PT. Sonpedia Publishing Indonesia. Dataset yang digunakan terdiri atas 299 record, diolah menggunakan perangkat lunak RapidMiner dengan penerapan dua metode validasi, yaitu Split Data (80:20) dan 10-Fold Cross Validation. Hasil eksperimen memperlihatkan bahwa algoritma C.45 dengan metode split data mampu mencapai akurasi sebesar 88.33%, presisi 94.29%, recall 86.84%, F-Score 90.41%. Namun, performanya mengalami penurunan dengan metode 10-fold cross validation dengan akurasi 86.60%, presisi 92.53%, recall 85.64%, dan F1-Score 88.99%. Sebaliknya, algoritma Naive Bayes menunjukan hasil yang lebih stabil dan unggul. Pada split data (80:20) tercatat akurasi sebesar 90.00%, presisi 90.00%, recall 94.74%, serta F1-Score 92.31%. Kinerja meningkat saat di uji dengan 10-fold cross validation, mencapai akurasi 91.29%, presisi 92.63%, recall 93.62%, serta F1-Score 93.10%. Dari temuan ini, dapat disimpulkan bahwa naive bayes memberikan hasil klasifikasi yang lebih konsisten dan akurat dibandingkan C4.5. Penelitian ini diharapkan menjadi acuan dalam pengembangan sistem prediksi penjualan buku yang mendukung efektivitas dan efisiensi pengambilan keputusan bisnis.

Kata kunci: Klasifikasi; C4.5; Naïve Bayes; Penjualan Buku; RapidMiner

Abstract

Book sales data is an important component in supporting marketing strategies and managerial decision-making. The objective of this research is to evaluate and compare the effectiveness of the C4.5 and Naive Bayes in processing book sales data at PT. Sonpedia Publishing Indonesia. The dataset used consists of 299 book sales records, processed using RapidMiner software with two validation methods, namely Split Data (80:20) and 10-fold cross validation. Experimental results reveal that the C4.5 algorithm with the split data method obtained an accuracy 88.33%, precision 94.29%, recall 86.84%, and F-Score 90.41%. Using 10-Fold Cross Validation, the performance decreased with an accuracy 86.60%, precision of 92.53%, recall 85.64%, and F-Score 88,99%. In contrast, the Naïve Bayes algorithm demonstrated better and consistent performance. With the Split Data method (80:20), it obtained an accuracy 90.00%, precision 90.00%, recall 94.74%, and an F-Score 92.31%. Furthermore, its performance improved with 10-Fold Cross Validation, achieving an accuracy 91.29%, precision 92.63%, recall 93.62%, and F1-Score of 93.10%. These findings suggest that naive bayes produces more consistent and accurate classification results compared to C4.5. The research is intended to act as a guide for the development of book sales prediction systems that support the effetiveness and efficiency of bussiness decision making.

Keywords: Classification; C4.5; Naive Bayes; Book Sales; RapidMiner

1. PENDAHULUAN

Kegiatan penjualan memiliki posisi penting dalam perusahaan karena menjadi sumber keuntungan yang mendukung keberlanjutan usaha. Untuk mencapai target penjualan, diperlukan pengendalian internal yang berfungsi mengevaluasi efektivitas aktivitas penjualan perusahaan. Dalam pengendalian internal, beberapa aspek yang harus diperhatikan meliputi kontrol terhadap lingkungan penjualan, aktivitas penjualan, serta ketersediaan barang atau stok. Di antara ketiga aspek tersebut, pengelolaan stok sangat penting, karena jika sering terjadi kekosongan barang, maka target penjualan bisa terhambat dan berpotensi menyebabkan kehilangan pelanggan [1].

Dalam bisnis yang bergerak pada sektor penjualan buku, ketersediaan stok sangat mempengaruhi tingkat penjualan dan kepuasan konsumen. Jika jumlah persediaan melebihi permintaan pasar, perusahaan akan menanggung kerugian berupa biaya penyimpanan dan risiko buku tidak laku, khususnya pada judul yang kurang dinikmati. Sebaliknya, kekurangan stok dapat membuat pembeli beralih ketempat lain karena tidak dapat menemukan buku yang diinginkan [2].

PT. Sonpedia Publishing Indonesia adalah perusahaan yang berfokus pada penerbitan dan distribusi buku secara daring dengan berbagai kategori seperti pendidikan, ekonomi, kesehatan, teknologi hukum dan lain-lain. Berdasarkan data tahun 2022–2024 perusahaan menerbitkan sekitar 1.000 judul buku dengan jumlah cetakan bervariasi sesuai permintaan pasar. Namun terdapat ketidakseimbangan antara ketersediaan stok dan volume penjualan, yang dalam beberapa periode menimbulkan penumpukan buku menjadi usang atau tidak terjual. Faktor-faktor lain yang turut mempengaruhi tingkat penjualan antara lain harga, jumlah penjualan, rating, diskon, stok dan kategori buku. Minimnya analisis prediktif terhadap data historis penjualan membuat perusahaan kesulitan mengenali pola permintaan secara akurat, sehingga pengambilan keputusan terkait produksi maupun distribusi kurang optimal.

Beberapa studi relevan telah dilakukan sebelumnya. Penelitian [3] mengenai penjualan obat dengan algoritma C4.5 menunjukan bahwa jumlah obat menjadi atribut paling berpengaruh. Penelitian [4] yang membandingkan algoritma Naive Bayes, Decision Tree, dan KNN untuk klasifikasi produk Adidas mendapati decision tree memiliki akurasi tertinggi. Di sisi lain penelitian [5] mengungkap bahwa naive bayes dan C4.5 sama-sama efektif untuk klasifikasi data

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

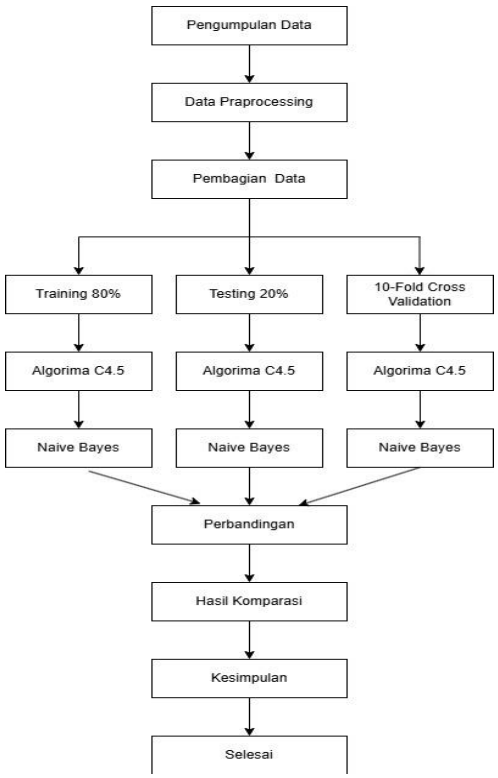
registrasi mahasiswa. Selanjutnya penelitian [6] tentang prediksi kelulusan mahasiswa memperlihatkan naive bayes lebih akurat dibandingkan C4.5. Sementara itu, penelitian [7] mengenai penjualan buah justru menunjukan C4.5 lebih unggul daripada naive bayes.

Walaupun berbagai penelitian telah membahas penggunaan algoritma C4.5 maupun naive bayes pada beragam bidang. Kajian yang membandingkan keduanya secara langsung dalam klasifikasi penjualan buku masih terbatas, terutama disektor penerbitan di Indonesia. Kebanyakan studi hanya menyoroti satu algoritma tanpa melihat kemungkinan bahwa algoritma lain mungkin lebih sesuai dengan karakteristik data. Mengingat data penjualan buku cukup kompleks karena dipengaruhi banyak atribut seperti harga, kategori, jumlah penjualan, rating dan stok. Pemilihan algoritma yang kurang tepat bisa menghasilkan analisis yang keliru dan berdampak kepada keputusan manajemen yang tidak optimal. Oleh sebab itu, penelitian ini secara khusus membandingkan kinerja algoritma C4.5 dan naive bayes dalam klasifikasi data penjualan buku untuk menentukan metode yang lebih efektif dan efisien. Hasil penelitian ini diharapkan mampu menjawab kesenjangan dalam literatur serta memberi kontribusi praktis bagi PT. Sonpedia Publishing Indonesia meningkatkan pengelolaan penjualan berbasis data.

2. METODOLOGI PENELITIAN

2.1 Alur Eksperimen

Alur eksperimen adalah urutan langkah-langkah yang terstruktur dan sistematis yang dilakukan dalam penelitian. Berikut alur eksperimen yang digunakan dalam penelitian seperti pada gambar 1:



Gambar 1. Alur Penelitian

Berdasarkan Gambar 1, tahapan eksperimen dapat dijabarkan sebagai berikut:

a. Pengumpulan Data

Data yang digunakan diperoleh dari PT. Sonpedia Publishing Indonesia dengan jumlah 300 entri. Setiap entri memiliki 9 atribut, yaitu ID Buku, Judul buku, Jumlah Penjualan, Harga, Tahun terbit, Rating, Diskon, Stok, Kategori buku.

b. Data Praprocessing

Data yang telah dikumpulkan dipersiapkan untuk proses pelatihan model. Tahapan meliputi seleksi data menghapus atribut yang tidak relevan (seperti ID Buku atau Judul Buku, Pembersihan data menangani data yang kosong dan menghapus data duplikat. Transformasi data mengubah data numerik menjadi kategori.

c. Pembagian Data

Dataset dipisahkan menggunakan dua pendekatan yaitu metode split data (80:20) yang membagi data menjadi data latih dan uji secara proposional. Serta metode 10-Fold cross validation yang membagi data menjadi 10 lipatan agar setiap bagian bergantian menjadi data uji maupun latih.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

d. Training 80% C4.5 dan Naive Bayes

Data latih digunakan unutk membangun model dengan dua algoritma. Algoritma C4.5 bentuk pohon keputuan berdasarkan informastion gain ratio untuk memilih atribut yang terbaik, sedangkan naive bayes menghitung probabilitas masing-masing kelas dengan asumsi antar atribut saling independen.

e. Testing 20% C4.5 dan Naive Bayes

Evaluasi model klasifikasi dapat dilakukan menggunakan metode confusion matrix dengan perhitungan metrik seperti accuray, precision, dan recall. Metode confusion matrix digunakan untuk mengevaluasi performa algoritma klasifikasi [8]. Pada klasifikasi biner, terdapat dua kelas keluaran, yang dijelaskan Tabel 1:

Tabel 1. *Confusion Matrix*

Kelas	Terklasifikasi Posistif	Terklasifikasi Negatip
Negatif	Nilai True Positive (TP)	False Negative (FN)
Fositif	False Positive (FP)	True Negatif (TN)

Nilai True Positive (TP), True Negatif (TN), False Positive (FP), False Negative (FN) digunakan dalam perhitungan metrik evaluasi berupa akurasi, precision, recall. Akurasi menunjukan tingkat ketepatan sistem dalam melakukan klasifikasi dengan benar, precision merepretasikan proporsi data yang teridentifikasi sebagai data positif secara tepat, sedangkan recall mengukur persentase data positif yang berhasil dikenali dengan benar oleh sistem. Perhitungan ketiga metrik tersebut ditunjukkan pada persamaan 1, persamaan 2, dan persamaan 3:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (1)$$

$$Precision = \frac{TP}{FP+TP} * 100\% \quad (2)$$

$$Recall = \frac{TP}{FN+TP} * 100\% \quad (3)$$

Keterangan:
TP : Data positif yang diklasifikasikan dengan benar
TN : Data negatif yang diklasifikasikan dengan benar
FN : Data positif yang salah diklasifikasikan sebagai negatif
FP : Data negatif yang salah diklasifikasikan sebagai positif

Kinerja algoritma juga dapat dinilai menggunakan F-score, yaitu rata-rata harmonik dari precision dan recall seperti persamaan 4 [9] :

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

f. Pengujian Menggunakan Cross Validation

Cross Validation merupakan metode evaluasi kinerja algoritma atau model yang dilakukan dengan membagi dataset menjadi dua bagian, yakni data latih dan data uji berdasarkan proporsi tertentu[10]. Sedangkan menurut [11] Metode K-Fold Cross validation merupakan teknik validasi dengan membagi dataset menjadi k lipatan, dimana satu lipatan berperan sebagai data uji dan sisanya sebagai data latih untuk memperoleh akurasi optimal.

g. Perbandingan Kinerja Model

Tahap ini bertujuan untuk melakukan perbandingan kinerja kedua algoritma berdasarkan hasil evaluasi. Nilai akurasi, precision, recall, dan F1-Score dari masing-masing model akan dianalisis dan dibandingkan untuk menentukan algoritma yang lebih efektif dalam klasifikasi buku terlaris dan tidak terlaris.

h. Hasil Komparasi

Pada tahap memfokuskan pada perbandingan performa antara algoritma C4.5 dan Naive Bayes disajikan dalam bentuk grafik. Hasil komparasi ini memberikan pemahaman mengenai keunggulan dan kelemahan masing-masing algoritma dalam hal keakuratan dan efisiensi.

i. Kesimpulan

Penelitian ini menyajikan kesimpulan dari hasil eksperimen mengenai algoritma yang lebih efektif untuk digunakan dalam proses klasifikasi buku terlaris di PT Sonpedia Publishing Indonesia serta saran untuk peningkatan model dan implementasi lebih lanjut.

2.2 Data Mining

Data mining merupakan bagian dari proses Knowledge Discovery in Database (KDD) yang berfokus pada pencarian pola dan informasi baru dalam kumpulan data. Proses ini melibatkan teknik ilmiah, analisis, interpretasi, hingga visualisasi. Tujuan utama KDD adalah mengekstrak informasi bernilai yang mudah dipahami , bermanfaat, dan bersifat baru dari dataset yang kompleks dan berukuran besar [12]. Sedangkan menurut [13] KDD merupakan suatu proses yang bertujuan untuk mengekstraksi informasi bernilai, mudah dipahami, serta bersifat baru dari kumpulan data besar dan beragam.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Proses ini mencakup interpretasi hasil diperoleh dari kumpulan data dengan mengintegrasikannya kedalam berbagai disiplin ilmu terkait.

2.3 Algoritma C4.5

Pada dasarnya, algoritma C4.5 dalam membentuk pohon keputusan memiliki tahapan sebagai berikut [14] [15] [16] [17]:
a. Pilih atribut sebagai akar
b. Buat cabang tiap-tiap nilai
c. Buat kasus dalam cabang
d. Ulangi proses setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

Untuk memilih atribut akar ditentukan berdasarkan gain tertinggi diantara semua atribut. Untuk menghitung nilai *gain* tertinggi digunakan persamaan 5:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si) \quad (5)$$

Keterangan:
S : Himpunan kasus
A : Atribut
N : Jumlah partisi A
|Si| : Jumlah kasus pada partisi ke-i
|S| : Jumlah kasus dalam S

Nilai entropy dihitung dengan persamaan 6:

$$Entropy(S) = \sum_{i=1}^{n} -Pi * log_2 pi \quad (6)$$

Dimana:
S : Himpunan kasus
n : Jumlah partisi S
pi : Proporsi dari Si terhadap S

2.4 Naive Bayes

Algoritma Naive Bayes tidak bergantung pada aturan tertentu, melainkan menggunakan konsep probabilitas dalam cabang matematika. Konsep ini digunakan untuk menentukan probabilitas tertinggi dalam proses klasfikasi melalui pengamatan terhadap frekuensi kemunculan setiap kelas pada data pelatihan [18] [20] [19]. Metode naive bayes didasarkan pada teorema probabilitas Bayes. Itu mengasumsikan bahwa atribut-artibut dalam data adalah independen satu sama lain. Naive bayes mengklasifikasikan data dengan menghitung probabilitas kemunculan target berdasarkan atribut-artibut terkait. Teorema Bayes secara umum dirumuskan pada persamaan 7 :

$$P(H \mid X) = \frac{P(H|X)\,P(H)}{P(X)} \quad (7)$$

Keterangan :
X = Data dengan class yang belum diketahui
H = Hipotesis data X merupakan suatu class spesifik
P(H|X) = Probabilitas hipotesis H *(Prior Probability*)
P(H) = Probabilitas hipotesis H *(Prior Probability*)
P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H
P(X) = Probabilitas X

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian dan pembahasan yang diperoleh berdasarkan metodologi penelitian yang telah diterapkan, serta memberikan penjelasan untuk menginterpretasikan temuan tersebut.

3.1 Pembahasan

a. Pengumpulan Data

Data dalam penelitian ini merupakan data penjualan buku PT. Sonpedia Publishing Indonesia yang diperoleh dari staf bagian penjualan dengan total sebanyak 300 data. Data tersebut mencakup penjualan buku pada tahun 2024 dan menjadi data awal sebelum dilakukan tahap praprocessing, yang terdiri dari 9 atribut yaitu id buku, Judul buku, jumlah penjualan, harga, tahun terbit, rating, diskon, stok, kategori buku. Berikut rincian dari masing-masing atribut.

b. Praprocessing Data

Praprocessing merupakan tahapan persiapan data sebelum dilakukan proses klasifikasi. Dalam penelitian ini, praprocessing meliputi tiga tahap utama.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

1. Data Selection
   Tahap seleksi dilakukan untuk mengekstrasi data yang dibutuhkan sehingga klasifikasi berjalan optimal. Pada penelitian ini, atribut id judul buku tidak digunakan dalam proses klasifikasi. Atribut id buku untuk menjelaskan nomor unik buku, atribut judul buku untuk menjelaskan identitas buku. Atribut id buku, judul buku akan dihapus menggunakan rapidminer dengan hasil pada gambar 2.

Row No.	JUMLAH PE...	HARGA (Rp)	TAHUN TER...	RATING	DISKON (%)	STOK	KATEGORI BUKU	STATUS TER...
1	85	120000	2024	3	20	100	Ekonomi dan Manajeme...	Tidak Terlaris
2	60	180000	2024	3	10	100	Komputer dan Teknologi	Tidak Terlaris
3	86	200000	2024	3	30	102	Ekonomi dan Manajeme...	Tidak Terlaris
4	101	160000	2024	4	30	115	Pendidikan	Tidak Terlaris
5	104	160000	2024	5	20	105	Ekonomi dan Manajeme...	Terlaris
6	100	80000	2024	3	20	107	Kesehatan	Tidak Terlaris
7	84	160000	2024	5	0	110	Komputer dan Teknologi	Tidak Terlaris
8	120	160000	2024	5	30	150	Komputer dan Teknologi	Terlaris
9	112	120000	2024	3	20	133	Komputer dan Teknologi	Terlaris
10	89	180000	2024	4	10	95	Pendidikan	Tidak Terlaris

Gambar 2. Atribut Penelitian Setelah Menghapus Id buku dan Judul buku

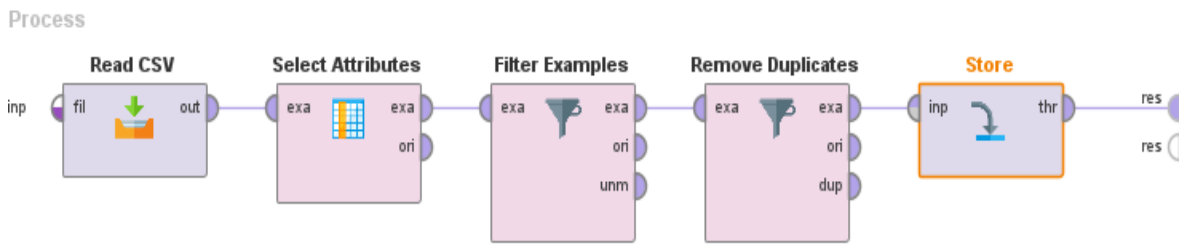
Gambar 2 menunjukkan dataset yang telah melalui tahap pembersihan data dengan menghapus atribut ID Buku dan Judul Buku yang dianggap tidak relevan dalam proses klasifikasi. Dataset ini terdiri dari 8 atribut, termasuk atribut target, yaitu Status Terlaris. Penelitian ini menggunakan atribut mencakup jumlah penjualan, harga buku, tahun terbit, rating, diskon, jumlah stok, kategori buku, dan status terlaris. Data ini akan digunakan sebagai dasar dalam proses klasifikasi untuk memprediksi apakah suatu buku tergolong "Terlaris" atau "Tidak Terlaris".

2. Data Cleaning
   Tahap ini dilakukan dengan cara menghapus data nilai yang kosong karena data tidak dapat dipakai dalam proses analisis. Gambar 2 menampilkan proses *data cleaning* yang dilakukan dengan menghapus data yang mengandung *missing value* karena tidak dapat digunakan dalam proses analisis. Data kosong ini dapat mempengaruhi akurasi hasil klasifikasi apabila tidak ditangani dengan tepat. Setelah proses pembersihan data dilakukan, jumlah data yang valid tersisa sebanyak 299 entri.

Row No.	JUMLAH PE...	HARGA (Rp)	TAHUN TER...	RATING	DISKON (%)	STOK	KATEGORI BUKU	STATUS TER...
87	117	60000	2024	4	10	158	Ekonomi dan Mana...	Terlaris
88	100	120000	2024	4	20	130	Kesehatan	Tidak Terlaris
89	75	160000	2024	4	5	90	Komputer dan Tekn...	Tidak Terlaris
90	114	120000	2024	5	10	141	Kesehatan	Terlaris
91	127	120000	2024	5	5	162	Komputer dan Tekn...	Terlaris
92	?	120000	2024	5	0	126	Ekonomi dan Mana...	Terlaris
93	124	120000	2024	4	30	150	Ekonomi dan Mana...	Terlaris
94	116	180000	2024	5	30	148	Pendidikan	Terlaris
95	73	120000	2024	4	0	83	Kesehatan	Tidak Terlaris
96	116	120000	2024	4	5	120	Pendidikan	Terlaris

Gambar 3. Data Missing Value

3. Remote Duplicate
   Gambar 4 menampilkan tahapan proses pra-pemrosesan data dalam RapidMiner. Proses dimulai dengan *Read CSV* untuk membaca file dataset, dilanjutkan dengan *Select Attributes* untuk memilih atribut yang relevan.



Gambar 4. Proses Pra-processing Data di RapidMiner

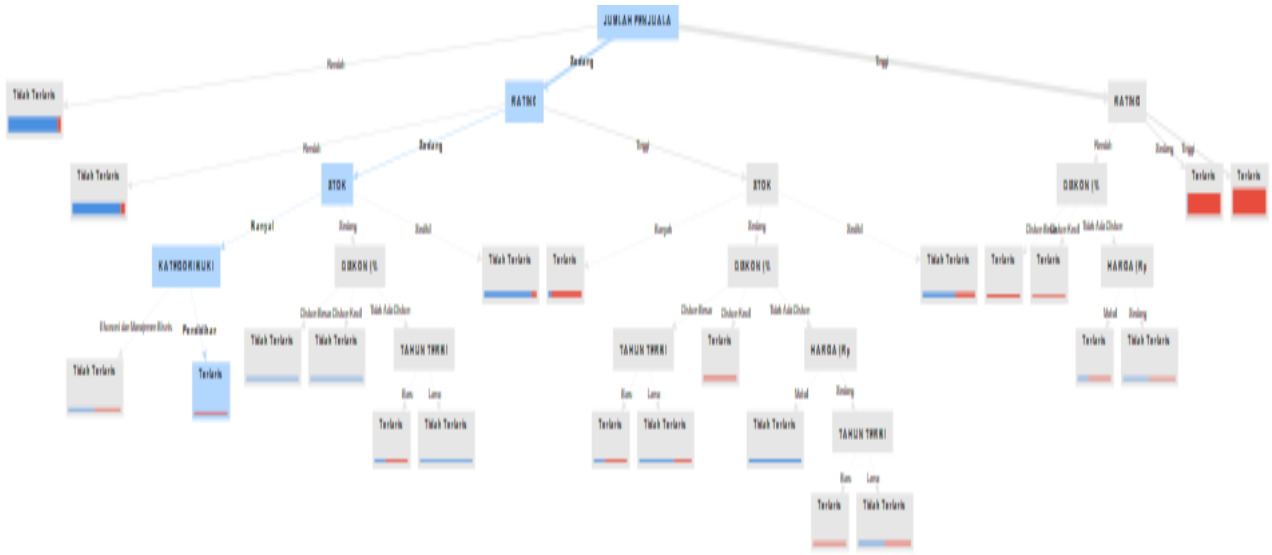
JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Kemudian, *Filter Examples* digunakan untuk menyaring data sesuai kriteria tertentu, dan *Remove Duplicates* berfungsi menghapus data ganda. Lalu data yang telah dibersihkan disimpan menggunakan operator *Store*. Langkah berikutnya melakukan *transformasi data* sesuai dengan kebutuhan proses *data mining* dalam tahap klasifikasi.

3.2 Training 80% Algoritma C4.5

Pada tahap ini, C4.5 dilatih menggunakan menggunakan pembagian (80:20), dengan 239 data sebagai data pelatihan dan 60 data sebagai data pengujian. Data pelatihan dimanfaatkan untuk menyusun model pohon keputusan, di mana perhitungan entropy dan information gain dilakukan guna mendapatkan atribut akar, kemudian proses dilanjutkan secara rekursif hingga terbentuk pohon keputusan.

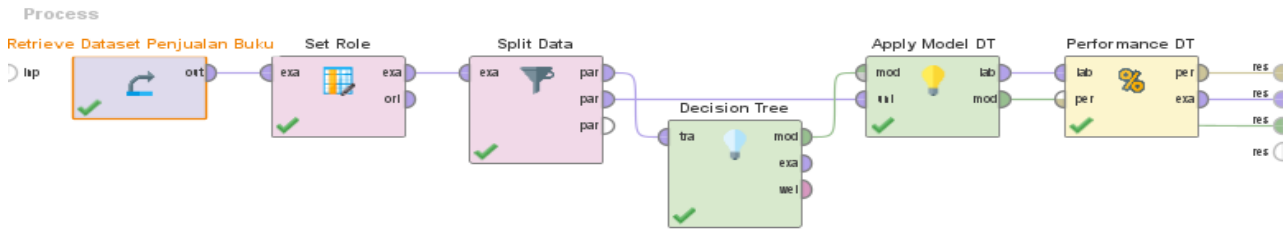
Gambar 5 memperlihatkan struktur pohon keputusan yang dibentuk algoritma C4.5 berdasarkan metode Split Data 80:20. Pohon keputusan ini dihasilkan dari proses pelatihan menggunakan 239 data latih dari total 299 data. Struktur pohon keputusan tersebut merepresentasikan aturan klasifikasi yang terbentuk untuk membedakan kategori "terlaris" dan "tidak terlaris" pada data penjualan buku.



Gambar 5. Pohon Keputusan Algoritma C4.5

Berdasarkan gambar 5 menunjukan atribut jumlah Penjualan berada pada posisi akar pohon, yang menunjukkan bahwa atribut ini memiliki nilai gain tertinggi dan paling berpengaruh dalam proses klasifikasi buku ke dalam kategori terlaris atau tidak terlaris. Dari akar ini, pohon bercabang ke atribut-atribut lain seperti Rating, Stok, Harga, Tahun Terbit, dan Kategori Buku, yang masing-masing dipilih berdasarkan nilai gain tertinggi berikutnya pada subset data. Struktur pohon ini mencerminkan urutan pengaruh atribut terhadap hasil klasifikasi, dengan Jumlah Penjualan sebagai faktor dominan dalam menentukan prediksi penjualan buku.

Berikut visualisasi hasil pengujian menggunakan split data 80:20 algoritma C4.5.



Gambar 6. Pengujian C4.5 Split Data 80:20 di RapidMiner

Gambar 6 menunjukkan alur proses klasifikasi data penjualan buku menggunakan algoritma Decision Tree (C4.5) di aplikasi RapidMiner. Tahapan diawali dengan operator Retrieve Dataset Penjualan Buku, yaitu untuk memanggil dataset yang berisi 299 data penjualan buku. Selanjutnya, digunakan operator Set Role untuk menentukan atribut target (label) yang akan diprediksi, misalnya kelas "Terlaris" dan "Tidak Terlaris".

Setelah itu, dataset diproses menggunakan operator Split Data dengan pembagian 80% sebagai data latih (239 data) dan 20% sebagai data uji (60 data). Data latih kemudian dimasukkan ke dalam operator Decision Tree, yang berfungsi membangun model klasifikasi berupa pohon keputusan melalui perhitungan entropy dan information gain. Model pohon keputusan yang dihasilkan diterapkan pada data uji melalui operator Apply Model, sehingga dapat memprediksi kelas pada data uji. Tahap terakhir adalah Performance DT, yang dipakai untuk menilai performa model dengan menghasilkan metrik evaluasi seperti akurasi, precision, recall, F1-Score, serta confusion matrix. Dengan alur ini,

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

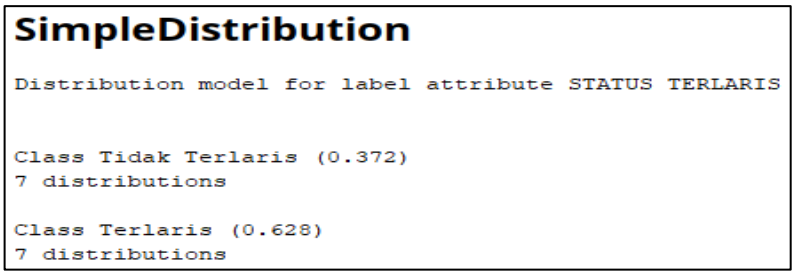
RapidMiner mampu secara sistematis membandingkan hasil klasifikasi antara data aktual dan hasil prediksi dari algoritma C4.5.

3.3 Training 80% Algoritma Naive Bayes

Pada tahap ini, algoritma Naïve Bayes membangun model dengan menghitung probabilitas prior dari setiap kelas (Laris/Tidak Laris) dan probabilitas kondisional setiap atribut terhadap kelas. Hasil training berupa model probabilistik yang digunakan untuk mengestimasi kelas data uji.

Tabel 2. Probabilitas Prior

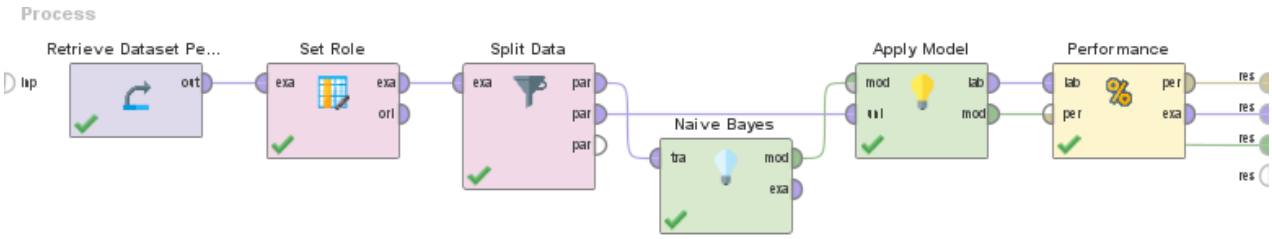
Status Terlaris	Jumlah Data	Probabilitas
Terlaris	188	0,628
Tidak Terlaris	111	0,372



Gambar 7. Perhitungan Probabilitas di RapidMiner

Dalam pemodelan ini, hasil perhitungan probabilitas menunjukkan bahwa kelas Terlaris memiliki probabilitas sebesar 0,628 atau 62,80%, sedangkan kelas Tidak Terlaris memiliki probabilitas sebesar 0,372 atau 37,20%. Jika kedua kelas tersebut digabungkan, total probabilitasnya menjadi 100%.

Berikut merupakan visualisasi hasil pengujian menggunakan split data (80:20) algoritma naive bayes.



Gambar 8. Pengujian Naive Bayes Split Data (80:20) di RapidMiner

Gambar 7. menunjukkan alur proses klasifikasi data penjualan buku menggunakan algoritma Naive Bayes di aplikasi RapidMiner. Tahapan diawali dengan operator Retrieve Dataset Penjualan Buku, yaitu untuk memanggil dataset yang berisi 299 data penjualan buku. Selanjutnya, digunakan operator Set Role untuk menentukan atribut target (label) yang akan diprediksi, misalnya kelas *"Terlaris"* dan *"Tidak Terlaris"*.

Selanjutnya, dataset diproses operator Split Data dengan pembagian 80% (239 data) untuk data latih dan 20% (60 data) untuk data uji. Data latih kemudian dimasukkan ke dalam operator Naive Bayes, yang berfungsi membangun model klasifikasi berdasarkan perhitungan probabilitas prior dan probabilitas kondisional dari setiap atribut terhadap kelas.

Model Naive Bayes yang dihasilkan kemudian diterapkan pada data uji melalui operator Apply Model, sehingga dapat memprediksi kelas pada data uji. Tahap terakhir adalah operator Performance, yang digunakan untuk mengevaluasi kinerja model dengan metrik evaluasi seperti akurasi, precision, recall, F1-Score, serta confusion matrix. Dengan alur ini, RapidMiner mampu secara sistematis membandingkan hasil klasifikasi antara data aktual dan hasil prediksi dari algoritma Naive Bayes.

3.4 Testing 20% Algoritma C.45

Setelah model dilatih menggunakan 239 data latih, tahap berikutnya adalah melakukan pengujian terhadap 60 data uji (20%). Hasil pengujian kemudian dianalisis menggunakan confusion matrix guna memperoleh metrik evaluasi yang mencakup akurasi, precision, recall dan F1-score.

Evaluasi hasil klasifikasi pada algoritma C4.5 menggunakan metode confusion matrix.

Tabel 3. Confusion Matrix C4.5 (Split Data 80:20)

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP = 33	FN = 5
Negatif	FP = 2	TN = 20

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Berdasarkan hasil klasifikasi yang ditampilkan pada tabel confusion matrix, dapat diamati bahwa:

a. True Positive (TP) adalah kondisi ketika data diklasifikasikan sebagai terlaris, dan memang benar merupakan data terlaris. Jumlah data dalam kategori ini adalah 33 data.
b. False Positive (FP) adalah kondisi ketika data diklasifikasikan sebagai terlaris, namun sebenarnya adalah tidak terlaris. Jumlah data dalam kategori ini adalah 2 data.
c. False Negative (FN) adalah kondisi ketika data diklasifikasikan sebagai tidak terlaris, padahal sebenarnya merupakan data terlaris. Jumlah data dalam kategori ini adalah 5 data.
d. True Negative (TN) adalah kondisi ketika data diklasifikasikan sebagai tidak terlaris, dan memang benar merupakan data tidak terlaris. Jumlah data dalam kategori ini adalah 20 data.

Perhitungan akurasi menggunakan rumus berikut

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$
$$= \frac{33 + 20}{33 + 20 + 2 + 5} * 100\%$$
$$= 88{,}33\ \%$$

Perhitungan *precision* menggunakan rumus berikut

$$Precision = \frac{TP}{FP + TP} * 100\%$$
$$= \frac{33}{2 + 33} * 100\%$$
$$= 94{,}29\%$$

Perhitungan *recall* menggunakan rumus berikut

$$Recall = \frac{TP}{FN + TP} * 100\%$$
$$= \frac{33}{5 + 33} * 100\%$$
$$= 86{,}84\%$$

Perhitungan *F1 Score* menggunakan rumus berikut

$$F1\ Score = 2 * \frac{Precision\ *\ Recall}{Precision\ +\ Recall}$$
$$= 2 * \frac{94{,}29\% * 86{,}84\%}{94{,}29\% + 86{,}84\%}$$
$$= 90{,}41\%$$

3.5 Testing 20% Algoritma Naive Bayes

Setelah model dilatih menggunakan 239 data latih, tahap berikutnya adalah pengujian terhadap 60 data uji (20%). Hasil dari proses pengujian tersebut dianalisis melalui confusion matrix guna mendapatkan metrik evaluasi yang mencakup akurasi, presisi, recall dan f1-score.

Hasil klasifikasi algoritma naive bayes menggunakan confusion matrix.

Tabel 4. Confusion Matrix Naive Bayes (Split Data 80:20)

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP = 36	FN = 2
Negatif	FP = 4	TN = 18

Mengacu pada tabel confusion matrix, hasil klasifikasi memperlihatkan bahwa:

1. True Positive (TP) adalah kondisi ketika data diklasifikasikan sebagai terlaris, dan memang benar merupakan data terlaris. Jumlah data dalam kategori ini adalah 36 data.
2. False Positive (FP) adalah kondisi ketika data diklasifikasikan sebagai terlaris, namun sebenarnya adalah tidak terlaris. Jumlah data dalam kategori ini adalah 4 data.
3. False Negative (FN) adalah kondisi ketika data diklasifikasikan sebagai tidak terlaris, padahal sebenarnya merupakan data terlaris. Jumlah data dalam kategori ini adalah 2 data.
4. True Negative (TN) adalah kondisi ketika data diklasifikasikan sebagai tidak terlaris, dan memang benar merupakan data tidak terlaris. Jumlah data dalam kategori ini adalah 18 data.

Perhitungan akurasi menggunakan rumus berikut

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$
$$= \frac{36 + 18}{36 + 18 + 4 + 2} * 100\%$$
$$= 90{,}00\ \%$$

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Perhitungan *precision* menggunakan rumus berikut

$$Precision = \frac{TP}{FP + TP} * 100\%$$
$$= \frac{36}{4 + 36} * 100\%$$
$$= 90{,}00\%$$

Perhitungan *recall* menggunakan rumus berikut

$$Recall = \frac{TP}{FN + TP} * 100\%$$
$$= \frac{36}{2 + 36} * 100\%$$
$$= 94{,}74\%$$

Perhitungan *F1 Score* menggunakan rumus berikut

$$F1\ Score = 2 * \frac{Precision\ *\ Recall}{Precision\ +\ Recall}$$
$$= 2 * \frac{90{,}00\% * 94{,}74\%}{90{,}00\% + 94{,}74\%}$$
$$= 93{,}31\%$$

3.6 Pengujian 10-Fold Cross Validatian C4.5

Berdasarkan hasil pengujian yang disajikan pada Tabel 5, algoritma C4.5 menunjukan tingkat akurasi yang relatif lebih stabil pada setiap fold. Meskipun terdapat sedikit variasi, perbedaan antar fold tidak bersifat signifikan. Akurasi rata-rata yang dihasilkan melalui 10-Fold Cross Validation sebesar 86,60%, sehingga dapat disimpulkan bahwa model memiliki kinerja klasifikasi yang cukup baik secara keseluruhan.

Tabel 5. Pengujian 10-Fold Cross Validation C4.5

K-Fold	*Akurasi*
1	90,00%
2	90,00%
3	83,33%
4	79,31%
5	86,67%
6	83,33%
7	86,67%
8	86,67%
9	96,67%
10	83,33%
Rata-rata	86,60%

Perhitungan *accuracy* menggunakan rumus berikut

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$
$$= \frac{161 + 98}{161 + 98 + 13 + 27} * 100\%$$
$$= 86{,}60\%$$

Perhitungan *precision* menggunakan rumus berikut

$$Precision = \frac{TP}{FP + TP} * 100\%$$
$$= \frac{161}{13 + 161} * 100\%$$
$$= 92{,}53\%$$

Perhitungan *recall* menggunakan rumus berikut

$$Recall = \frac{TP}{FN + TP} * 100\%$$
$$= \frac{161}{27 + 161} * 100\%$$
$$= 85{,}64\%$$

Perhitungan *F1 Score* menggunakan rumus berikut

$$F1\ Score = 2 * \frac{Precision\ *\ Recall}{Precision\ +\ Recall}$$
$$= 2 * \frac{92{,}53\% * 85{,}64\%}{92{,}53\% + 85{,}64\%}$$

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$= 88{,}99\%$

3.7 Pengujian 10-Fold Cross Validatian Naive Bayes

Mengacu pada hasil pengujian yang ditampilkan pada Tabel 6, akurasi algoritma Naïve Bayes menunjukkan performa yang cukup stabil pada setiap fold. Meskipun terdapat variasi nilai akurasi antar fold, perubahan tersebut tidak terlalu signifikan dan masih berada dalam rentang yang konsisten. *Mean* akurasi yang diperoleh melalui pengujian 10-fold cross validation adalah sebesar 91,29%, yang mencerminkan bahwa model memiliki kemampuan klasifikasi yang baik secara keseluruhan.

Tabel 6. Pengujian 10-Fold Cross Validation Naive Bayes

K-Fold	Akurasi
1	86,67%
2	90,00%
3	93,33%
4	86,21%
5	96,67%
6	93,33%
7	86,67%
8	90,00%
9	93,33%
10	96,67%
Rata-rata	**91,29%**

Perhitungan akurasi menggunakan rumus berikut

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$
$$= \frac{176 + 98}{176 + 97 + 12 + 14} * 100\%$$
$$= 91{,}29\%$$

Perhitungan *precision* menggunakan rumus berikut

$$Precision = \frac{TP}{FP + TP} * 100\%$$
$$= \frac{176}{14 + 176} * 100\%$$
$$= 92{,}63\%$$

Perhitungan *recall* menggunakan rumus berikut

$$Recall = \frac{TP}{FN + TP} * 100\%$$
$$= \frac{176}{12 + 176} * 100\%$$
$$= 93{,}62\%$$

Perhitungan *F1 Score* menggunakan rumus berikut

$$F1\ Score = 2 * \frac{Precision\ *\ Recall}{Precision\ +\ Recall}$$
$$= 2 * \frac{92{,}63\% * 93{,}62\%}{92{,}63\% + 93{,}62\%}$$
$$= 93{,}10\%$$

3.8 Perbandingan Kinerja Model

Berdasarkan evaluasi klasifikasi, diperoleh kinerja algoritma C4.5 dan Naive Bayes yang diukur melalui nilai akurasi, precision, recall dan F-Score. Berikut hasil pengukuran performa kedua algoritma.

Tabel 7. Perbandingan performa algoritma C4.5 dan Naive Bayes

Algoritma	Metode Pengujian	Akurasi	Precision	Recall	F1-Score
C4.5	Split Data (80:20)	88,33%	94,29%	86,84%	90,41%
Naive Bayes	Split Data (80:20)	90,00%	90,00%	94,74%	92,31%
C4.5	10-Fold CV	86,60%	92,53%	85,64%	88,99%
Naive Bayes	10-Fold CV	91,29%	92,63%	93,62%	93,10%

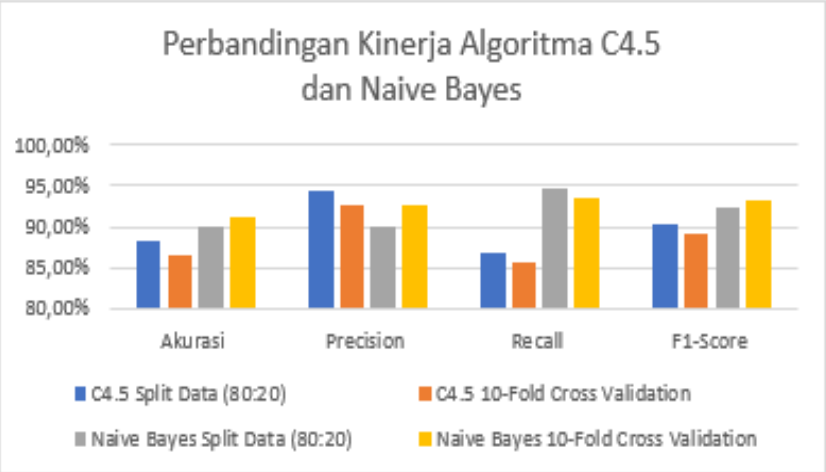
Hasil Tabel 7. menunjukkan bahwa algoritma C4.5, dengan metode Split Data (80:20) diperoleh akurasi sebesar 88,33%, precision sebesar 94,29%, recall sebesar 86,84%, dan F1-score sebesar 90,41%. Sementara itu, dengan metode 10-Fold Cross Validation, nilai akurasi sedikit menurun menjadi 86,60%, precision 92,53%, recall 85,64%, dan F1-Score

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9345
Hal 793-804
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

88,99%. Berdasarkan hasil tersebut, dapat diinterpretasikan bahwa C4.5 cukup baik mengenali data positif (recall tinggi), namun akurasinya relatif lebih rendah bila dibandingkan dengan naive bayes. Sebaliknya, algoritma naive bayes memperlihatkan performa konsisten dan unggul pada hampir semua metrik evaluasi. Berdasarkan pengujian menggunakan metode Split Data (80:20), algoritma Naive Bayes mencapai akurasi 90,00%, precision 90,00%, recall 94,74%, dan F1-score 92,31%. Sedangkan pengujian dengan metode 10-fold cross validation, hasil evaluasi menunjukan peningkatan performa dengan akurasi 91,29%, precision 92,63%, recall 93,62%, dan F1-score 93,10%.

3.9 Hasil Komparasi

Hasil komparasi ini menyajikan perbandingan kinerja algoritma C4.5 dan naive bayes berdasarkan metrik evaluasi berupa akurasi, precision, dan F1-Score.



Gambar 9. Grafik Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes

Gambar 9. Naive bayes dengan metode 10-Fold Cross validation menghasilkan performa terbaik dengan akurasi 91,29%, lebih unggul dibandingkan algoritma C4.5 yang memperoleh 86,60%. Hasil penelitian ini mengindikasikan bahwa naive bayes memiliki keandalan lebih optimal dalam melakukan klasifikasi secara menyeluruh. Pada metrik precision, nilai tertinggi diperoleh oleh algoritma C4.5 dengan metode Split Data (80:20), yaitu sebesar 94,29%. Hal ini menunjukkan bahwa pada skema pembagian data tersebut, C4.5 lebih mampu meminimalkan kesalahan dalam melakukan prediksi terhadap kelas positif dibandingkan Naïve Bayes. Sementara itu, pada metrik recall, algoritma Naïve Bayes dengan metode Split Data (80:20) menghasilkan nilai tertinggi sebesar 94,74%, melampaui algoritma C4.5 yang hanya mencapai 86,84%. Temuan penelitian ini memperlihatkan bahwa naive bayes mampu lebih baik dalam mengidentifikasi seluruh data positif. Pada metrik F1- Score, naive bayes melalui pengujian 10-Fold CV kembali memperoleh nilai tertinggi sebesar 93,10%, sedangkan C4.5 hanya mencapai 88,99%. Hal ini menunjukan bahwa naive bayes mampu mencapai keseimbangan yang lebih optimal antara precision dan recall dibandingkan C4.5. Secara keseluruhan, proses evaluasi memperlihatkan bahwa naive bayes memiliki keunggulan dibandingkan C4.5 pada sebagian besar metrik evaluasi. Meskipun demikian, keunggulan C4.5 pada precision dengan metode Split Data (80:20) mengindikasikan bahwa pemilihan algoritma klasifikasi perlu mempertimbangkan karakteristik data serta tujuan analisis yang ingin dicapai dalam penelitian.

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode C4.5 dan naive bayes menunjukan kinerja yang optimal dalam mengklasifikasikan data penjualan buku di PT. Sonpedia Publishing Indonesia, meskipun terdapat perbedaan signifikan pada beberapa metrik evaluasi. Algoritma C4.5 menunjukkan keunggulan pada aspek precision, terutama ketika menggunakan metode Split Data (80:20). Hal ini mengindikasikan bahwa model C4.5 lebih efektif dalam meminimalkan kesalahan klasifikasi pada kelas positif, sehingga cocok diterapkan pada kasus-kasus di mana kesalahan positif palsu (false positive) harus dihindari. Namun demikian, performa C4.5 sedikit menurun pada skema 10-Fold Cross Validation, yang memperlihatkan bahwa algoritma ini kurang konsisten ketika diuji dengan data yang lebih bervariasi. Sebaliknya, algoritma Naïve Bayes tampil lebih stabil dan konsisten pada kedua metode evaluasi. Dengan Split Data (80:20), Naive Bayes mampu mencapai akurasi yang cukup tinggi sekaligus menghasilkan recall dan F1-Score yang seimbang. Bahkan, pada 10-Fold Cross Validation, performanya meningkat lebih baik lagi dengan akurasi mencapai 91,29% serta hasil recall dan F1-Score yang lebih optimal dibandingkan C4.5. Hal ini membuktikan bahwa Naive Bayes mampu menangani variasi data dengan lebih baik serta menjaga keseimbangan antara precision dan recall. Dengan demikian, naive bayes menunjukan performa tinggi dibandingkan C4.5 dalam klasifikasi penjualan buku, baik dari sisi akurasi maupun konsistensi kinerja.

REFERENCES

[1] A. Rachmanto, "Sistem Informasi Akuntansi Penjualan Perusahaan Dagang," *J. Ris. Akunt.*, vol. 5, no. 1, 2017, doi: 10.34010/jra.v5i1.506.
[2] Ahmad Thariq, "Implementasi Market Basket Analysis Menggunakan Algoritma Apriori pada Data Penjualan Buku," *J. Kolaboratif Sains*, vol. 6, no. 3, hal. 154–163, 2023, doi: 10.56338/jks.v6i3.3333.
[3] R. Sovia, A. Muhammad, S. Arlis, Guslendra, dan S. Defit, "Analysis of sales levels of pharmaceutical products by using data mining algorithm C45," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 1, hal. 476–484, 2021, doi: 10.11591/ijeecs.v22.i1.pp476-484.
[4] L. Firdaus dan T. Setiadi, "Perbandingan Algoritma Naive Bayes, Decision Tree, dan KNN untuk Klasifikasi Produk Populer Adidas US dengan Confusion Matrix," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, hal. 185, 2023, doi: 10.30865/json.v5i2.6124.
[5] N. Yahya dan A. Jananto, "Komparasi Kinerja Algoritma C.45 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaan mahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang)," *Pros. SENDI*, no. 2014, hal. 978–979, 2019.
[6] F. Solikhah, M. Febianah, A. L. Kamil, W. A. Arifin, dan Shelly Janu Setyaning Tyas, "Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan," *Tematik*, vol. 8, no. 1, hal. 96–103, 2021, doi: 10.38204/tematik.v8i1.576.
[7] H. P. Herlambang, F. Saputra, M. H. Prasetiyo, D. Puspitasari, dan D. Nurlaela, "Perbandingan Klasifikasi Tingkat Penjualan Buah di Supermarket dengan Pendekatan Algoritma Decision Tree, Naive Bayes dan K-Nearest Neighbor," *J. Insa. - J. Inf. Syst. Manag. Innov.*, vol. 3, no. 1, hal. 21–28, 2023, doi: 10.31294/jinsan.v3i1.2097.
[8] M. Kamil dan W. Cholil, "Analisis Perbandingan Algoritma C4.5 dan Naive Bayes pada Lulusan Tepat Waktu Mahasiswa di Universitas Islam Negeri Raden Fatah Palembang," *J. Inform.*, vol. 7, no. 2, hal. 97–106, 2020, doi: 10.31294/ji.v7i2.7723.
[9] A. Tripathy, A. Agrawal, dan S. K. Rath, "Classification of Sentiment Reviews using N-gram Machine Learning Approach Classification of sentiment reviews using n-gram machine learning approach," *Expert Syst. Appl.*, vol. 57, no. October 2017, hal. 117–126, 2016, doi: 10.1016/j.eswa.2016.03.028.
[10] M. F. Kurniawan, N. Bayes, dan N. Bayes, "Komparasi Algoritma Data Mining untuk Klasifikasi Penyakit Kanker Payudara," *J. stmik wp*, no. 1, hal. 1–8, 2014.
[11] A. Supriyadi, "Perbandingan Algoritma Naive Bayes dan Decision Tree (C4.5) dalam Klasifikasi Dosen Berprestasi," *Gener. J.*, vol. 7, no. 1, hal. 39–49, 2023.
[12] T. Jurnal, S. Dan, R. Rayuwati, dan K. Koko, "Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode Apriori (studi kasus : Kota Medan)," vol. 16, no. 1, hal. 69–82, 2020.
[13] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, hal. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
[14] Gellysa Urva dkk, *Penerapan Data Mining di Berbagai Bidang*. Jambi: PT Sonpedia Publishing Indonesia, 2023.
[15] T. Azhima dan Y. Siswa, "Analisis Penerapan Optimasi Perbandingan Kinerja Algoritma C4 . 5 dan Naïve Bayes Berbasis Particle Swarm Optimization (PSO) Untuk Mendeteksi Kanker Payudara," vol. 2, no. Vii, hal. 1–9, 2018.
[16] M. Kantardzic, *Data Mining Concepts, Model, Methods, and Algotithms*. Wiley - IEEE Press, 2020.
[17] F. Fatmawati dan N. Narti, "Perbandingan Algoritma C4.5 dan Naive Bayes Dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 1, hal. 1–12, 2022, doi: 10.35746/jtim.v4i1.196.
[18] A. Julianto dan S. Andayani, "Penerapan Data Mining Untuk Klasifikasi Produk Terlaris Menggunakan Algoritma Naive Bayes Pada Bengkel Motor," *J. Sist. dan Teknol. Inf. Komun.*, 2024.
[19] L. Rifky, Z. Nugraha, B. Saputra, D. Pratama, E. Raswir, dan Y. Pratama, "Implementasi Data Mining Untuk Penjualan Mobil Menggunakan Metode Naive Bayes," *J. Inform. Dan Rekayasa Komputer(JAKAKOM)*, vol. 2, no. 2, hal. 225–230, 2022, doi: 10.33998/jakakom.2022.2.2.109.
[20] Randi Farmana Putra dkk, *Data Mining Algoritma dan Penerapannya*. Jambi: PT Sonpedia Publishing Indonesia, 2023.