KOMPARASI ALGORITMA NAIVE BAYES DAN DECISION TREE PADA KLASIFIKASI PENERIMAAN PESERTA DIDIK BARU

Kesi Aprilia[1], Ulfa Khaira[2], Benedika Ferdian Hutabarat[3]
[1,2,3]Program Studi Sistem Informasi, FST, Universitas Jambi
*e-mail : *[1]kesiaprillia9@gmail.com, [2]ulfakhaira@unja.ac.id, [3]benedika@unja.ac.id*

Penerimaan Peserta Didik Baru (PPDB) merupakan tahapan penting dalam menyeleksi calon peserta didik yang memenuhi kriteria lembaga pendidikan. Di Yayasan Sahabat Qur'an Al-Karim Jambi, peningkatan jumlah pendaftar tidak sebanding dengan kapasitas penerimaan, sehingga dibutuhkan proses seleksi yang lebih tepat dan objektif. Penelitian ini bertujuan untuk mengaplikasikan metode data mining dengan melakukan perbandingan antara algoritma Naïve Bayes dan Decision Tree CART untuk mengetahui algoritma yang paling efektif dalam klasifikasi penerimaan peserta didik baru. Proses penelitian mencakup tahap praproses data, pembagian dataset dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian, penerapan kedua algoritma, serta dilakukan evaluasi mengggunakan k-fold cross validation dan confusion matrix. Temuan penelitian menunjukkan bahwa algoritma Decision Tree memberikan performa paling unggul dengan akurasi 98,77%, presisi 99,19%, recall 97,56%, dan f1-score 98,34%, sedangkan algoritma Naïve Bayes memperoleh akurasi 94,44%, presisi 96,54%, recall 89,02%, dan f1-score 92,04%.

***Kata Kunci*—Penerimaan Peserta Didik Baru, Data Mining, Naïve Bayes, Decision Tree CART, Klasifikasi**

I. PENDAHULUAN

Penerimaan peserta didik baru merupakn tahapan pendaftaran bagi siswa yang akan memulai jenjang pendidikan dan menjadi pintu awal bagi mereka dalam memasuki lingkungan pendidikan, baik formal maupun nonformal. Kegiatan ini merupakan tahapan awal yang memiliki peran signifikan dalam menjamin kelancaran pelaksanaan proses pendidikan di sekolah maupun lembaga, dengan dukungan tenaga pengajar serta fasilitas dan sarana prasarana yang memadai guna menghasilkan peserta didik yang kompeten dan berwawasan luas. Selain itu, sekolah atau yayasan diharapkan mampu melaksanakan proses seleksi secara optimal agar memperoleh calon siswa yang benar-benar memenuhi standar kualifikasi yang telah ditetapkan [1].

Yayasan Sahabat Qur'an merupakan suatu Lembaga pendidikan yang memiliki beberapa cabang yang terletak di kota jambi. Lembaga ini berorientasi pada pengembangan pendidikan yang berlandaskan nilai-nilai Islam, dengan penekanan pada pendalaman Al-Qur'an sebagai dasar utama. Selain itu, lembaga ini berfokus pada pembinaan karakter generasi muda agar tidak hanya unggul dalam aspek akademik, tetapi juga mampu menerapkan prinsip-prinsip Qur'ani dalam kehidupan sehari-hari.

Penerimaan peserta didik baru di Yayasan Sahabat Qur'an Al-Karim Jambi menghadapi peningkatan jumlah pendaftar setiap tahun dengan daya tampung yang terbatas sehingga para penguji harus lebih teliti dalam melakukan seleksi penerimaan agar tidak rentang terhadap kesalahan data dan penilaian subjektif yang dapat menimbulkan hasil dari tes menjadi tidak akurat. Selain itu, terdapat sejumlah fitur atau atribut yang digunakan dalam proses seleksi, namun hingga kini belum dapat dipastikan fitur atau atribut mana yang memberikan pengaruh paling signifikan terhadap hasil seleksi.

Klasifikasi adalah salah satu teknik paling umum yang digunakan dalam proses data *mining*. Metode ini digunakan untuk menemukan pola dalam data dan menempatkan objek penelitian ke dalam kelompok tertentu berdasarkan ciri-ciri yang diketahui [2]. Dalam proses klasifikasi, terdapat berbagai metode yang sering digunakan, antara lain *Naive Bayes Classifier*, *Decision Tree, K-Nearest Neighbor, Neural Network*, *Random Forest, Support Vector Machine*, *Logistic Regression*, *Fuzzy Logic*, dan lainnya. Dari berbagai metode yang tersedia, peneliti memilih algoritma Naïve Bayes Classifier dan Decision Tree CART sebagai metode analisis karena keduanya menunjukkan tingkat akurasi yang tinggi.

Dalam rujukan [3], klasifikasi prestasi belajar mahasiswa dengan menerapkan algoritma Naive Bayes, CART, Random Forest, dan suport vector machine menunjukkan bahwa algoritma CART merupakan model terbaik dengan akurai sebesar 89,00%. Dalam rujukan [4] Implementasi algoritma *Naïve Bayes Classifier* pada proses klasifikasi judul skripsi berdasarkan konsentrasi menunjukkan bahwa algoritma ini mampu menghasilkan tingkat akurasi yang cukup baik, yaitu sebesar 80%.

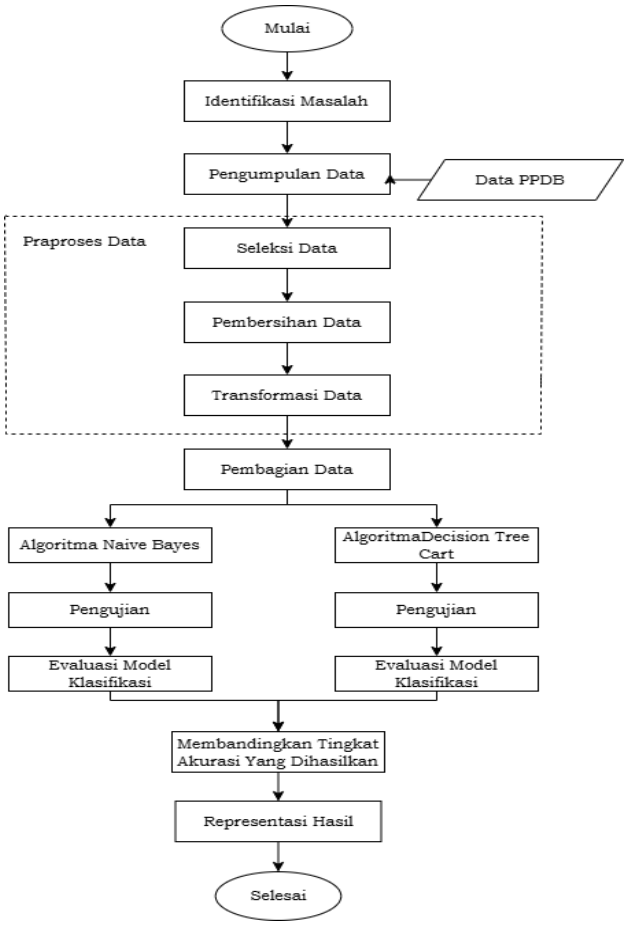
Berdasarkan temuan penelitian yang telah dilakukan oleh peneliti sebelumnya, dapat disimpulkan bahwa algoritma *Naive Bayes* dan *Decision Tree* CART merupakan algoritma yang cukup baik pada proses klasifikasi. Meskipun kedua algoritma tersebut memiliki kemampuan yang serupa, namun performa akurasinya dapat berbeda tergantung pada data yang digunakan. Oleh sebab itu, penulis bermaksud untuk membandingkan kedua algoritma tersebut melalui penerapannya pada data penerimaan peserta didik baru di Yayasan Sahabat Qur'an Al-Karim selama periode 2022 hingga 2024.

Penelitian ini menghasilkan analisis data yang bertujuan untuk mengkaji penerapan algoritma *Naïve Bayes* dan *Decision Tree* CART serta membandingkan akurasi keduanya dalam klasifikasi penerimaan peserta didik baru.

II. METODE PENELITIAN

A. Tahapan Penelitian

Penelitian ini menerapkan pendekatan kuantitatif dengan melakukan perbandingan antara algoritma *Naïve Bayes* dan *Decision Tree* untuk mengklasifikasikan penerimaan siswa baru berdasarkan data historis penerimaan di Yayasan Sahabat Qur'an Al-Karim Jambi. Dalam penelitian ini, Bahasa pemrograman *Python* digunakan untuk melakukan pemrosesan data, analisis, dan pembuatan model *machine learning*. Adapun kerangka dalam pengerjaan penelitian ini yang disusun dalam gambar berikut:



Gambar 1. Kerangka Penelitian

B. Objek dan Variabel Penelitian

Objek yang diangkat dalam penelitian ini adalah data calon peserta didik baru yang diperoleh dari Yayasan Sahabat Qur'an Al-Karim yang berlokasi di Provinsi Jambi.

Variabel atau atribut yang digunakan dalam penelitian ini berasal dari data penerimaan peserta didik baru di Yayasan Sahabat Qur'an Jambi untuk periode tahun 2022, 2023, dan 2024. Data dalam penelitian ini memiliki 6 atribut yang terdiri dari Nama, Jenis Kelamin, Asal Sekolah, Tes Bacaan Qur'an, Tes Kelancaran Hafalan, dan Hasil Wawancara. Sedangkan atribut Status merupakan variabel target yang menunjukkan apakah peserta diterima atau tidak diterima di Yayasan Sahabat Qur'an Al-Karim Jambi berdasarkan keseluruhan hasil seleksi.

C. Praproses Data

Parproses data adalah tahap awal dalam pengolahan yang bertujuan untuk menyiapkan serta membersihkan data sebelum dilakukan analisis. Tahapan ini meliputi proses pemilihan data, pembersihan, dan transformasi data.

Seleksi data adalah tahap pemilihan data yang dianggap relevan dari kumpulan data operasional yang tersedia, di mana data operasional biasanya masih bersifat umum, bervariasi, dan mengandung informasi yang tidak semuanya diperlukan [5].

Pembersihan data merupakan tahap dimana akan dilakukan penghapusan pada data duplikat serta mengidentifikasi dan menangani nilai kosong (*missing value*).

Transformasi data merupakan upaya mengonversi bentuk, struktur, maupun format data mentah ke dalam wujud yang lebih sesuai dan siap diterapkan pada fase pemodelan data [6]. Tahapan yang dilaksanakan pada penelitian ini yaitu mengubah semua data yang bersifat kategorikal menjadi numerik, sehingga data yang digunakan menjadi lebih akurat dan mudah diinterpretasikan oleh algoritma.

D. Algoritma Naïve Bayes

Naïve Bayes Classifier merupakan algoritma klasifikasi yang berpijak pada konsep probabilitas dengan dasar pokok dari Teorema Bayes. Algoritma ini memanfaatkan teknik probabilistik dan statistik yang dibuat oleh Thomas Bayes, seorang ilmuwan asal Inggris untuk meramalkan peluang suatu kejadian dimasa mendatang berdasarkan data atau pengalaman masa silam [7]. Proses pengklasifikasian dengan algoritma *Naïve Bayes* melibatkan dua tahap utama, yaitu tahap pelatihan (*training set*) dan tahap pengujian (*testing set*) [8]. Algoritma *Naïve Bayes* diimplementasikan dengan menggunakan rumus sebagai berikut:

$$P(C|X) = \frac{P(X|C)\times P(C)}{P(X)} \quad (1)$$

Keterangan:

X : prediksi data masukan

C : Hipotesa data masuk pada suatu kelas label

$P(C|X)$: Probabilitas hipotesa berdasarkan kondisi

$P(C)$: Probabilitas hipotesa
$P(X)$: Probabilitas dari *C*

E. Algoritma Decision Tree CART

Decision Tree adalah salah satu metode klasifikasi yang memiliki bentuk menyerupai struktur pohon. Algoritma ini berfungsi untuk membantu dalam proses pengambilan keputusan dengan menyajikan alur logika yang divisualisasikan dalam bentuk cabang dan daun pada pohon keputusan. [9]. Model *Decision Tree* terdiri dari *node* dan cabang, di mana setiap *node* menggambarkan fitur atau atribut dari kategori yang akan diklasifikasikan, sedangkan setiap cabang menunjukkan nilai yang dimiliki oleh *node* tersebut [10].

Algoritma Clasification and Regression Trees (CART) merupakan salah satu metode decision tree yang digunakan untuk melakukan proses klasifikasi dan regresi dengan membangun pohon keputusan biner [11]. Proses penyusunan pohon keputusan pada algoritma CART dilakukan dengan menentukan nilai Gini Index pada setiap kelas sebagai dasar pemisahan data [9]. Bentuk persamaan algoritma CART dinyatakan seperti yang diuraikan berikut:

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2 \quad (2)$$

Nilai P_i diperoleh dengan membandingkan jumlah data pada atribut kelas C_i terhadap keseluruhan jumlah data. Proses pemisahan data menjadi dua subset, yaitu D1 dan D2, dilakukan berdasarkan nilai Gini Index terendah yang dapat dirumuskan sebagai berikut:

$$Gini_A = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (3)$$

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang dimanfaatkan dalam penelitian ini bersumber dari calon murid baru pada tahun 2022, 2023, serta 2024, yang didapat dari Yayasan Sahabat Qur'an Al-Karim Jambi dengan jumlah total mencapai 168 data. Selanjutnya, berikut ditampilkan beberapa sampel data yang dipakai dalam penelitian ini.

Tabel 1. Data Penelitian

Jenis_Kelamin	Asal_Sekolah	Tes_Bacaan_Qur'an	Tes_Kelancaran_Hafalan	Hasil_Wawancara
Perempuan	MIN 1 Merangin	Kurang	Kurang	Tidak Siap
Perempuan	MIN 1 Merangin	Kurang	Kurang	Tidak Siap
Laki-laki	MI 2 Tanjabtim	Kurang	Kurang	Siap
Perempuan	MIN 1 Merangin	Kurang	Kurang	Siap
Perempuan	SDIT Permata Insani Islamic School	Baik	Baik	Siap
Laki-laki	Ponpes Tahfidz Permata Nusantara	Sangat Baik	Baik	Siap

Sumber: Data PPDB Tahun 2022-2024

B. Praproses Data

Data yang diambil sebelumnya merupakan data mentah sehingga diperlukan praproses data sebelum dilakukannya proses klasifikasi agar data menjadi lebih bersih dan hanya fokus pada fitur-fitur yang relevan. Pada penelitian ini, tahap prapemrosesan data mencakup proses seleksi data, pembersihan data, serta transformasi data.

1. Seleksi Data

Proses seleksi data yang dilakukan dalam penelitian ini yaitu menghapus kolom nama yang merupakan data kategorikal dimana setiap nama adalah identitas yang bersifat unik sehingga tidak memberikan kontribusi terhadap proses klasifikasi. Selanjutnya, dilakukan juga pengelompokan jenis asal sekolah yang dikelompokkan menjadi sekolah umum dan sekolah agama. Nama sekolah yang mengandung kata SDN, SD Negeri, SMPN, atau SMAN, maka sekolah tersebut dikategorikan sebagai sekolah umum, namun jika tidak ditemukan kata kunci tersebut, maka sekolah dikategorikan sebagai sekolah agama.

2. Pembersihan Data

Tahap pembersihan data dilaksanakan guna menjamin bahwa himpunan data yang dipakai tidak mengandung inkonsistensi dan telah siap dimanfaatkan dalam proses analisa serta pelatihan model. Dalam penelitian ini baris atau kolom yang mengandung nilai kosong (*missing value*) ditangani dengan cara dihapus. Dari 168 data terdapat 6 data yang memiliki nilai kosong. Setelah dilakukan proses pembersihan yaitu menghapus 6 data yang memiliki nilai kosong, jumlah data efektif yang digunakan dalam penelitian ini adalah 162 data.

3. Transformasi Data

Pada tahap preprocessing, data yang bersifat kategorikal dikonversi menjadi bentuk numerik agar dapat diolah oleh algoritma *Naïve Bayes* dan *Decision Tree*. Proses konversi ini dilaksanakan dengan memanfaatkan *label encoder* yang secara otomatis mengubah setiap nilai kategori unik pada suatu kolom menjadi representasi angka. Berikut merupakan penerapan *label encoder* pada *library scikit-learn.*

Listing 1. Penerapan *Label Encoder*

```
from sklearn.preprocessing import
LabelEncoder

label_encoders = {}
for column in df.columns:
    if df[column].dtype == 'object':
        le = LabelEncoder()
        df[column] =
        le.fit_transform(df[column])
            label_encoders[column] = le
```

Dari gambar diatas, penerapan *label encoder* dilakukan dengan mengubah setiap kategori unik berdasarkan urutan alfabet pada setiap nilai atribut tanpa mengubah makna aslinya. Nilai angka yang akan diterapkan dimulai dari angka 0.

C. Pembagian Data

Setelah tahap praproses data selesai, langkah berikutnya adalah melakukan pembagian data menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*testing set*). Pada penelitian ini, data dibagi dengan proporsi 80% untuk data training dan 20% untuk data

testing. Hasil pembagian ini ditunjukkan pada tabel berikut.

Tabel 2. Hasil Pembagian Data

Jumlah Data	Data *Training*	Data *Testing*
162	129	33

D. Penerapan Algoritma Naïve Bayes

Penggunaan algoritma *Naïve Bayes* didasarkan pada prinsip probabilitas yang mengukur peluang suatu data tergolong ke dalam kelas spesifik sesuai dengan distribusi fitur-fiturnya. Berikut merupakan penerapan algoritma *naive bayes* menggunakan *syntax python*.

Listing 2. Penerapan Algoritma *Naive Bayes*

```
from sklearn.model_selection import
train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score,
classification_report, confusion_matrix

X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2,
random_state=42)
model = GaussianNB()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
print("Akurasi:", accuracy_score(y_test, y_pred))
print("\nConfusion Matrix:\n",
confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n",
classification report(y test, y pred))
```

Algoritma *Naïve Bayes* dikembangkan menggunakan pustaka *scikit-learn* pada bahasa pemrograman *Python*. Berdasarkan hasil penerapan, algoritma ini memperoleh tingkat akurasi sebesar 90,91%, yang mengindikasikan bahwa *Naïve Bayes* mampu melakukan klasifikasi dengan ketepatan yang tinggi. Selanjutnya, dilakukan tahap pengujian dan evaluasi untuk menilai performa model klasifikasi melalui langkah-langkah berikut.

1. Pengujian

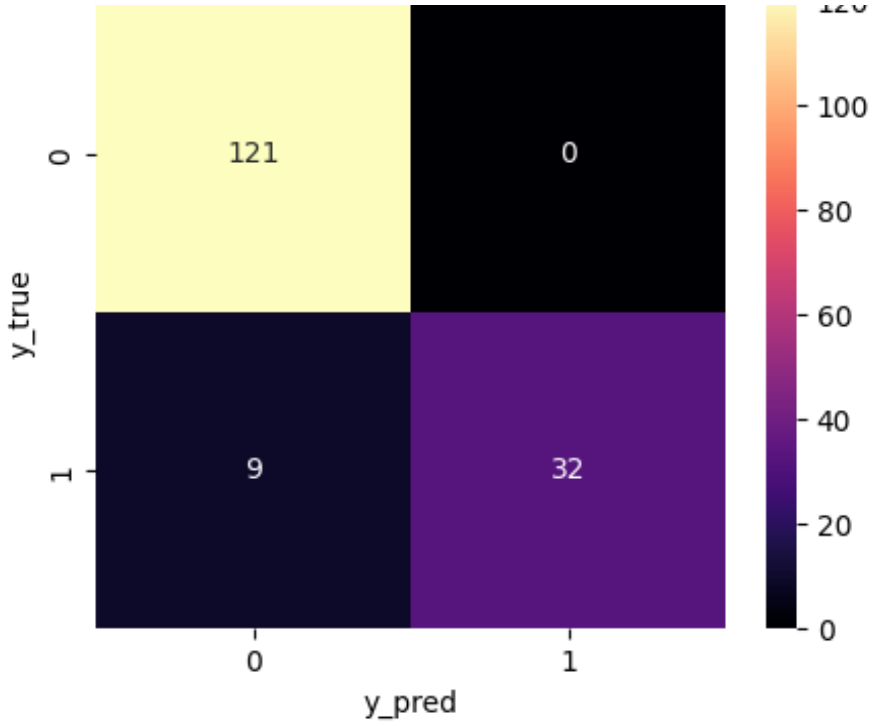
Pengujian terhadap algoritma *Naïve Bayes* dilaksanakan menggunakan metode *cross-validation*. Pada penelitian ini, diterapkan pendekatan *k-fold cross-validation* dengan k=10, sehingga sembilan bagian data dimanfaatkan sebagai data pelatihan dan satu bagian yang tersisa sebagai data pengujian. Berikut disajikan hasil evaluasi menggunakan algoritma *Naïve Bayes*.

Tabel 3. Hasil Pengujian Algoritma *Naive Bayes*

K-Fold	Hasil Akurasi
1	82,35 %
2	100 %
3	87,50 %
4	93,75 %
5	93,75 %
6	93,75 %
7	93,75 %
8	100 %
9	100 %
10	100 %
Rata-rata	**94,49 %**

2. Evaluasi

Setelah pengujian terhadap algoritma *naive bayes* dilakukan, selanjutnya adalah mengevaluasi performa dari algoritma tersebut. Pada penelitian ini, *confusion matrix* digunakan untuk menghitung nilai akurasi, presisi, *recall*, serta *f1-score.* Akurasi adalah metrik evaluasi yang dimanfaatkan untuk mengukur rasio antara jumlah data siswa yang berhasil diprediksi secara akurat oleh model terhadap seluruh data yang terlibat dalam proses. Presisi menggambarkan proporsi antara kasus positif yang diklasifikasikan dengan benar oleh model terhadap total kasus positif yang telah diklasifikasikan. Recall menggambarkan sejauh mana model mampu memprediksi dengan benar data aktual pada kelas positif atau negatif, sedangkan f1-score merepresentasikan keseimbangan antara nilai precision dan recall dalam proses klasifikasi [12]. Berikut merupakan evaluasi hasil klasifikasi menggunakan agoritma *Naive Bayes.*



Gambar 2. *Confussion Matrix* Algoritma *Naive Bayes*

Mengacu pada *confussion matrix* diatas, Hasil akurasi yang didapatkan adalah sebesar 94,44%. Nilai presisi dari data yang diterima dan tidak diterima adalah 93,08% dan 100%, nilai recall dari data yang diterima dan tidak diterima adalah 100% dan 78,05%, dan nilai f1-score dari masing-masing data adalah sebesar 96,41% dan 87,67%.

Dari hasil evaluasi, diketahui bahwa nilai recall yang rendah pada kelas tidak diterima menunjukkan bahwa masih ada sebagian data yang sebenarnya tidak diterima namun terklasifikasi sebagai diterima. Kondisi ini menunjukkan bahwa model lebih sensitif terhadap kelas diterima dibandingkan dengan kelas yang tidak diterima.

E. Penerapan Algoritma Decision Tree CART

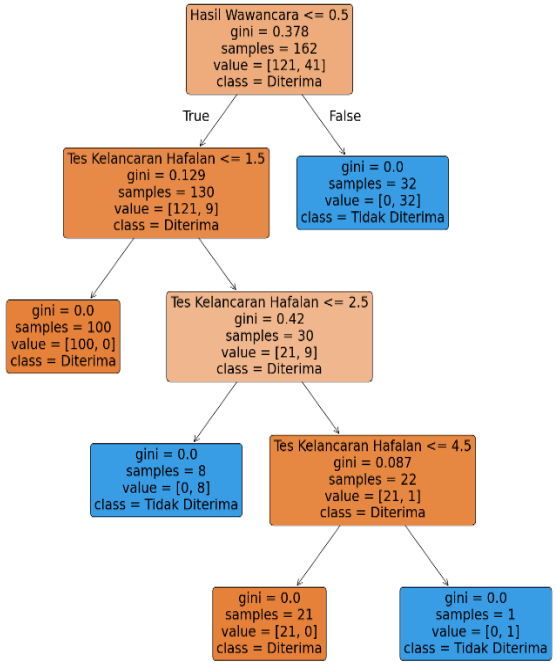
Selain algoritma *Naïve Bayes*, penelitian ini selanjutnya mengimplementasikan algoritma *Decision Tree* dengan pendekatan CART sebagai model klasifikasi. Realisasi algoritma tersebut dilaksanakan melalui sintaks yang disajikan berikut ini.

Listing 3. Penerapan Algoritma *Decision Tree* CART

```
from sklearn.model_selection import
train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score,
classification_report, confusion_matrix

X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2,
random_state=42)
model = DecisionTreeClassifier(criterion='gini',
max_depth=3, random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
print("Akurasi:", accuracy_score(y_test, y_pred))
print("\nConfusion Matrix:\n",
confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n",
classification_report(y_test, y_pred))
```

Berdasarkan listing kode diatas, algoritma CART dibentuk dengan memanfaatkan kelas *DecisionTreeClassifier* dan parameter criterion='gini' untuk menentukan metode pengukuran ketidakmurnian data. Berikut merupakan pohon keputusan dalam menentukan klasifikasi calon peserta didik baru.



Gambar 3. Pohon Keputusan

Mengacu pada gambar di atas, atribut *hasil wawancara* yang dijadikan sebagai akar pada pohon keputusan merupakan faktor yang memiliki pengaruh terbesar dalam proses seleksi penerimaan peserta didik baru di Yayasan Sahabat Qur'an Al-Karim. Atribut lain yang juga berpengaruh dalam penerimaan adalah atribut 'Tes Kelancaran Hafalan' yang dimana atribut ini ditetapkan sebagai *node* dalam pohon keputusan. Sementara itu, atribut yang tidak muncul pada pohon keputusan menandakan bahwa atribut tersebut memiliki tingkat pengaruh yang lebih kecil terhadap hasil klasifikasi, sesuai dengan nilai *gain* yang dimiliki oleh masing-masing atribut.

Setelah pohon keputusan dibentuk, selanjutnya ialah menentukan *rule*. Berikut merupakan aturan-aturan keputusan yang dibuat berdasarkan pohon keputusan.

Tabel 4. Aturan Keputusan

IF Hasil Wawancara = 'Tidak Siap' *THEN* 'Tidak Diterima'
IF Hasil Wawancara = 'Siap' *AND* Tes Kelancaran Hafalan = 'Sangat Baik' *THEN* 'Diterima'
IF Hasil Wawancara = 'Siap' *AND* Tes Kelancaran Hafalan = 'Baik' *THEN* Diterima'
IF Hasil Wawancara = 'Siap' *AND* Tes Kelancaran Hafalan = 'Kurang' *THEN* 'Tidak Diterima'
IF Hasil Wawancara = 'Siap' *AND* Tes Kelancaran Hafalan = 'Cukup' *THEN* 'Diterima'

1. Pengujian

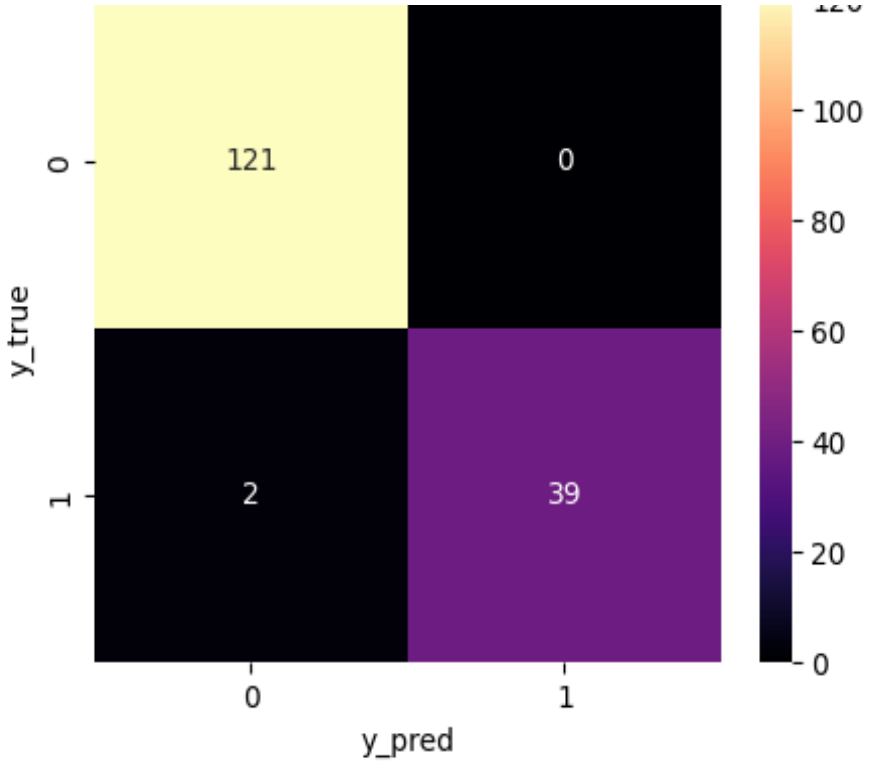
Pengujian terhadap algoritma *Decision Tree* dilaksanakan dengan menerapkan metode cross-validation menggunakan nilai k=10, di mana *dataset* dibagi menjadi 10 bagian, kemudian 9 bagian digunakan sebagai data *training* dan 1 bagian sisanya sebagai data *testing*. Adapun hasil akurasi dari pengujian menggunakan algoritma *Decision Tree* CART disajikan sebagai berikut.

Tabel 5. Hasil Pengujian Algoritma *Decision Tree* CART

K-Fold	Hasil Akurasi
1	94,12 %
2	100 %
3	93,75 %
4	100 %
5	100 %
6	100 %
7	100 %
8	100 %
9	100 %
10	100 %
Rata-rata	**98,79 %**

2. Evaluasi

Hasil evaluasi proses klasifikasi dengan algoritma *Decision Tree* CART ditampilkan pada diagram berikut.



Gambar 4. *Confussion Matrix* Algoritma *Decision Tree* CART

Kinerja algoritma Decision Tree CART dapat dievaluasi melalui metrik akurasi, presisi, recall, serta F1-score yang diperoleh dari *confussion* matriks. Algoritma ini menghasilkan akurasi sebesar 98,77%. Nilai presisi untuk data yang diterima dan tidak diterima masing-masing adalah 98,37% dan 100%, sedangkan nilai *recall* untuk kedua kelas tersebut adalah 100% dan 95,12%. Sementara itu, nilai *f1-score* untuk masing-masing kelas mencapai 99,18% dan 97,50%. Hasil penilaian tersebut menunjukkan bahwa algoritma *Decision Tree* CART berhasil mencapai kinerja optimal untuk kedua kelas, sehingga dianggap sesuai untuk diterapkan dalam penelitian ini.

F. Perbandingan Algoritma Naïve Bayes dan Decision Tree CART

Setelah dilakukan pengujian dan evaluasi terhadap algoritma *Naïve Bayes* dan *Decision Tree* CART, langkah selanjutnya yaitu melakukan perbandingan antara keduanya guna menentukan algoritma yang paling efektif dalam mengklasifikasikan penerimaan peserta didik baru di Yayasan Sahabat Qur'an Al-Karim. Hasil perbandingan tersebut disajikan pada tabel berikut.

Tabel 6. Perbandingan Algoritma Naive Bayes dan Decision Tree

Algoritma	Akurasi	Presisi	Recall	F1-Score
Naive Bayes	94,44%	96,54%	89,02%	92,04%
Decision Tree C4.5	98,77%	99,19%	97,56%	98,34%

Berdasarkan tabel di atas, terlihat bahwa algoritma *Decision Tree* CART menunjukkan nilai akurasi yang lebih tinggi yaitu sebesar 98,77%, presisi 99,19%, recall 97,56%, dan f1-score 98,34%. Sementara itu, algoritma *Naïve Bayes* memperoleh tingkat akurasi sebesar 94,44%, presisi 96,54%, recall 89,02%, dan f1-score 92,04%. Hal ini menunjukkan bahwa algoritma *Decision Tree* CART memiliki kinerja paling unggul dalam mengklasifikasikan penerimaan peserta didik baru di Yayasan Sahabat Qur'an Al-Karim Jambi.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil perbandingan antara kedua algoritma tersebut, dapat disimpulkan bahwa algoritma *Decision Tree* CART menunjukkan keunggulan dibandingkan dengan algoritma *Naïve Bayes*, terbukti dari nilai akurasi dan metrik evaluasi lainnya yang lebih tinggi. Oleh karena itu, algoritma *Decision Tree* CART direkomendasikan sebagai metode yang lebih optimal untuk mendukung proses seleksi penerimaan peserta didik baru di Yayasan Sahabat Qur'an Al-Karim Jambi. Meski demikian, algoritma *Naïve Bayes* tetap layak digunakan karena perbedaan akurasinya dengan *Decision Tree* tidak terlalu signifikan.

B. Saran

Temuan dari penelitian ini, sebaiknya tidak hanya berhenti pada tahap analisis akademis, tetapi juga diimplementasikan secara langsung pada sistem penerimaan peserta didik baru di Yayasan Sahabat Qur'an Al-Karim Jambi. Selain itu, Penelitian berikutnya disarankan untuk menggunakan dataset dengan jumlah yang lebih besar dan bervariasi, sehingga hasil evaluasi yang diperoleh dapat lebih representatif serta mampu menguji konsistensi kinerja algoritma pada berbagai kondisi data.

DAFTAR PUSTAKA

[1] A. Yusnita, S. Lailiyah, and K. Saumahudi, "Penerapan Algoritma Naïve Bayes Untuk Penerimaan Peserta Didik Baru," *J. Inform. Wicida*, vol. 10, no. 1, pp. 11–16, 2021, doi: 10.46984/inf-wcd.1194.

[2] M. H. H. Calderon, E. B. B. Palad, and M. S. Tangkeko, "Filipino Online Scam Data Classification using Decision Tree Algorithms," *2020 Int. Conf. Data Sci. Its Appl. ICoDSA 2020*, 2020, doi: 10.1109/ICoDSA50139.2020.9212929.

[3] A. Rahmadeyan and Mustakim, "Seleksi Fitur pada Supervised Learning: Klasifikasi Prestasi Belajar Mahasiswa Saat dan Pasca Pandemi COVID-19," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 1, pp. 21–32, 2023, doi: 10.25077/teknosi.v9i1.2023.21-32.

[4] S. Dania, R. Ishak, M. Kom, and H. Dalai, "Penerapan Algoritma Naive Bayes Classifier Untuk Klasifikasi Judul Skripsi Berdasarkan Konsentrasi," *J. Ilm. Ilmu Komput. Banthayo Lo Komput.*, vol. 3, no. 1, pp. 15–22, 2024.

[5] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 452–458, 2023, doi: 10.36040/jati.v7i1.6303.

[6] D. K. Lee, "Data transformation: A focus on the interpretation," *Korean J. Anesthesiol.*, vol. 73, no. 6, pp. 503–508, 2020, doi: 10.4097/kja.20137.

[7] A. F. Watratan, A. P. B, and D. Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1 SE-Articles, Jul. 2020, doi: 10.52158/jacost.v1i1.9.

[8] S. Sinaga, R. W. Sembiring, and S. Sumarno, "Penerapan Algoritma Naive Bayes untuk Klasifikasi Prediksi Penerimaan Siswa Baru," *J. Mach. …*, vol. 1, no. 1, pp. 55–64, 2022, [Online]. Available: https://journal.fkpt.org/index.php/malda/article/view/162%0Ahttps://journal.fkpt.org/index.php/malda/article/download/162/115

[9] D. B. Stiawan and Y. S. Nugroho, "Perbandingan Performa Algoritma Decision Tree untuk Klasifikasi Penerima Beasiswa Bank Indonesia," *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2023, [Online]. Available: http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135

[10] W. A. Firmansyach, U. Hayati, and Y. Arie Wijaya, "Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.

[11] F. M. Hana, W. C. Wahyudin, S. Ulya, and D. S. Negara, "IMPLEMENTASI ALGORITMA CART DALAM KLASIFIKASI PENYAKIT DIABETES," *J. Ilmu Komput. dan Matemtika*, pp. 1–8, 2023.

[12] S. Tosun and D. B. Kalaycıoğlu, "Data mining approach for prediction of academic success in open and distance education," *J. Educ. Technol. Online Learn.*, vol. 7, no. 2, pp. 168–176, 2024, doi: 10.31681/jetol.1334687.