

Analisis Komparasi Algoritme C4.5 dan Naïve Bayes dalam Kasus Klasifikasi Kecenderungan Gaya Belajar Visual Auditori Kinestetik (VAK)

Alvindo Tri Jatmiko¹, Wibisono Sukmo Wardhono², Satrio Hadi Wijoyo³

Program Studi Pendidikan Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹alvindotri@student.ub.ac.id, ²wibiwardhono@ub.ac.id, ³satriohadi@ub.ac.id

Abstrak

Kebutuhan data dalam identifikasi gaya belajar merupakan hal yang penting dan berpengaruh dalam menentukan preferensi belajar seseorang. Proses identifikasi gaya belajar menggunakan pendekatan sensorik untuk menggali kebutuhan data yang diperlukan. Salah satu model yang sering digunakan adalah Fleming-VAK (Visual, Auditori, Kinestetik). Data yang dapat diperoleh dan diukur adalah indikator gaya belajar yang dalam masing-masing tipe terdapat 10 indikator, dengan variasi tipe gaya belajar ada 3 yaitu visual, auditori, dan kinestetik. Metode yang digunakan pada saat prediksi gaya belajar secara sederhana menggunakan metode penentuan nilai tertinggi dari salah satu gaya belajar yang dilakukan secara sederhana, sehingga tidak terdapat otomatisasi prediksi gaya belajar dan terdapat masalah ketika dari ketiga gaya belajar terdapat nilai yang sama dari dua atau tiga gaya belajar. Oleh karena itu dilakukan pengujian untuk melakukan komparasi bagaimana hasil yang didapat menggunakan metode komparasi nilai terbesar dari ketiga gaya belajar dengan hasil pengolahan data menggunakan Algoritme C4.5 yang ditujukan untuk klasifikasi. Data dummy digunakan sebagai contoh kasus dalam penelitian, data tersebut dibangkitkan menggunakan software microsoft excel. Hasil pengujian dilakukan menggunakan tahapan preprocessing data, klasifikasi, dan evaluasi. Hasil evaluasi menunjukkan Algoritme Naïve Bayes memiliki nilai akurasi 90% dibandingkan Algoritme C4.5 dengan nilai akurasi 79%. Akurasi didapatkan dari komposisi pembagian data 90% data latih dan 10% data tes. Dengan demikian, penggunaan Algoritme Naïve Bayes memberikan hasil yang lebih akurat dalam prediksi gaya belajar dibandingkan dengan Algoritme C4.5.

Kata kunci: gaya belajar Fleming-VAK, otomatisasi prediksi gaya belajar, data mining, Algoritme C4.5, Algoritme Naïve Bayes

Abstract

The need for data in identifying learning styles is very important and influential in determining individual learning preferences. The process of identifying learning styles uses a sensory approach to extract the required data. One model that is often used is the Fleming-VAK model (Visual, Auditory, Kinesthetic). The data obtained and measured consists of learning style indicators, with each type having 10 indicators. The three variations of learning styles are visual, auditory, and kinesthetic. Previously, a simple learning style determination method was used by selecting the highest score among the less automatic styles and encountered problems when two or three styles had the same score. Therefore a test was carried out to compare the results obtained using the method of comparing the highest scores of the three learning styles by processing the data using the C4.5 Algorithm designed for classification. Data dummy is used as a case study in research produced using Microsoft Excel. The evaluation results show that the Naïve Bayes Algorithm achieves an accuracy rate of 90% compared to the C4.5 Algorithm with an accuracy rate of 79%. Accuracy is obtained by using a data distribution of 90% for training and 10% for testing. As a result, the use of the Naïve Bayes Algorithm provides more accurate results in predicting learning styles compared to the C4.5 Algorithm.

Keywords: Fleming-VAK learning styles, automated prediction of learning styles, data mining, C4.5 Algorithm, Naïve Bayes Algorithm

1. PENDAHULUAN

Gaya belajar adalah cara individu memperoleh pemahaman dengan metodenya sendiri (Rahmi & Samsudi, 2020). Model Fleming-VAK mengategorikan individu ke dalam gaya belajar visual, auditori, dan kinestetik (Fleming, 2001). Gaya belajar visual cenderung menyukai visualisasi dan membaca, serta memahami dengan cepat melalui pengamatan (Lestari & Djuhan, 2021). Individu dengan gaya belajar visual fokus pada penglihatan dan tertarik dengan informasi grafis, video, dan tulisan (Ritonga & Rahma, 2021). Gaya belajar kinestetik memberikan hasil belajar yang baik jika individu dalam kondisi jasmani yang sehat (Irawati, et al., 2021). Gaya belajar dapat diklasifikasikan dengan menggunakan 10 indikator dalam setiap tipe gaya belajar, yaitu visual, auditori, dan kinestetik. Setiap indikator dapat dikelompokkan ke dalam himpunan sangat tidak setuju, tidak setuju, setuju, dan sangat setuju. Himpunan tersebut penting untuk penentuan gaya belajar (Maulid & Arifin, 2022). Prediksi gaya belajar secara sederhana menggunakan metode penentuan nilai tertinggi dari salah satu gaya belajar, sehingga tidak terdapat otomatisasi prediksi gaya belajar dan terdapat masalah ketika dari ketiga gaya belajar terdapat nilai yang sama dari dua atau tiga gaya belajar. Oleh karena itu diperlukan penggunaan algoritme untuk prediksi gaya belajar dan evaluasi kelayakannya.

Dalam bidang pendidikan, data mining digunakan untuk memberikan informasi kepada lembaga pendidikan agar dapat mengambil langkah-langkah yang diperlukan dalam mengatasi masalah siswa (Arcinas, et al., 2021). Metode dalam data mining digunakan untuk memprediksi hasil berdasarkan model data latih yang telah dibuat (Febriani & Sulistiani, 2021). Teknik *supervised learning* membutuhkan dataset yang dilengkapi dengan label, sehingga teknik klasifikasi dan regresi dapat digunakan. Di sisi lain, teknik *unsupervised learning* menggunakan dataset tanpa label, sehingga teknik kluster dapat digunakan (Arcinas, et al., 2021). Dalam kasus prediksi gaya belajar, yang melibatkan data yang bersifat kategorikal, teknik klasifikasi diperlukan untuk melakukan prediksi tipe gaya belajar berdasarkan indikator-indikator gaya belajar (Febriani & Sulistiani, 2021).

Klasifikasi adalah proses mengorganisasi suatu hal tertentu dalam kelompok melalui

model yang telah dibuat. Dataset merupakan sebuah kumpulan data yang masih mentah atau belum terproses, diperlukan tahap *preprocessing* data, klasifikasi dan evaluasi. *Preprocessing* data merupakan kegiatan untuk menghilangkan data *noise*, yaitu data yang memiliki nilai variabel tidak tepat. Klasifikasi merupakan kegiatan untuk melakukan prediksi dari sampel data tes. Evaluasi merupakan kegiatan untuk menguji kemampuan dari algoritme dalam melakukan klasifikasi. Tujuan dari proses klasifikasi adalah untuk melakukan prediksi terhadap suatu data yang tidak memiliki label (Febriani & Sulistiani, 2021). Klasifikasi dapat dilakukan dengan menggunakan Algoritme C4.5 (pohon keputusan) dan *Naïve Bayes*.

Berdasarkan penelitian Badriah, et al. (2021) Algoritme C4.5 digunakan untuk menentukan penerima bantuan covid-19 memiliki akurasi mencapai 79,54%, dan berdasarkan penelitian Yunita & Ikasari (2021) yang membandingkan Algoritme C4.5 dengan *Naïve Bayes* dengan hasil lebih akurat Algoritme C4.5 dengan nilai 94,17%, sedangkan *Naïve Bayes* memiliki akurasi nilai 85,83%. Berdasarkan penelitian Jijo & Abdulazeez (2021) yang membandingkan Algoritme C4.5 dengan algoritme lain seperti KNN (*K-Nearest Neighbor*), LR (*Logistic Regression*), SVM (*Support Vector Machine*), dan NB (*Naïve Bayes*) dalam penerapan pada machine learning didapatkan akurasi Algoritme C4.5 adalah 99,9% dan algoritme yang lain mendapatkan akurasi di bawahnya. Kedua algoritme tersebut tidak dijalankan secara bersama dalam sistem, namun perlu dilakukan komparasi untuk mengetahui algoritme mana yang lebih unggul.

Penerapan algoritme menggunakan sebuah pustaka dari PHP-ML (*PHP Machine Learning*) yang berisikan kumpulan kode dalam bahasa PHP untuk kebutuhan machine learning (Kondas, 2023). PHP (*Hypertext Preprocessor*) adalah bahasa pemrograman *server-side* yang dirancang khusus untuk pengembangan perangkat lunak *web* (PHP, 2023). Data yang digunakan berbentuk CSV (*Comma-Separated Values*). CSV (*Comma-Separated Values*) adalah format berkas yang digunakan untuk menyimpan data dalam bentuk teks dengan pemisah koma.

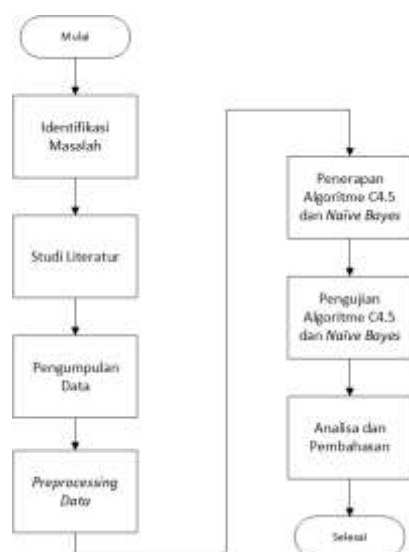
Langkah awal penelitian adalah melakukan identifikasi masalah didapat dari *stakeholder* dari pengembang permainan *Personalized Learning Environment Top Down Version*. Studi literatur dilakukan untuk memahami solusi dari

masalah yang didapat. Pengumpulan data dengan membuat *data dummy*. Dilakukan *preprocessing data* untuk menghilangkan data *noise*. Penerapan algoritme dilakukan menggunakan *data dummy* yang telah diproses. Pengujian dilakukan menggunakan metode *confusion matrix*. Dan tahap terakhir adalah analisis dan pembahasan karakteristik algoritme berdasarkan hasil dari tahap pengujian.

2. METODE PENELITIAN

Metode pengembangan yang digunakan secara linier, metode tersebut dipakai untuk mengaplikasikan Algoritme C4.5 dan Naïve Bayes ke dalam sebuah sistem. Keluaran dari sistem adalah sebuah pesan yang merupakan hasil dari pemrosesan data melalui Algoritme C4.5. Gambar 1 merupakan tahapan penelitian dimulai dari tahap identifikasi masalah, studi literatur, pengumpulan data, *preprocessing data*, penerapan Algoritme C4.5 dan Naïve Bayes, pengujian, dan tahap terakhir adalah analisis dan pembahasan.

Penelitian dilakukan sesuai dengan alur berikut:



Gambar 1. Diagram Alir Metodologi Penelitian

2.1. Identifikasi Masalah

Identifikasi masalah didapat dari *stakeholder* pengembang permainan *Personalized Learning Environment Top Down Version*. Masalah yang dihadapi adalah bagaimana cara melakukan klasifikasi gaya belajar secara otomatis dilakukan oleh sistem. Masalah tersebut dirumuskan dalam rumusan masalah dan diselesaikan dengan tujuan dari penelitian, yaitu alat bantu berupa algoritme

yang dapat melakukan prediksi gaya belajar berdasarkan 30 indikator gaya belajar dan menampilkan hasil klasifikasi dalam bentuk teks dengan cara pengguna mengunggah berkas dalam format CSV.

2.2. Studi Literatur

Studi literatur ditujukan untuk memahami karakteristik dari Algoritme C4.5 dan *Naïve Bayes* apakah cocok untuk digunakan dalam melakukan prediksi terhadap data yang memiliki kesamaan karakteristik. Studi literatur dipilih berdasarkan tingkat evaluasi terhadap algoritme yang baik. Studi literatur ini memberikan pemahaman yang lebih mendalam tentang bagaimana Algoritme C4.5 dan *Naïve Bayes* dapat diterapkan secara efektif dalam prediksi data.

2.3. Pengumpulan Data

Pengumpulan data digunakan untuk melakukan analisis komparasi performa dari algoritme dan mendapatkan informasi yang relevan dan akurat. Metode pengumpulan data berdasarkan tahap studi literatur adalah dengan membangkitkan data berdasarkan indikator gaya belajar. Dengan analisis yang cukup data tersebut digunakan sebagai contoh kasus dalam proses klasifikasi algoritme. Data penelitian yang digunakan merupakan data dummy, yaitu data yang dibuat secara acak dan digunakan. Data baru didapat dari *database* yang ada pada permainan *Personalized Learning Environment Top Down Version* sebagai contoh untuk melakukan prediksi data baru dalam format berkas CSV. Pengumpulan *data dummy* menggunakan software *microsoft excel* untuk melakukan pembentukan data dengan 30 indikator gaya belajar pada setiap baris dengan 10 indikator pertama adalah gaya belajar visual, 10 indikator berikutnya adalah auditori, dan 10 terakhir adalah kinestetik. Pengumpulan data menggunakan fungsi *random* pada excel antara angka 1 sampai 4. Diperlukan atribut tambahan yaitu visual, auditori, dan kinestetik untuk menampung nilai total dari masing-masing indikator. Setelah nilai total dari keseluruhan atribut visual, auditori, dan kinestetik terpenuhi kemudian dilakukan perbandingan nilai terbesar dari ketiga atribut tersebut. Atribut gaya belajar diperlukan untuk menampung hasil dari nilai terbesar pada ketiga atribut tersebut dan mengubahnya menjadi label yang sesuai dengan nilai terbesarnya (Hamu, 2022).

2.4. Preprocessing Data

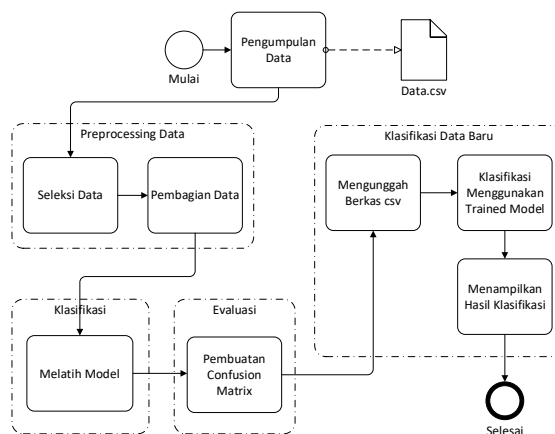
Preprocessing data adalah proses membersihkan, mentransformasi, dan memformat data mentah agar dapat digunakan untuk analisis. Tujuan dari preprocessing data adalah membuat data lebih konsisten, akurat, dan lengkap. Tahap yang dilakukan adalah dengan melakukan seleksi data dan pembagian data (Badriah, et al., 2021). Seleksi data dilakukan untuk memilih atribut yang digunakan dalam pengembangan algoritme.

Seleksi pertama adalah dengan menghapus baris data yang memiliki gaya belajar berlabel “*unknown*”, yaitu ketika penentuan nilai tertinggi dari ketiga tipe gaya belajar tidak memiliki nilai yang dominan atau terdapat nilai yang sama dari dua tipe gaya belajar. Seleksi kedua adalah dengan memilih atribut yang berguna yaitu atribut indikator gaya belajar dan hasil klasifikasinya dari metode sederhana. Pembagian data dilakukan dengan mengubah data menjadi data latih dan data uji, untuk kemudian dilakukan pembentukan model algoritme dan pengujian dari performa Algoritme C4.5 dan *Naïve Bayes*.

2.5. Penerapan Algoritme C4.5 dan Naïve Bayes

Dalam penerapan Algoritme C4.5 dan *Naïve Bayes* pada sistem menggunakan bahasa pemrograman PHP, terdapat beberapa tahapan yang harus dilakukan. Tahapan awal adalah melakukan pengolahan data untuk menghasilkan data latih dan data uji yang nantinya akan digunakan dalam algoritme. Setelah itu, pembuatan training model dilakukan dengan memanfaatkan data latih yang telah diolah sebelumnya. Selanjutnya, model yang dihasilkan akan diuji menggunakan data uji untuk mengevaluasi performa model. Tahapan terakhir adalah penerapan Algoritme C4.5 dan *Naïve Bayes* (Badriah, et al., 2021).

Penerapan algoritme dalam bahasa php menggunakan tahapan dari diagram blok berikut:



Gambar 2. Diagram Blok Penerapan Algoritme

Gambar 2 merupakan diagram alir penerapan Algoritme C4.5 dan *Naïve Bayes*. Diagram tersebut digunakan sebagai tahapan dalam pengkodean pada web menggunakan bahasa PHP. Diagram tersebut juga menjelaskan proses dari preprocessing data, klasifikasi, evaluasi, dan klasifikasi data baru. Tahap *preprocessing* dan klasifikasi menggunakan pustaka dari PHP-ML.

Tahap pre-processing dilakukan dengan seleksi atribut yang digunakan, menghilangkan data yang tidak dibutuhkan, dan membagi data menjadi data latih dan data uji. Tahap klasifikasi dilakukan dengan cara melatih model yang dibuat oleh algoritme. Tahap evaluasi dilakukan dengan membuat tabel confusion matrix. Dan tahap klasifikasi baru dilakukan dengan cara mengunggah berkas baru yang tidak memiliki label dengan kolom 1 sampai 30, kemudian dilakukan klasifikasi terhadap model yang telah dibuat dan menampilkan indikator serta prediksi gaya belajar dalam sistem.

Menurut Quinlan (1993) Algoritme C4.5 adalah model prediksi yang digunakan untuk melakukan prediksi suatu nilai menggunakan model pohon keputusan, dalam penerapan algoritme tersebut diperlukan beberapa langkah yaitu:

1. Menyiapkan data latih yang diambil dari yang pernah terjadi sebelumnya atau data yang sudah dikelompokkan.
2. Penentuan akar dari pohon dilakukan dengan menghitung nilai *gain* dari setiap atribut yang dipilih, yang memiliki nilai *gain* terbesar akan menjadi akar dari pohon tersebut, nilai *gain* memerlukan nilai *entropy* agar dapat dihitung. Untuk menghitung nilai *entropy* dapat dilakukan dengan persamaan 1 berikut:

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

Keterangan:

S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi S_i terhadap S

- Melakukan perhitungan nilai gain menggunakan persamaan 2.

$$Gain(S,A) = E(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} E(S_i) \quad (2)$$

Keterangan:

E = Entropy

S = Himpunan kasus

A = Fitur

n = Jumlah pembagian atribut A

$|S_i|$ = Proporsi S_i terhadap S

$|S|$ = Jumlah kasus dalam S

- Lakukan langkah 2 dan 3 hingga objek terakhir terbagi.
- Pembagian pohon keputusan dihentikan apabila:
 - Seluruh objek dalam simpul N mendapat kelas yang sama
 - Tidak terdapat atribut pada objek yang dibagi.

Objek didalam cabang tidak boleh kosong.

Menurut Quinlan (1993) kelemahan dari Algoritme C4.5 adalah sensitivitas terhadap dataset yang memiliki banyak atribut dan nilai. Hal ini dapat menyulitkan dalam pemilihan atribut yang paling informatif dan mempengaruhi kualitas pohon keputusan yang dihasilkan. Meskipun demikian, Algoritme C4.5 tetap menjadi salah satu algoritme yang populer dan sering digunakan dalam *machine learning*.

Menurut Bayes (1763) Algoritme Naïve Bayes menggunakan prinsip probabilitas untuk melakukan klasifikasi. Prinsip dasarnya adalah bahwa kelas dari suatu objek atau data dapat ditentukan berdasarkan probabilitas kemunculan atribut-atribut pada kelas tersebut. Metode ini bergantung pada perhitungan probabilitas kondisional dan probabilitas bersyarat.

Naïve Bayes dapat dijelaskan dengan menggunakan persamaan 3 berikut:

$$P(\text{Kelas}|\text{Atribut}) = \frac{P(\text{Atribut}|\text{Kelas}) \times P(\text{Kelas})}{P(\text{Atribut})} \quad (3)$$

Keterangan:

$P(\text{Kelas}|\text{Atribut})$ adalah probabilitas kelas yang diberikan atribut.

$P(\text{Atribut}|\text{Kelas})$ adalah probabilitas atribut yang diberikan kelas.

$P(\text{Kelas})$ adalah probabilitas prior dari kelas.

$P(\text{Atribut})$ adalah probabilitas dari atribut.

Menurut Bayes (1763) asumsi independensi fitur sering tidak realistis, sehingga kinerjanya kurang maksimal. Algoritme Naïve Bayes secara keseluruhan tetap menjadi metode klasifikasi yang sederhana, namun efektif dalam banyak kasus. Ketika asumsi independensi fitur diterima algoritme tersebut memberikan hasil yang baik.

2.6. Pengujian Algoritme C4.5 dan Naïve Bayes

Pengujian Algoritme C4.5 dan *Naïve Bayes* merupakan tahap untuk melakukan pengujian Algoritme C4.5 dan *Naïve Bayes*, pengujian dilakukan dengan membuat tabel *confusion matrix*. *Confusion Matrix* adalah tabel yang digunakan untuk mengevaluasi model yang dibuat oleh algoritme (Azhari, et al., 2021). Pengujian dimulai dengan menggunakan data uji untuk mengevaluasi performa model yang dihasilkan oleh Algoritme C4.5. Selanjutnya, dilakukan perhitungan *confusion matrix* berdasarkan hasil prediksi model terhadap data uji dengan membandingkan dengan nilai sebenarnya. Dari *confusion matrix*, dapat dihitung nilai akurasi, presisi, recall, dan F1-score untuk mengevaluasi performa model secara lebih terperinci.

Evaluasi dari algoritme dibuat dalam bentuk tabel berikut:

Tabel 1. Tabel *Confusion Matrix*

Confusion Matrix		
Data Aktual	Data Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan:

TP (*True Positive*) adalah jumlah sampel yang secara benar diprediksi sebagai positif oleh model.

FN (*False Negative*) adalah jumlah sampel yang seharusnya diprediksi sebagai positif, tetapi

secara salah diprediksi sebagai negatif oleh model.

FP (*False Positive*) adalah jumlah sampel yang seharusnya diprediksi sebagai negatif, tetapi secara salah diprediksi sebagai positif oleh model.

TN (*True Negative*) adalah jumlah sampel yang secara benar diprediksi sebagai negatif oleh model.

Confusion Matrix diperlukan dalam evaluasi kinerja sebuah algoritme. Matriks tersebut berguna untuk menentukan seberapa baik model yang dibuat oleh algoritme. Tabel 1 merupakan sebuah matriks yang dapat digunakan untuk mendapatkan nilai akurasi, presisi, *recall*, dan *f1-score*.

Nilai akurasi didapat melalui perhitungan berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (4)$$

Persamaan 4 merupakan rumus untuk mencari nilai dari akurasi berdasarkan *confusion matrix*. Persamaan tersebut melakukan perhitungan terhadap keseluruhan hasil kelas prediksi dengan kelas aktual. Nilai akurasi digunakan untuk mengukur seberapa baik algoritme dalam melakukan klasifikasi, yaitu seberapa baik algoritme melakukan prediksi benar.

Nilai presisi didapat melalui perhitungan berikut:

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

Persamaan 5 merupakan rumus untuk mencari nilai dari presisi berdasarkan tabel *confusion matrix*. Persamaan tersebut melakukan perhitungan terhadap setiap tipe gaya belajar. Nilai presisi digunakan untuk mengukur seberapa baik algoritme dapat menentukan gaya belajar yang sebenarnya, dikarenakan dapat meminimalkan kesalahan prediksi positif palsu, yaitu ketika algoritme melakukan klasifikasi positif namun sebenarnya adalah negatif.

Nilai *recall* didapat melalui perhitungan berikut:

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

Persamaan 6 merupakan rumus untuk mencari nilai dari *recall* berdasarkan tabel *confusion matrix*. Persamaan tersebut melakukan perhitungan terhadap setiap tipe gaya belajar. Nilai *recall* digunakan untuk mengukur

seberapa baik algoritme mampu menemukan semua sampel positif yang sebenarnya.

Nilai *f1-score* didapat melalui perhitungan berikut:

$$F1 - Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \times 100\% \quad (7)$$

Persamaan 7 merupakan rumus untuk mencari nilai dari *f1-score* berdasarkan tabel *confusion matrix*. Persamaan tersebut melakukan perhitungan terhadap setiap tipe gaya belajar. Nilai *f1-score* digunakan untuk mengukur seberapa baik keseimbangan nilai presisi dan *recall* pada algoritme.

2.7. Analisis dan Pembahasan

Analisis dan pembahasan merupakan tahap untuk melakukan pengujian Algoritme C4.5 dan *Naïve Bayes* pada sistem menggunakan metode *confusion matrix*, tahap berikutnya adalah analisis dan pembahasan hasil pengujian. Pada tahap analisis, dilakukan evaluasi terhadap nilai akurasi, presisi, *recall*, dan *f1-score* untuk mengevaluasi performa model. Selanjutnya, dilakukan pembahasan terhadap karakteristik Algoritme C4.5 dan *Naïve Bayes*. Jika terdapat perbedaan nilai akurasi, presisi, *recall*, dan *f1-score* antara kedua algoritme, maka dilakukan analisis penyebab berdasarkan karakteristiknya.

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan penelitian didasarkan pada metodologi yang digunakan. Hasil penelitian mencakup implementasi algoritme dalam bentuk kode serta pengujian yang dilakukan dengan menggunakan *confusion matrix* pada berbagai pembagian data. Pembahasan penelitian berfokus pada karakteristik algoritme yang dianalisis berdasarkan nilai akurasi, presisi, *recall*, dan *f1-score* yang dihasilkan.

3.1. Deskripsi Data

Data penelitian merupakan *data dummy* yang dibuat secara acak sebanyak 500 baris dengan variabel sesuai dengan indikator gaya belajar untuk kemudian digunakan sebagai data latih pada Algoritme C4.5 dan *Naïve Bayes*. Variabel gaya belajar diperoleh dengan prediksi menggunakan metode komparasi nilai terbesar dari variabel visual, auditori, dan kinestetik.

Apabila indikator memiliki nilai 1 memiliki arti sangat tidak setuju, nilai 2 memiliki arti tidak setuju, nilai 3 memiliki arti setuju, dan nilai 4

memiliki arti sangat setuju. Dibutuhkan variabel lain untuk menentukan gaya belajar secara sederhana dari indikator gaya belajar, yaitu variabel gaya belajar, visual, auditori, dan kinestetik. Variabel visual digunakan untuk menampung hasil dari penjumlahan variabel VIS1 sampai VIS10. Variabel auditori digunakan untuk menampung hasil penjumlahan variabel AUD1 sampai AUD10. Variabel kinestetik digunakan untuk menampung hasil penjumlahan variabel KIN1 sampai KIN10. Variabel gaya belajar digunakan untuk melakukan seleksi dari nilai tertinggi dari ketiga tipe gaya belajar, apabila terdapat dua tipe yang memiliki nilai yang sama maka label dari variabel gaya belajar adalah *unknown* (tidak diketahui).

3.2. Preprocessing Data

Preprocessing data dilakukan menggunakan pustaka dari PHP-ML, dalam pustaka tersebut tidak dicantumkan fitur seleksi data, sehingga perlu ditambahkan fitur untuk seleksi gaya belajar dengan label “*unknown*”. Label tersebut diabaikan dan tidak dimasukkan ke dalam pemrosesan data. Pembagian data menggunakan kelas dari pustaka PHP-ML dengan menggunakan fitur *randomsplit*.

Berdasarkan data yang dikumpulkan terdapat data yang tidak sesuai dengan kebutuhan dan menghasilkan nilai yang seimbang atau tidak memiliki satu dominan gaya belajar. Tabel 2 merupakan cuplikan dari sumber data yang memiliki gaya belajar seimbang. Data yang memiliki gaya belajar seimbang diberikan label “*unknown*” (tidak diketahui). Data tersebut dihilangkan untuk menghindari hasil yang buruk dari pohon keputusan.

Contoh seleksi data tahap pertama adalah sebagai berikut:

Tabel 2. Cuplikan Data Yang Tidak Sesuai Kebutuhan

Gaya Belajar	visual	auditori	kinestetik
visual	29	17	20
kinestetik	19	23	24
unknown	29	17	29
visual	33	26	25
visual	32	23	22
auditori	23	29	24
unknown	25	23	25

Tabel 3 menjelaskan tentang cuplikan atribut yang dipakai dalam penelitian untuk mengembangkan model algoritme. Atribut tersebut digunakan sebagai masukan untuk melakukan klasifikasi. Atribut pada kolom

visual, auditori, dan kinestetik dihilangkan untuk kemudian dibuat dalam format *comma-separated value*.

Contoh seleksi data tahap kedua adalah sebagai berikut:

Tabel 3. Cuplikan Pemilihan Atribut Klasifikasi Gaya Belajar

KIN1	KIN2	KIN3	KIN4	KIN5	KIN6	KIN7	KIN8	KIN9	KIN10	Gaya Belajar
2	1	2	4	1	3	2	2	2	1	visual
1	2	4	1	3	3	1	3	3	3	kinestetik
3	2	4	3	2	4	2	4	3	2	unknown
2	1	2	4	3	4	2	2	4	1	visual
2	1	1	4	2	2	2	2	2	4	visual
1	2	2	3	2	1	4	4	1	4	auditori
4	1	1	2	4	4	4	1	1	3	unknown

Data penelitian dibagi menjadi data latih dan data uji ditujukan untuk mengurangi kesalahan algoritme ketika model yang dibuat terlalu rumit, sehingga sulit untuk memberikan hasil yang akurat. Tabel 4 merupakan komposisi dari berapa persen data sekunder yang dipakai untuk mengembangkan pohon keputusan dan berapa persen data primer yang digunakan untuk menguji akurasi dari hasil klasifikasi. Komposisi berikut yang kemudian diujikan pada tahap pengujian pohon keputusan.

Pengujian dari performa algoritme menggunakan tabel berikut:

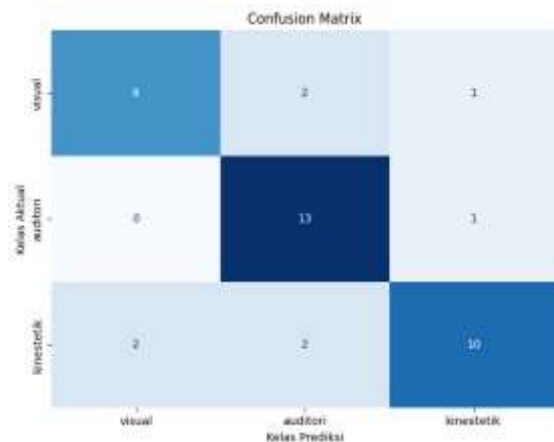
Tabel 4. Komposisi pembagian data sekunder

Data Latih	Data Uji
60%	40%
70%	30%
80%	20%
90%	10%

3.3. Hasil Penerapan Algoritme C4.5

Hasil penerapan ditampilkan dalam bentuk teks dari tipe gaya belajar, sehingga hasil pembentukan pohon keputusan tidak perlu ditampilkan dalam sistem. Gambar 3 merupakan hasil dari evaluasi Algoritme C4.5 dengan mencari akurasi menggunakan metode *confusion matrix*. Bentuk dari *confusion matrix* dibuat dengan memetakan hasil dari kelas prediksi yang benar terhadap kelas aktual. Cuplikan tersebut merupakan hasil dari evaluasi dengan pembagian data 90% untuk data latih dan 10% untuk data tes.

Pengujian menggunakan confusion matrix dari Algoritme C4.5 sebagai berikut:

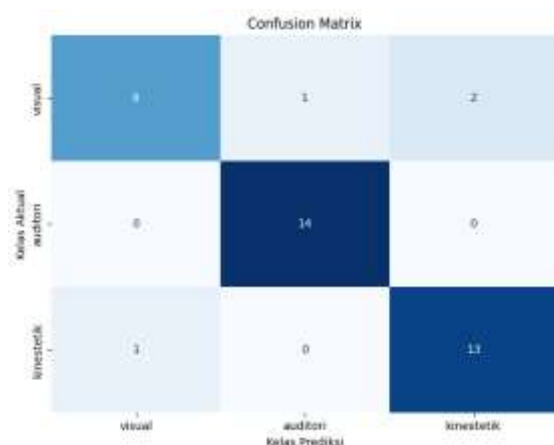


Gambar 3. Cuplikan Hasil Evaluasi Algoritme C4.5

3.4. Hasil Penerapan Algoritme Naïve Bayes

Hasil dari penerapan Algoritme Naïve Bayes sama dengan penerapan Algoritme C4.5 dilakukan dengan kondisi yang sama, sehingga dapat dilakukan analisis komparasi akurasi, presisi, *recall*, dan *f1-score* dari kedua algoritme. Gambar 6 merupakan hasil dari evaluasi Algoritme Naïve Bayes dengan memetakan seberapa banyak data prediksi yang sesuai dengan data aktual. Hasil dibuat dengan menggunakan metode *confusion matrix*. Cuplikan tersebut merupakan hasil dari evaluasi dengan pembagian data 90% untuk data latih dan 10% untuk data tes.

Pengujian menggunakan confusion matrix dari Algoritme Naïve Bayes sebagai berikut:



Gambar 4. Cuplikan Hasil Evaluasi Algoritme Naïve Bayes

3.5. Pengujian Confusion Matrix

Berdasarkan hasil penerapan Algoritme

C4.5 dan *Naïve Bayes* didapatkan nilai akurasi, presisi, *recall*, dan *f1-score* dari *confusion matrix*. Hasil tersebut didapatkan dengan membandingkan kelas prediksi dan kelas aktual. Kelas prediksi merupakan data yang didapatkan setelah melakukan klasifikasi dengan algoritme. Kelas aktual merupakan data yang didapatkan dengan mencari nilai terbesar dari keseluruhan jumlah nilai dari masing-masing gaya belajar.

Komparasi hasil akurasi dari Algoritme C4.5 dan *Naïve Bayes* sebagai berikut:

Tabel 5. Hasil pengujian akurasi algoritme

Komposisi Pembagian Data		Akurasi	
Data Latih	Data Uji	Algoritme C4.5	Algoritme Naïve Bayes
60%	40%	66%	84%
70%	30%	65%	87%
80%	20%	72%	85%
90%	10%	79%	90%
Rata-rata hasil:		71%	86%

Tabel 5 merupakan hasil pengujian berdasarkan komposisi pembagian data pada tabel 4 yang menjelaskan hasil akurasi yang berbeda. Setelah dilakukan pengukuran rata-rata dari hasil akurasi pada tiap komposisi pembagian data, didapatkan hasil berupa Algoritme Naïve Bayes memiliki akurasi sebesar 86% dan Algoritme C4.5 memiliki akurasi sebesar 71%. Terjadi peningkatan akurasi dari Algoritme C4.5 dari 66% hingga 79% dan Algoritme Naïve Bayes yang cenderung stabil diatas 80%. Hasil tersebut menunjukkan bahwa akurasi dari algoritme dipengaruhi oleh banyaknya data latih yang digunakan untuk membuat model prediksi.

Hasil pengujian presisi berdasarkan komposisi pembagian pada tabel 4. Setelah dilakukan pengukuran rata-rata dari setiap tipe gaya belajar, didapatkan bahwa Algoritme C4.5 memiliki presisi pada tipe gaya belajar visual 73%, auditori 75%, dan kinestetik 66%. Algoritme *Naïve Bayes* memiliki presisi pada tipe gaya belajar visual 85%, auditori 85%, dan kinestetik 90%. Algoritme C4.5 pada setiap komposisi pembagian data memiliki presisi yang tidak stabil untuk setiap tipe gaya belajarnya, sedangkan Algoritme *Naïve Bayes* pada setiap komposisi pembagian data memiliki presisi yang stabil diatas 80% untuk setiap tipe gaya belajarnya. Hasil tersebut menunjukkan bahwa Algoritme C4.5 memiliki kesalahan prediksi palsu yang tinggi dikarenakan memiliki tingkat presisi yang rendah.

Hasil pengujian *recall* berdasarkan komposisi pembagian pada tabel 4. Setelah dilakukan pengukuran rata-rata dari setiap tipe

gaya belajar, didapatkan bahwa Algoritme C4.5 memiliki *recall* pada tipe gaya belajar visual 62%, auditori 70%, dan kinestetik 79%. Algoritme *Naïve Bayes* memiliki *recall* pada tipe gaya belajar visual 89%, auditori 83%, dan kinestetik 87%. Algoritme C4.5 pada setiap komposisi pembagian data memiliki *recall* yang tidak stabil untuk setiap tipe gaya belajarnya, sedangkan Algoritme *Naïve Bayes* pada setiap komposisi pembagian data memiliki presisi yang stabil diatas 80% untuk setiap tipe gaya belajarnya. Algoritme C4.5 tidak dapat mendeteksi kelas positif dengan baik dikarenakan nilai *recall* yang rendah.

Hasil pengujian *f1-score* berdasarkan komposisi pembagian pada tabel 4. Setelah dilakukan pengukuran rata-rata dari setiap tipe gaya belajar, didapatkan bahwa Algoritme C4.5 memiliki *f1-score* pada tipe gaya belajar visual 68%, auditori 67%, dan kinestetik 75%. Algoritme *Naïve Bayes* memiliki *recall* pada tipe gaya belajar visual 85%, auditori 85%, dan kinestetik 90%. Algoritme C4.5 pada setiap komposisi pembagian data memiliki *f1-score* yang tidak stabil untuk setiap tipe gaya belajarnya, sedangkan Algoritme *Naïve Bayes* pada setiap komposisi pembagian data memiliki presisi yang stabil diatas 80% untuk setiap tipe gaya belajarnya. Algoritme C4.5 memiliki tingkat keseimbangan rendah dikarenakan nilai presisi dan *recall* yang tidak seimbang.

3.6. Analisis dan Pembahasan

Evaluasi nilai akurasi, presisi, *recall*, dan *f1-score* digunakan sebagai pembandingan dari Algoritme C4.5 dan *Naïve Bayes*. Nilai akurasi digunakan untuk mengukur seberapa baik algoritme dalam melakukan prediksi, namun diperlukan nilai presisi dan *recall* untuk mengetahui apakah terdapat kesalahan prediksi palsu yang dilakukan oleh algoritme. Apabila algoritme memiliki nilai presisi yang tinggi, namun nilai *recall* yang rendah menandakan bahwa banyak sampel positif yang tidak terdeteksi. Apabila algoritme memiliki nilai *recall* yang tinggi, namun nilai presisi yang rendah menandakan bahwa banyak sampel positif yang salah diprediksi. Nilai *f1-score* digunakan untuk menjadikan nilai dari presisi dan *recall* menjadi tunggal, sehingga untuk menentukan kehandalan dari setiap kelas pada algoritme dapat menggunakan nilai tersebut.

Algoritme *Naïve Bayes* memiliki akurasi sebesar 79% memiliki arti bahwa algoritme

memiliki kebenaran dalam prediksi mencapai 79% benar dan memiliki tingkat kesalahan yang sebesar 21%. Berdasarkan karakteristik Algoritme C4.5 faktor yang menjadikan akurasinya menjadi rendah adalah ketika model pohon keputusan yang dibuat memiliki kedalaman yang kurang, sehingga algoritme tidak dapat dengan baik melakukan prediksi terhadap data baru. Hal tersebut ditunjukkan oleh peningkatan komposisi data latih yang seimbang dengan naiknya dari akurasi yang dihasilkan.

Algoritme *Naïve Bayes* memiliki akurasi sebesar 90% memiliki arti bahwa algoritme memiliki kebenaran dalam prediksi mencapai 90% benar dan memiliki tingkat kesalahan yang sebesar 10%. Berdasarkan karakteristik Algoritme *Naïve Bayes* faktor yang menjadikan akurasinya tinggi adalah asumsi independensi fitur, dikarenakan Algoritme *Naïve Bayes* menghitung probabilitas setiap fitur. Apabila fitur tersebut memiliki nilai yang sama secara banyak, maka akan meningkatkan probabilitas dari fitur tersebut.

4. KESIMPULAN

1. Kebutuhan untuk melakukan klasifikasi gaya belajar secara otomatis terselesaikan dengan penggunaan Algoritme *Naïve Bayes* dalam melakukan prediksi gaya belajar, algoritme tersebut dibuat memakai pustaka PHP-ML sehingga dapat dijalankan pada situs secara daring.
2. Perbandingan hasil penggunaan algoritme sebagai alat bantu daripada menggunakan metode penentuan nilai tertinggi dari salah satu gaya belajar memiliki tingkat akurasi dan efisiensi yang tinggi, serta meminimalkan kesalahan prediksi dalam prosesnya, sehingga algoritme layak dipakai untuk melakukan klasifikasi.
3. Hasil pengujian menggunakan komposisi pembagian data yang dipakai adalah ketika performa dari kedua algoritme memiliki nilai akurasi yang paling tinggi, yaitu pada komposisi 90% data latih dan 10% data tes. Perbandingan akurasi antara Algoritme C4.5 dan *Naïve Bayes* adalah 79%:90%. Nilai akurasi digunakan untuk mengetahui tingkat kebenaran dari algoritme ketika melakukan prediksi. Pengujian presisi dan *recall* menunjukkan Algoritme *Naïve Bayes* lebih unggul daripada Algoritme

C4.5. Nilai presisi dan *recall* digunakan untuk mengetahui tingkat kesalahan prediksi palsu. Pengujian *f1-score* menunjukkan Algoritme Naïve Bayes lebih unggul dari Algoritme C4.5. Nilai *f1-score* digunakan untuk mengetahui ketidakseimbangan antara kelas positif dan negatif dalam dataset.

5. DAFTAR PUSTAKA

- Arcinas, M. M. et al., 2021. Role of data mining in education for improving students performance for social change. *Turkish Journal of Physiotherapy and Rehabilitation*, 32(3), pp. 6519-6526.
- Azhari, M., Situmorang, Z. & Rosnelly, R., 2021. Komparasi Akurasi, Recall, dan Presisi Klasifikasi pada Algoritme C4.5, Random Forest, SVM dan Naive Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), pp. 640-651.
- Badriah, S. et al., 2021. Klasifikasi Algoritma C4.5 dalam Menentukan Penerima Bantuan Covid-19 (Studi Kasus: Desa di Karawang). *Jurnal Informatika Polinema*, 7(3), pp. 23-28.
- Bayes, T., 1763. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*, (53), pp. 370-418.
- Febriani, S. & Sulistiani, H., 2021. Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritme C4.5. *Jurnal Teknologi dan Sistem Informasi*, 2(4), pp. 89-96.
- Fleming, N. D., 2001. *Teaching and Learning Styles: VARK Strategies*. Christchurch: Neil Fleming.
- Hamu, M. A., 2022. Hasil Belajar Siswa Ditinjau Dari Gaya Belajar. *AKSARA: Jurnal Ilmu Pendidikan Nonformal*, 8(1), pp. 547-562.
- Irawati, I., Nasruddin & Ilhamdi, M. L., 2021. Pengaruh Gaya Belajar Terhadap Hasil Belajar IPA. *J. Pijar MIPA*, 16(1), pp. 44-48.
- Jijo, B. T. & Abdulazeez, A. M., 2021. Classification Based on Decision Tree Algorithm for Machine Learning. *JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS*, 2(01), pp. 20-28.
- Kondas, A., 2023. Home - PHP-ML - Machine Learning library for PHP. [Online] Available at: <https://php-ml.readthedocs.io/en/latest/> [Accessed 15 Juni 2023].
- Lestari, S. & Djuhan, M. W., 2021. Analisis Gaya Belajar Visual, Auditori dan Kinestetik Dalam Pengembangan Prestasi Belajar Siswa. *IIPSI: Jurnal Ilmiah Ilmu Pengetahuan Sosial Indonesia*, 1(1), pp. 79-90.
- Maulid, M. I. & Arifin, T., 2022. Pengembangan Sistem Pakar Gaya Belajar Anak dengan Metode Fuzzy Logic Berbasis Android. *E-PROSIDING TEKNIK INFORMATIKA*, 3(1), pp. 11-22.
- Php, 2023. PHP: Hypertext Preprocessor. [Online] Available at: <https://www.php.net/> [Accessed 15 Juni 2023].
- Quinlan, J. R., 1993. *C4. 5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann Publishers.
- Rahmi, M. N. & Samsudi, M. A., 2020. Pemanfaatan Media Pembelajaran Berbasis Teknologi Sesuai dengan Karakteristik Gaya Belajar. *Jurnal Edumaspul*, 4(2), pp. 355-363.
- Ritonga, N. C. & Rahma, I. F., 2021. Analisis Gaya Belajar VAK pada Pembelajaran Daring Terhadap Minat Belajar Siswa. *Jurnal Analisis*, 7(1), pp. 76-88.
- Yunita, D. & Ikasari, I. H., 2021. Komparasi Metode Klasifikasi C4.5 dan Naïve Bayes untuk Mengukur Kepuasan Pelanggan. *Jurnal Informatika Universitas Pamulang*, 6(3), pp. 456-462.