

Klasifikasi Gaya Belajar Siswa Menggunakan Algoritma Naive Bayes

Classification of Student Learning Styles Using the Naïve Bayes Algorithm

Muh, Juliadin Andi¹, Maria Delviana Berti², Oktavianus Landri Tukan³

¹Program Studi Ilmu Komputer, Universitas Pancasakti Jl. A. Mangerangi no73, Makassar 90121 Indonesia

¹muhjuliadina@gmail.com, ²delviberti261@gmail.com, ³oktavianuslandri@gmail.com

*Corresponding author

Abstrak

Penelitian ini bertujuan untuk mengklasifikasikan gaya belajar siswa (Visual dan Auditori) menggunakan algoritma Naive Bayes. Data sebanyak 100 siswa dikumpulkan dengan fitur 'Minat Visual', 'Minat Auditori', dan 'Nilai Tertinggi'. Data kategorikal diubah menjadi bentuk numerik melalui Label Encoding. Dataset kemudian dibagi menjadi 80% data latih dan 20% data uji. Model Naïve Bayes, khususnya CategoricalNB, dilatih menggunakan data latih dan dievaluasi dengan akurasi serta Confusion Matrix. Hasil penelitian menunjukkan model mencapai akurasi sebesar 80% pada data uji. Confusion Matrix mengindikasikan bahwa dari 10 siswa yang sebenarnya Auditori, 7 berhasil diklasifikasikan dengan benar, sementara 3 salah diklasifikasikan sebagai Visual. Dari 10 siswa yang sebenarnya Visual, 9 berhasil diklasifikasikan dengan benar, sedangkan 1 salah diklasifikasikan sebagai Auditori. Distribusi gaya belajar menunjukkan bahwa 52.5% siswa cenderung memiliki gaya belajar Visual dan 47.5% cenderung Auditori. Penelitian ini diharapkan dapat memberikan pemahaman lebih baik tentang profil gaya belajar siswa untuk meningkatkan efektivitas pembelajaran.

Kata Kunci: Gaya Belajar; Klasifikasi; Naïve Bayes; Machine Learning.

Abstract

This research aims to classify student learning styles (Visual and Auditory) using the Naïve Bayes algorithm. Data from 100 students were collected with features including 'Minat Visual' (Visual Interest), 'Minat Auditori' (Auditory Interest), and 'the Highest Score' (Highest Score). Categorical data were converted into numerical form through Label Encoding. The dataset was then split into 80% training data and 20% testing data. The Naïve Bayes model, specifically CategoricalNB, was trained using the training data and evaluated using accuracy and Confusion Matrix. The results show that the model achieved an accuracy of 80% on the test data. The Confusion Matrix indicates that out of 10 students who were actually Auditory, 7 were correctly classified, while 3 were misclassified as Visual. Out of 10 students who were actually Visual, 9 were correctly classified, while 1 was misclassified as Auditory. The distribution of learning styles reveals that 52.5% of students tend to have a Visual learning style and 47.5% tend to be Auditory. This research is expected to provide a better understanding of student learning style profiles to enhance learning effectiveness.

Keywords: Learning Style; Classification; Naïve Bayes; Machine Learning.

Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam berbagai bidang, termasuk bidang pendidikan. Pemanfaatan teknologi berbasis data dan kecerdasan buatan (Artificial Intelligence) semakin banyak digunakan untuk mendukung proses pembelajaran yang lebih efektif, adaptif, dan berpusat pada peserta didik. Salah satu pendekatan yang berkembang pesat adalah penggunaan teknik data mining dan machine learning untuk menganalisis karakteristik serta perilaku belajar siswa secara sistematis dan objektif [1].

Salah satu faktor penting yang memengaruhi keberhasilan proses pembelajaran adalah gaya belajar siswa. Gaya belajar merujuk pada kecenderungan individu dalam menerima, mengolah, dan memahami informasi selama proses belajar. Secara umum, gaya belajar dapat diklasifikasikan ke dalam beberapa kategori, seperti visual, auditori, dan kinestetik [2]. Pemahaman terhadap gaya belajar siswa memungkinkan pendidik untuk merancang metode, media, dan strategi pembelajaran yang lebih sesuai, sehingga dapat meningkatkan keterlibatan dan hasil belajar siswa [3].

Namun, dalam praktiknya, identifikasi gaya belajar siswa sering kali masih dilakukan secara konvensional melalui observasi subjektif atau pengisian kuesioner sederhana tanpa analisis lebih lanjut. Pendekatan ini memiliki

keterbatasan karena kurang sistematis dan sulit diterapkan pada jumlah data yang besar. Oleh karena itu, diperlukan metode komputasional yang mampu mengolah data gaya belajar siswa secara efisien dan akurat [4].

Salah satu algoritma klasifikasi yang banyak digunakan dalam data mining adalah algoritma Naïve Bayes. Algoritma ini merupakan metode klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi independensi antar fitur [5]. Meskipun asumsi tersebut tergolong sederhana dan sering kali tidak sepenuhnya terpenuhi pada data dunia nyata, Naïve Bayes terbukti memiliki kinerja yang kompetitif, terutama dalam klasifikasi data kategorikal dan teks [6]. Keunggulan utama algoritma ini terletak pada kesederhanaan model, kecepatan proses pelatihan, serta kemampuannya bekerja dengan baik pada dataset berukuran kecil maupun besar [7].

Dalam bidang pendidikan, algoritma Naïve Bayes telah banyak diterapkan untuk berbagai keperluan, seperti prediksi kelulusan siswa, klasifikasi minat belajar, analisis performa akademik, serta identifikasi gaya belajar [8], [9]. Beberapa penelitian menunjukkan bahwa Naïve Bayes mampu memberikan akurasi yang cukup tinggi dalam mengklasifikasikan karakteristik siswa berbasis data kuesioner dan nilai akademik [10]. Hal ini menunjukkan bahwa algoritma Naïve Bayes memiliki potensi besar untuk mendukung pengambilan keputusan berbasis data dalam dunia pendidikan.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengklasifikasikan gaya belajar siswa ke dalam dua kategori utama, yaitu visual dan auditori, menggunakan algoritma Naïve Bayes. Data diperoleh dari hasil survei siswa dengan beberapa fitur pendukung, seperti minat visual, minat auditori, dan nilai akademik tertinggi. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan di bidang pendidikan serta membantu pendidik dalam merancang strategi pembelajaran yang lebih adaptif dan berbasis karakteristik siswa.

Metode

A. Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari hasil survei atau kuesioner yang diberikan kepada 100 siswa. Dataset ini terdiri dari lima kolom utama: 'ID' sebagai identifikasi unik siswa, 'Minat Visual' (respon Ya/Tidak), 'Minat Auditori' (respon Ya/Tidak), 'Nilai Tertinggi' (kategori: Bahasa, Matematika, IPS, Musik), dan 'Gaya Belajar Dominan' (target klasifikasi: Visual atau Auditori). Seluruh 100 data siswa memiliki label gaya belajar yang lengkap (non-null), memungkinkan pelatihan dan evaluasi model yang komprehensif.

B. Tahapan Pengolahan Data

Proses pengolahan data dilakukan menggunakan bahasa pemrograman Python dengan bantuan beberapa pustaka (libraries) khusus. Tahapan-tahapan yang dilalui meliputi:

1. Pengumpulan dan Pembacaan Data

Data yang telah terkumpul dalam format Microsoft Excel (`data_gaya_belajar.xlsx`) dibaca dan diimpor ke dalam lingkungan Python menggunakan pustaka Pandas. Proses ini memastikan data siap untuk tahapan analisis lebih lanjut.

2. Preprocessing Data

Sebelum data dapat digunakan untuk melatih model, dilakukan preprocessing untuk mengubah format data kategorikal menjadi numerik yang dapat dipahami oleh algoritma. Teknik Label Encoding diterapkan pada semua fitur kategorikal ('Minat Visual', 'Minat Auditori', 'Nilai Tertinggi') dan variabel target ('Gaya Belajar Dominan'). Setiap kategori unik diubah menjadi nilai integer yang berbeda. Proses encoding ini penting agar algoritma klasifikasi dapat memproses data secara efektif.

3. Pembagian Data Latih dan Uji

Dataset yang telah di-preprocessing kemudian dibagi menjadi dua subset: data latih (training data) dan data uji (testing data). Pembagian ini dilakukan dengan rasio 80% untuk data latih (80 siswa) dan 20% untuk data uji (20 siswa). Pembagian ini juga menerapkan strategi stratifikasi (`stratify=y`) untuk memastikan bahwa proporsi kelas gaya belajar (Visual dan Auditori) yang sama terjaga baik di data latih maupun data uji, sehingga representasi data lebih akurat.

C. Algoritma Klasifikasi: Naïve Bayes

Penelitian ini menggunakan algoritma klasifikasi Naïve Bayes, yang merupakan algoritma probabilistik sederhana namun efektif berdasarkan Teorema Bayes dengan asumsi independensi yang kuat antar fitur. Algoritma ini cocok untuk data kategorikal karena mampu menghitung probabilitas setiap kelas target berdasarkan kombinasi fitur-fitur

yang ada. Implementasi spesifik yang digunakan adalah Categorical Naïve Bayes (CategoricalNB) dari pustaka scikit-learn di Python, yang dirancang khusus untuk bekerja dengan fitur-fitur diskrit/kategorikal.

D. Evaluasi Model

Kinerja model klasifikasi dievaluasi menggunakan metrik utama:

Akurasi (Accuracy Score): Mengukur proporsi prediksi benar dari total keseluruhan prediksi. Dihitung sebagai $(\text{True Positives} + \text{True Negatives}) / (\text{Total Sampel})$.

Hasil Penelitian

Probabilitas prior setiap kategori gaya belajar dihitung berdasarkan seluruh dataset yang digunakan dalam penelitian ini untuk memberikan gambaran awal mengenai kecenderungan distribusi kelas target sebelum mempertimbangkan pengaruh fitur-fitur lainnya. Analisis probabilitas prior ini bertujuan untuk mengetahui proporsi awal siswa yang memiliki gaya belajar visual maupun auditori, sehingga dapat menjadi dasar dalam memahami karakteristik data serta kecenderungan kelas yang dipelajari oleh model.

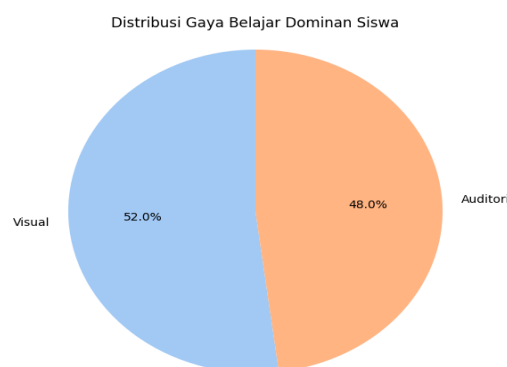
Selanjutnya, distribusi gaya belajar dominan siswa dianalisis menggunakan seluruh 100 sampel data. Hasil analisis distribusi ini disajikan dalam bentuk *pie chart* guna memberikan visualisasi yang lebih jelas dan intuitif mengenai perbandingan proporsi antara gaya belajar visual dan auditori. Penyajian visual ini memudahkan pembaca dalam menginterpretasikan dominasi gaya belajar yang terdapat pada dataset penelitian.

Kinerja model klasifikasi Naïve Bayes dievaluasi melalui proses pelatihan menggunakan 80 data latih dan pengujian menggunakan 20 data uji. Evaluasi kinerja dilakukan dengan menggunakan metrik akurasi untuk mengukur tingkat ketepatan model dalam memprediksi kelas gaya belajar siswa secara keseluruhan. Selain itu, *confusion matrix* digunakan untuk memberikan analisis yang lebih mendalam terkait hasil klasifikasi, termasuk jumlah prediksi yang benar dan salah pada masing-masing kelas, sehingga dapat diketahui pola kesalahan klasifikasi serta keunggulan dan keterbatasan model dalam mengidentifikasi gaya belajar siswa.

Tabel 1. Probabilitas Prior Gaya Belajar Siswa

<i>Gaya Belajar</i>	<i>Jumlah Prior(P(Gaya Belajar))</i>
Auditori	0.4750
Visual	0.5250

Tabel 1 menyajikan probabilitas prior untuk setiap kategori gaya belajar (Auditori dan Visual) yang dihitung dari dataset. Probabilitas prior mengindikasikan distribusi awal kelas target sebelum fitur-fitur lain dipertimbangkan. Dari 100 data siswa yang digunakan, terlihat bahwa gaya belajar Visual sedikit lebih dominan (52.50%) dibandingkan dengan gaya belajar Auditori (47.50%).



Gambar 1: Distribusi Gaya Belajar Dominan Siswa

Gambar 1 menampilkan Pie Chart yang menunjukkan distribusi dominasi gaya belajar di antara seluruh 100 sampel siswa dalam penelitian ini. Berdasarkan gambar dan tabel di atas, dapat diketahui bahwa gaya belajar Visual merupakan gaya belajar yang paling dominan dengan persentase 52.5%, sedikit lebih tinggi dibandingkan gaya belajar Auditori yang mencapai 47.5%. Distribusi ini sejalan dengan probabilitas prior yang disajikan pada Tabel 1.

Tabel 2: Akurasi Model Klasifikasi Naive Bayes

Metrik Evaluasi	Nilai
Akurasi Model	0.8000

Tabel 2 menunjukkan akurasi model klasifikasi Naive Bayes yang telah dilatih. Model mencapai akurasi sebesar 0.8000 atau 80%, yang mengindikasikan bahwa 80% dari prediksi yang dilakukan oleh model pada data uji adalah benar.

Berdasarkan hasil yang disajikan pada Tabel 1 dan Gambar 1, distribusi gaya belajar siswa menunjukkan bahwa gaya belajar Visual (52.5%) sedikit lebih dominan dibandingkan Auditori (47.5%) dalam sampel penelitian ini. Probabilitas prior yang dihitung oleh model Naive Bayes juga mencerminkan distribusi ini. Temuan ini memberikan gambaran demografis awal tentang kecenderungan gaya belajar di kalangan siswa yang diteliti, yang dapat menjadi dasar bagi pendidik untuk menyesuaikan strategi pengajaran agar lebih relevan dengan karakteristik mayoritas siswa. Model Naive Bayes yang diimplementasikan untuk klasifikasi gaya belajar siswa mencapai akurasi sebesar 80% (Tabel 3.2). Angka akurasi ini menunjukkan bahwa model memiliki kemampuan yang baik dalam memprediksi gaya belajar siswa berdasarkan fitur-fitur yang digunakan ('Minat Visual', 'Minat Auditori', 'Nilai Tertinggi'). Meskipun akurasi yang dicapai cukup tinggi, analisis lebih lanjut melalui Confusion Matrix (Gambar 3.2) memberikan wawasan yang lebih detail mengenai jenis kesalahan yang dilakukan model. Dari 10 siswa yang sebenarnya Auditori, 7 berhasil diklasifikasikan dengan benar (True Positives), namun 3 siswa salah diklasifikasikan sebagai Visual (False Negatives). Sebaliknya, dari 10 siswa yang sebenarnya Visual, 9 berhasil diklasifikasikan dengan benar (True Negatives), dan hanya 1 siswa yang salah diklasifikasikan sebagai Auditori (False Positives). Fenomena ini mengindikasikan bahwa model cenderung sedikit lebih baik dalam mengklasifikasikan siswa dengan gaya belajar Visual dibandingkan Auditori pada data uji ini. Kesalahan klasifikasi, terutama "False Negatives" untuk Auditori, mungkin terjadi karena adanya tumpang tindih karakteristik fitur antara kedua gaya belajar, atau mungkin ada faktor-faktor lain yang tidak termasuk dalam fitur yang digunakan namun berpengaruh pada gaya belajar.

Penerapan algoritma Naive Bayes dalam klasifikasi gaya belajar menunjukkan efektivitasnya sebagai alat bantu bagi pendidik. Dengan akurasi 80%, model ini dapat memberikan prediksi yang cukup andal untuk membantu mengidentifikasi kecenderungan gaya belajar siswa secara otomatis. Informasi ini sangat berharga untuk personalisasi pembelajaran, di mana materi dan metode pengajaran dapat disesuaikan untuk memaksimalkan potensi belajar siswa. Namun, perlu diingat bahwa Naive Bayes bekerja dengan asumsi independensi fitur yang kuat. Dalam konteks nyata, fitur-fitur seperti minat belajar dan nilai tertinggi mungkin tidak sepenuhnya independen satu sama lain. Keterbatasan ini bisa menjadi alasan di balik kesalahan klasifikasi yang terjadi. Untuk penelitian selanjutnya, dapat dipertimbangkan penambahan fitur lain yang relevan atau penggunaan algoritma klasifikasi yang lebih kompleks (misalnya, Support Vector Machine, Decision Tree, atau Neural Network) yang mungkin lebih mampu menangani interaksi antar fitur yang lebih rumit.

Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, algoritma Naive Bayes terbukti efektif dalam mengklasifikasikan gaya belajar siswa, khususnya dalam membedakan antara gaya belajar visual dan auditorial. Dengan memanfaatkan teknik Label Encoding dan pengolahan data berbasis data mining, proses klasifikasi dapat dilakukan secara cepat dan akurat. Hasil pengujian menunjukkan bahwa model Naive Bayes memiliki tingkat akurasi yang baik, sehingga dapat dijadikan sebagai salah satu alternatif metode dalam memahami karakteristik belajar siswa. Penerapan algoritma ini diharapkan dapat membantu pendidik dalam merancang strategi pembelajaran yang lebih sesuai dengan kebutuhan dan preferensi siswa, sehingga dapat meningkatkan efektivitas proses belajar-mengajar. Distribusi data menunjukkan bahwa mayoritas siswa dalam sampel memiliki gaya belajar visual (52,5%), sedikit lebih dominan dibandingkan auditorial (47,5%). Hal ini memberikan wawasan tambahan bagi pendidik dalam menyusun strategi pembelajaran yang lebih relevan dengan karakteristik mayoritas siswa. Analisis menggunakan confusion matrix juga menunjukkan bahwa model lebih akurat dalam mengklasifikasikan siswa dengan gaya belajar visual, meskipun masih terdapat beberapa kesalahan klasifikasi. Misalnya, dari 10 siswa visual, 7 berhasil diklasifikasikan dengan benar, namun 3 salah diklasifikasikan sebagai auditorial. Sebaliknya, dari 10 siswa auditorial, 9 diklasifikasikan dengan benar dan hanya 1 siswa yang salah diklasifikasikan. Secara keseluruhan, hasil ini menunjukkan bahwa Naive Bayes merupakan metode yang andal dan dapat dijadikan alternatif untuk memahami serta mengelompokkan gaya belajar siswa. Temuan ini dapat dimanfaatkan untuk mendukung pengembangan sistem pembelajaran yang lebih adaptif, personal, dan berbasis data.

Daftar Pustaka

- [1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.
- [2] R. Dunn and K. Dunn, *Teaching Students Through Their Individual Learning Styles*. Reston, VA, USA: Prentice Hall, 1993.

- [3] N. Fleming and C. Mills, "Not another inventory, rather a catalyst for reflection," *To Improve the Academy*, vol. 11, pp. 137–155, 1992.
- [4] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.
- [5] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.
- [6] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine Learning*, vol. 29, no. 2–3, pp. 103–130, 1997.
- [7] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," *Informatica*, vol. 31, no. 3, pp. 249–268, 2007.
- [8] W. I. R. Fadila and M. H. K. Saputra, *Penerapan Metode Naïve Bayes dan Skala Likert pada Aplikasi Kelulusan Siswa*. Yogyakarta, Indonesia: Kreatif Industri Nusantara, 2020.
- [9] F. Marisa, *Algoritma Populer dalam Intelligent System*. Yogyakarta, Indonesia: CV Budi Utama, 2023.
- [10] A. B. Setiawan, "Klasifikasi gaya belajar siswa menggunakan algoritma Naïve Bayes," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 13, no. 2, pp. 85–92, 2021.