

Penerapan Algoritma *K-Nearest Neighbor* Dan *Naive Bayes* Dalam Klasifikasi Minat Serta Gaya Belajar Mahasiswa

Jirjis De Rhinaldie¹

¹Pendidikan Teknik Informatika dan Komputer, Institut Prima Bangsa
Jl. Brigjend Dharsono No.20, Cirebon
Email: ¹jirjisderhinaldie16@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengklasifikasikan mahasiswa berdasarkan minat dan gaya belajar dengan menerapkan algoritma *K-Nearest Neighbor* (KNN) dan *Naive Bayes*, serta membandingkan kinerja kedua algoritma tersebut. Data penelitian diperoleh melalui penyebaran kuesioner kepada 112 mahasiswa Program Studi Pendidikan Teknik Informatika dan Komputer di Institut Prima Bangsa dengan variabel yang mencakup minat belajar (pemrograman, desain grafis, jaringan komputer, dan lain-lain), gaya belajar (*visual*, *auditori*, dan *kinestetik*), durasi belajar, tingkat keaktifan organisasi, serta penggunaan perangkat pembelajaran. Pendekatan yang digunakan adalah pendekatan kuantitatif dengan tahapan *data preprocessing* (pembersihan, normalisasi, dan *encoding*), pembagian data menjadi 80% data latih dan 20% data uji, serta evaluasi kinerja model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma KNN memiliki akurasi yang lebih tinggi dibandingkan *Naive Bayes*, dengan capaian 70% pada klasifikasi gaya belajar dan 43% pada klasifikasi minat belajar, sedangkan *Naive Bayes* hanya mencapai 22% pada kedua jenis klasifikasi tersebut. Dengan demikian, algoritma KNN terbukti lebih konsisten dan efektif dalam mengelola data serta memberikan hasil klasifikasi yang lebih akurat terhadap minat dan gaya belajar mahasiswa.

Kata kunci: *machine learning*, *K-Nearest Neighbor*, *Naive Bayes*, klasifikasi, gaya belajar, minat belajar

Abstract

This study aims to classify students based on their learning interests and styles using the K-Nearest Neighbor (KNN) and Naive Bayes algorithms, and to compare the performance of both algorithms. The data were obtained from questionnaires completed by 112 students from the Informatics and Computer Engineering Education Study Program at Institut Prima Bangsa. The variables include learning interests (programming, graphic design, computer networks, etc.), learning styles (visual, auditory, kinesthetic), study duration, organizational activity, and use of learning tools. The research employed a quantitative approach with data preprocessing, dataset splitting (80:20), and model evaluation using accuracy, precision, recall, and F1-score metrics. The results show that the KNN algorithm performed better than Naive Bayes. For learning style classification, KNN achieved 70% accuracy while Naive Bayes achieved 22%. For learning interest classification, KNN achieved 43% accuracy and Naive Bayes again 22%. Therefore, KNN demonstrates better consistency and suitability for classifying students' learning interests and styles.

Keywords: *machine learning, K-Nearest Neighbor, Naive Bayes, classification, learning style, learning interest*

1. PENDAHULUAN

Pendidikan merupakan salah satu aspek penting dalam pengembangan kualitas sumber daya manusia. Dalam proses pembelajaran di perguruan tinggi, setiap mahasiswa memiliki karakteristik yang berbeda dalam menentukan ketertarikan terhadap bidang pembelajaran maupun dalam menerima dan memahami informasi. Minat belajar berkaitan dengan ketertarikan dan dorongan mahasiswa terhadap bidang atau aktivitas pembelajaran tertentu, sedangkan gaya belajar menggambarkan cara individu dalam menerima, mengolah, dan memahami informasi yang diberikan [1]. Pemahaman terhadap kedua aspek tersebut penting karena dapat menjadi salah satu pertimbangan dalam menentukan strategi pembelajaran yang lebih sesuai dengan karakteristik mahasiswa[2].

Namun, identifikasi minat dan gaya belajar mahasiswa secara manual masih menghadapi beberapa kendala, seperti keterbatasan waktu, subjektivitas penilaian, serta jumlah mahasiswa yang relatif besar dalam satu kelas. Kondisi tersebut mendorong kebutuhan terhadap pendekatan berbasis teknologi yang mampu mengolah data mahasiswa secara lebih cepat dan objektif. Perkembangan kecerdasan buatan (Artificial Intelligence) dan machine learning memberikan peluang untuk melakukan analisis terhadap data pendidikan melalui pendekatan educational data mining. Salah satu penerapannya adalah klasifikasi karakteristik mahasiswa menggunakan algoritma machine learning, seperti K-Nearest Neighbor (KNN) dan Naive Bayes.

Penelitian terdahulu telah menunjukkan bahwa algoritma Naive Bayes dapat digunakan untuk klasifikasi gaya belajar mahasiswa. Ramadandi dan Jahring menerapkan metode Naive Bayes Classifier untuk mengklasifikasikan gaya belajar mahasiswa pada Program Studi Pendidikan Bahasa Indonesia dan Pendidikan Bahasa Inggris[3]. Penelitian tersebut memperoleh akurasi sebesar 90%, yang menunjukkan bahwa Naive Bayes mampu digunakan untuk mengidentifikasi gaya belajar mahasiswa berdasarkan data yang digunakan. Selanjutnya,

Penelitian lain mulai membandingkan KNN dan Naive Bayes secara langsung pada klasifikasi gaya belajar. Moerti dan Setiawan melakukan komparasi Naive Bayes dan KNN untuk klasifikasi gaya belajar mahasiswa menggunakan model Honey-Mumford yang terdiri atas Activist, Reflector, Theorist, dan Pragmatist. Data yang digunakan berasal dari riwayat akademik dan kuesioner perilaku mahasiswa pada mata kuliah Interpersonal Skill. Hasil penelitian menunjukkan bahwa Naive Bayes memperoleh akurasi sebesar 93,33% dan mengungguli KNN. Temuan tersebut menunjukkan bahwa perbedaan karakteristik dataset dan model gaya belajar dapat memengaruhi performa algoritma yang digunakan[4].

Perbandingan algoritma untuk klasifikasi gaya belajar juga dilakukan oleh Yuliansyah et al[5]. dengan menggunakan data kuesioner yang terdiri atas 1.170 data mahasiswa dan tiga kelas gaya belajar, yaitu visual, auditori, dan kinestetik. Penelitian tersebut membandingkan Naive Bayes, Decision Tree, Random Forest, dan K-Nearest

Neighbors. Hasil penelitian menunjukkan bahwa Naive Bayes memperoleh performa paling optimal dengan akurasi 90,60%.

Selain gaya belajar, aspek minat belajar juga telah menjadi objek penerapan machine learning. Penelitian sebelumnya menerapkan algoritma KNN untuk memprediksi minat belajar siswa berdasarkan nilai akademik dan intensitas bermain game. Penelitian tersebut menunjukkan bahwa KNN dapat digunakan untuk melakukan klasifikasi minat belajar berdasarkan karakteristik data yang digunakan. Pada konteks mahasiswa, Wibowo et al. menerapkan Naive Bayes untuk pemilihan konsentrasi mahasiswa berdasarkan data akademik. Penelitian tersebut memperoleh akurasi sebesar 84,27% dan menunjukkan bahwa algoritma Naive Bayes dapat digunakan untuk membantu mengidentifikasi kecenderungan pilihan bidang mahasiswa berdasarkan data yang tersedia.

Berdasarkan penelitian terdahulu tersebut, dapat diketahui bahwa KNN dan Naive Bayes telah digunakan dalam klasifikasi karakteristik belajar, baik untuk gaya belajar maupun aspek yang berkaitan dengan minat atau kecenderungan mahasiswa. Bahkan, penelitian Moerti dan Setiawan [4] telah membandingkan KNN dan Naive Bayes secara langsung pada klasifikasi gaya belajar mahasiswa. Penelitian Yuliansyah et al. [5] juga menunjukkan bahwa performa Naive Bayes dan algoritma klasifikasi lainnya dapat berbeda sesuai dengan karakteristik dataset dan model gaya belajar yang digunakan. Namun, penelitian terdahulu yang membandingkan KNN dan Naive Bayes dengan menggunakan dataset mahasiswa yang sama untuk mengklasifikasikan dua aspek sekaligus, yaitu minat belajar dan gaya belajar, masih terbatas. Penelitian sebelumnya cenderung berfokus pada klasifikasi gaya belajar saja atau minat belajar saja, dengan karakteristik objek, atribut, model gaya belajar, dan skenario pengujian yang berbeda.

Berdasarkan celah tersebut, penelitian ini menerapkan dan membandingkan algoritma K-Nearest Neighbor (KNN) dan Naive Bayes untuk klasifikasi minat dan gaya belajar mahasiswa. Kedua algoritma diterapkan pada dataset mahasiswa yang sama dengan tahapan preprocessing dan skema pembagian data yang sama, sehingga kinerja kedua algoritma dapat dibandingkan secara langsung. Evaluasi dilakukan menggunakan accuracy, precision, recall, dan F1-score untuk mengetahui kemampuan masing-masing algoritma dalam melakukan klasifikasi. Kontribusi penelitian ini adalah memberikan analisis komparatif KNN dan Naive Bayes pada dua aspek karakteristik belajar mahasiswa, yaitu minat belajar dan gaya belajar, sehingga dapat diketahui algoritma yang lebih sesuai berdasarkan karakteristik dataset penelitian. Hasil penelitian ini diharapkan dapat menjadi salah satu dasar dalam pengembangan pendekatan pembelajaran yang lebih adaptif terhadap karakteristik mahasiswa.

2. TINJAUAN PUSTAKA

2.1 Minat Belajar

Minat belajar merupakan dorongan internal yang muncul dari dalam diri individu untuk terlibat aktif dalam kegiatan pembelajaran. Menurut [6], minat belajar dapat diartikan sebagai rasa ketertarikan dan perhatian yang tinggi terhadap suatu bidang studi

atau aktivitas belajar tertentu. Minat yang tinggi akan mendorong mahasiswa untuk berpartisipasi aktif dalam proses pembelajaran dan memengaruhi tingkat pemahaman terhadap materi yang diberikan [7]. Faktor-faktor yang dapat memengaruhi minat belajar antara lain pengalaman sebelumnya, persepsi terhadap mata kuliah, lingkungan belajar, serta dukungan sosial [8]. Dengan demikian, pemahaman terhadap minat belajar mahasiswa dapat membantu pendidik menyesuaikan strategi pengajaran agar lebih efektif dan relevan.

2.2 Gaya Belajar

Gaya belajar merupakan cara individu dalam menerima, mengolah, dan memahami informasi yang diperoleh selama proses pembelajaran [9]. Secara umum, gaya belajar diklasifikasikan menjadi tiga kategori, yaitu: (1) *Visual*, yaitu individu yang lebih mudah memahami informasi melalui penglihatan seperti gambar, grafik, atau diagram; (2) *Auditori*, yaitu individu yang lebih mudah belajar melalui pendengaran seperti penjelasan verbal atau diskusi; dan (3) *Kinestetik*, yaitu individu yang belajar melalui pengalaman langsung dan aktivitas fisik [10]. Mengetahui gaya belajar mahasiswa sangat penting bagi pendidik untuk menyesuaikan metode penyampaian materi agar sesuai dengan preferensi dan kebutuhan belajar masing-masing individu.

2.3 Machine Learning Dalam Bidang Pendidikan

Perkembangan teknologi *machine learning* telah membawa perubahan signifikan dalam bidang pendidikan, terutama dalam hal analisis perilaku belajar dan prediksi hasil akademik [11]. *Machine learning* memungkinkan sistem untuk mempelajari pola dari data historis mahasiswa dan menghasilkan prediksi atau klasifikasi tanpa perlu diprogram secara eksplisit [12]. Salah satu penerapan utamanya adalah *educational data mining*, yaitu teknik penggalian data untuk mengidentifikasi pola belajar, minat, dan performa akademik mahasiswa. Pendekatan ini dapat membantu lembaga pendidikan dalam mengambil keputusan berbasis data guna meningkatkan efektivitas pembelajaran [13].

2.4 Algoritma K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor* (KNN) merupakan salah satu metode klasifikasi yang sederhana namun efektif dalam mengelompokkan data berdasarkan kemiripan antar atribut [14]. Prinsip dasar KNN adalah menentukan sejumlah tetangga terdekat dari data uji berdasarkan jarak tertentu, seperti jarak Euclidean, kemudian menentukan kelas data tersebut berdasarkan mayoritas kelas dari tetangga terdekat [15]. KNN tidak memerlukan asumsi distribusi data dan mampu bekerja baik pada dataset yang bersifat non-linear. Kelebihan lainnya adalah kemudahan implementasi serta kemampuan adaptif terhadap perubahan pola data [16]. Namun, kelemahannya terletak pada waktu komputasi yang relatif tinggi ketika jumlah data besar, serta sensitivitas terhadap fitur yang tidak relevan.

2.5 Algoritma Naïve Bayes

Naive Bayes merupakan algoritma klasifikasi berbasis probabilistik yang menggunakan teorema Bayes untuk menghitung kemungkinan suatu data termasuk ke dalam kelas tertentu [17]. Algoritma ini mengasumsikan bahwa setiap atribut bersifat independen terhadap atribut lainnya. Meskipun asumsi tersebut jarang terpenuhi sepenuhnya dalam data dunia nyata, *Naive Bayes* tetap menunjukkan performa yang baik pada berbagai kasus klasifikasi karena efisiensi perhitungannya

[18]. Dalam konteks pendidikan, *Naive Bayes* sering digunakan untuk menganalisis data survei atau kuesioner mahasiswa karena kemampuannya menangani variabel kategorikal dengan cepat dan akurat [19].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komputasi. Data diperoleh dari kuesioner yang diisi oleh 112 mahasiswa aktif Program Studi Pendidikan Teknik Informatika dan Komputer.

Tahapan penelitian meliputi:

1. Pengumpulan Data: Data dikumpulkan menggunakan kuesioner yang diberikan kepada mahasiswa. Informasi yang diperoleh kemudian digunakan sebagai dataset untuk membangun model klasifikasi. Variabel yang digunakan mencakup minat belajar, gaya belajar, durasi belajar, aktivitas organisasi, serta penggunaan perangkat pembelajaran.
2. Pra-pemrosesan Data: Meliputi pembersihan data, penghapusan duplikasi, penanganan *missing values*, normalisasi, dan *label encoding*.
3. Pembagian Dataset: Dataset yang telah melalui proses pra-pemrosesan dibagi menjadi dua bagian, yaitu data latih dan data uji. Proporsi yang digunakan adalah 80% untuk data latih dan 20% untuk data uji. Data latih digunakan dalam pembentukan model, sedangkan data uji digunakan untuk mengetahui kemampuan model ketika melakukan klasifikasi terhadap data yang tidak digunakan pada tahap pelatihan.
4. Penerapan Algoritma: Dua algoritma digunakan dalam eksperimen, yaitu *KNN* dan *Naive Bayes*. Implementasi dilakukan menggunakan Google Colaboratory. Kedua algoritma diterapkan pada dataset yang sama sehingga hasil performanya dapat dibandingkan dengan dasar pengujian yang sama.
5. Evaluasi Model: Evaluasi dilakukan dengan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* digunakan untuk melihat proporsi prediksi yang berhasil diklasifikasikan dengan benar. *Precision* menggambarkan ketepatan prediksi pada kelas tertentu, sedangkan *recall* menunjukkan kemampuan model dalam menemukan data yang termasuk dalam suatu kelas. *F1-score* digunakan untuk melihat keseimbangan antara *precision* dan *recall*.

Untuk karakteristik datasetnya adalah sebagai berikut:

Tabel 3.1

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 335 entries, 0 to 334

Data columns (total 18 columns):

#	Column	Non-Null Count		Dtype
0	Timestamp	125	non-null	datetime[ms]
1	Nama	125	non-null	object
2	Nim	125	non-null	int64
3	Jenis Kelamin	125	non-null	object
4	Semester	125	non-null	int64
5	Pada Minat Belajar, Saya lebih tertarik pada bidang :	125	non-null	object
6	Ketika Belajar Saya lebih suka :	125	non-null	object
7	Berapa lama waktu rata-rata Anda belajar perminggu diluar jam kuliah? (...jam/minggu)	125	non-null	object
8	Apakah anda aktif dalam organisasi kampus? (ya/tidak)	125	non-null	object
9	Apakah saat ini anda berkuliah sambil bekerja? (ya/tidak)	125	non-null	object
10	Perangkat apa yang anda gunakan untuk belajar? (boleh lebih dari satu)	125	non-null	object

dtypes: datetime64[ns](1), int64(2), object(8)

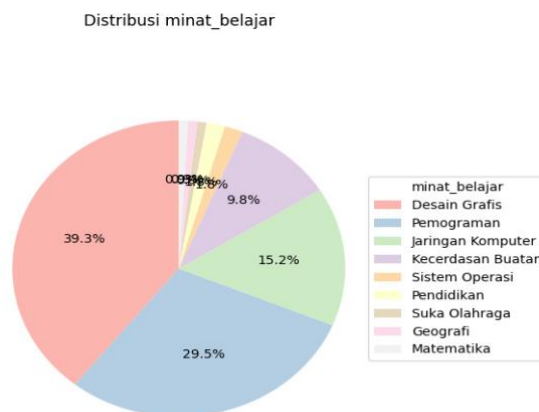
memory usage: 10.9+ KB

Data dalam penelitian ini dikumpulkan menggunakan instrumen kuesioner daring yang disebarluaskan kepada mahasiswa Pendidikan Teknik Informatika dan Komputer IPB CIREBON. Kuesioner dirancang untuk menggali informasi mengenai profil demografis, preferensi belajar, serta aktivitas ekstrakurikuler responden.

Pada tahap awal pengumpulan data, terdapat 335 respons yang masuk ke dalam sistem. Namun, sebelum dilakukan analisis lebih lanjut, dilakukan tahap *preprocessing* (pra-pemrosesan) dan pembersihan data (*data cleaning*). Sebanyak 210 respons dibuang (*dropped*) karena terindikasi tidak lengkap (*missing values*) atau tidak mengisi seluruh item pertanyaan. Dengan demikian, sampel final yang valid dan digunakan dalam penelitian ini berjumlah 125 responden.

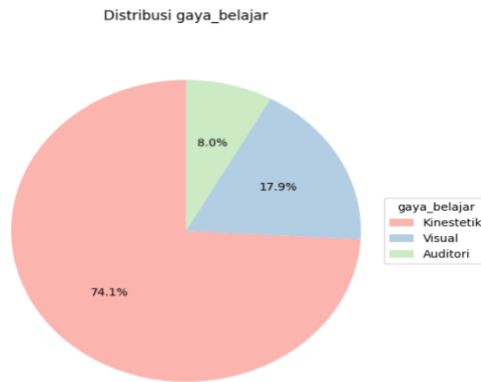
Instrumen penelitian ini mengukur beberapa variabel yang dikategorikan menjadi tiga kelompok utama. Pertama, variabel profil demografis yang meliputi jenis kelamin dan semester aktif. Kedua, variabel preferensi dan kebiasaan belajar, yang mencakup minat bidang studi, gaya belajar, estimasi durasi belajar mandiri di luar jam kuliah, serta perangkat teknologi yang digunakan. Ketiga, variabel status aktivitas mahasiswa, yang mengukur keaktifan responden dalam organisasi kampus dan status mereka sebagai mahasiswa yang juga bekerja. Untuk menjaga etika penelitian dan privasi responden, data identitas pribadi seperti Nama dan Nomor Induk Mahasiswa (NIM) telah dianonimkan dalam proses analisis.

Dalam penelitian ini, EDA dilakukan setelah data melalui proses pembersihan, untuk memastikan bahwa data siap untuk dianalisis lebih lanjut dan dapat memberikan insight yang bermanfaat. Analisis univariate di terapkan pada tahap ini untuk melihat distribusi data pada tiap fitur.



Gambar 3.1 Distribusi minat belajar

Dari visualisasi pada gambar 3.2 diata dapat disimpulkan bahwa distribusi fitur minat belajar. Responden memiliki beragam minat dalam bidang TIK. Sebagian besar responden berada pada Desain grafis sebanyak 44 orang atau 39,3%, pemrograman 29,5%, jaringan computer 15,2 kecerdasan buatan 9,8%, system operasi 1,8%. Dan beberapa topik diluar TIK seperti matematika, Pendidikan,dan geografi.



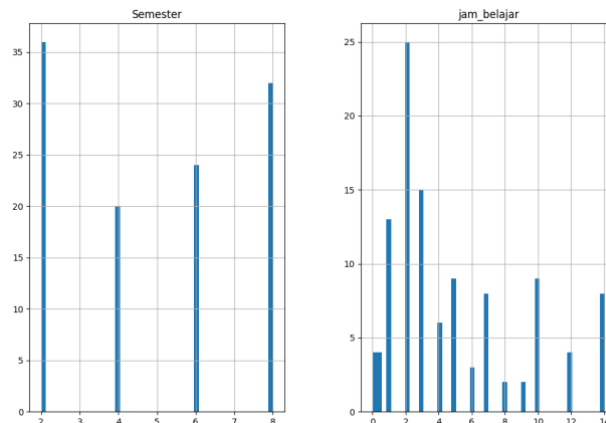
Gambar 3.2 Distribusi gaya belajar

Dari visualisasi gambar 3.3 diatas dapat disimpulkan bahwa distribusi fitur gaya belajar. Sebagian besar responden memiliki gaya belajar kinestetik sebanyak 83 responden atau sebesar 74,1%, visual 17,9% dan diikuti auditori sebesar 8%.

Untuk distribusi fitur organisasi, Sebagian besar responden tidak mengikuti organisasi dengan sebanyak 76 responden dan yang mengikuti organisasi sebanyak 36 responden atau sebesar 32,1%.

Untuk fitur bekerja, Sebagian responden tidak bekerja sebesar 78,6% dan responden yang bekerja sebesar 21,4%.

Sedangkan untuk fitur perangkat bantu, mayoritas responden menggunakan lebih dari satu perangkat, khususnya laptop dan ponsel sebanyak 77 responden atau sebesar 68,8%, responden yang menggunakan laptop saja sebanyak 21 responden, dan sisanya menggunakan kombinasi dengan computer desktop atau tablet.



Gambar 3.3 Distribusi fitur numerik

Dari hasil visualisasi pada gambar 3.7 menggunakan bar chart untuk melihat sebaran jumlah jam belajar per minggu dan semester responden. Pada fitur semester Sebagian besar responden berada disemester awal dan akhir, sedangkan fitur jam belajar mayoritas responden belajar antara 1 hingga 3 jam per minggu, dan hanya sedikit yang belajar lebih dari 10 jam.

Parameter algoritma *K-Nearest Neighbors* (KNN) yang digunakan dalam penelitian ini adalah nilai $K=3$, yang berarti bahwa klasifikasi data uji dilakukan berdasarkan tiga data latih terdekat yang memiliki tingkat kemiripan tertinggi dengan data baru. Untuk mengukur jarak kedekatan antar data, penelitian ini menggunakan metrik jarak Minkowski dengan parameter $p=2$, yang dikenal sebagai jarak Euclidean. Sementara itu, untuk algoritma *Naive Bayes*, penelitian ini mengimplementasikan varian *Gaussian Naive Bayes* yang sesuai untuk data dengan distribusi normal.

Pembagian data latih dan data uji dilakukan dengan rasio 80:20 menggunakan fungsi *train_test_split* dari pustaka *scikit-learn*. Proses pembagian data ini dilakukan secara acak (*random splitting*) untuk memastikan bahwa distribusi data pada kedua subset bersifat representatif dan menghindari bias dalam proses pelatihan dan pengujian model.

Confusion Matrix - KNN

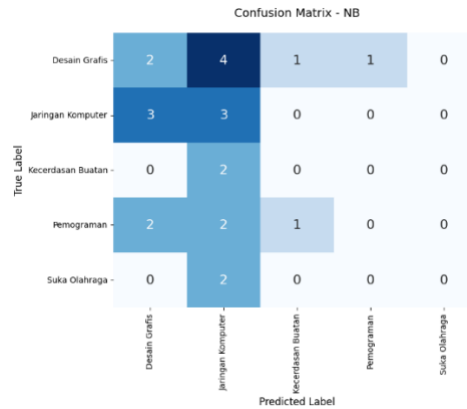
True Label	Desain Grafis	6	1	0	1	0
	Jaringan Komputer	3	1	2	0	0
	Kecerdasan Buatan	1	1	0	0	0
	Pemrograman	2	0	0	3	0
	Suka Olahraga	0	1	0	1	0
		Predicted Label				
		Desain Grafis	Jaringan Komputer	Kecerdasan Buatan	Pemrograman	Suka Olahraga

Gambar 3.4 Confusion matrix KNN minat belajar

Dari hasil *Confusion matrix* menunjukkan bahwa prediksi model paling akurat pada kelas Desain Grafis (6 dari 8 benar), dan Pemrograman (3 dari 5 benar). Banyak kesalahan klasifikasi terjadi pada label lain, terutama Jaringan Komputer dan Kecerdasan Buatan, yang banyak salah diklasifikasikan sebagai Desain Grafis.

Sedangkan klasifikasi gaya belajar menggunakan *KNN*, akurasi sebesar 70% dan macro average F1-score 0.46. Model menunjukkan kinerja sangat baik dalam memprediksi label Kinestetik dengan F1-score 0.82, yang juga merupakan kelas mayoritas. Namun, kelas Auditori tidak dapat diprediksi sama sekali ($F1 = 0$), yang menunjukkan bahwa model cenderung bias terhadap kelas mayoritas.

Dan confusion matrix nya mengilustrasikan bahwa dari 17 data aktual untuk Kinestetik, sebanyak 14 diklasifikasikan dengan benar. Namun untuk Auditori dan Visual, prediksi model masih kurang akurat, dengan adanya kesalahan klasifikasi ke kelas Kinestetik.



Gambar 3.5 Confusion matrix naïve bayes minat belajar

Berdasarkan gambar 3.5 menunjukkan confusion matrix hasil klasifikasi minat belajar menggunakan algoritma Gaussian Naive Bayes. Baris pada matrix menunjukkan label sebenarnya (actual label), sedangkan kolom menunjukkan hasil prediksi model (predicted label). Dari matrix tersebut, terlihat bahwa:

- Desain Grafis (baris pertama) paling sering diprediksi dengan benar oleh model, yaitu sebanyak 2 dari 8 data.
- Jaringan Komputer (baris kedua) juga menunjukkan adanya keberhasilan model dalam memprediksi 3 data secara benar dari total 6, ini berarti memiliki recall yang cukup tinggi dibanding kelas lain.
- Untuk Kecerdasan Buatan, Pemrograman, dan Sistem Operasi, model tidak berhasil memprediksi satu pun data dengan benar. Semua data dari kelas-kelas tersebut salah diklasifikasikan ke kelas lain yang menunjukkan bahwa model tidak mampu membedakan ciri khas dari kelas tersebut.

Secara umum, tampak bahwa persebaran prediksi cukup buruk, dengan mayoritas prediksi terkonsentrasi ke kelas Desain Grafis dan Jaringan Komputer, meskipun kenyataannya data memiliki lebih banyak variasi kelas. Ini juga menjelaskan mengapa nilai f1-score untuk beberapa kelas adalah 0.

4. PEMBAHASAN

Hasil pengujian menunjukkan bahwa algoritma *KNN* memiliki performa lebih baik dibandingkan *Naive Bayes*.

- Pada klasifikasi gaya belajar, *KNN* memperoleh akurasi sebesar **70%**, sementara *Naive Bayes* hanya mencapai **22%**.
- Pada klasifikasi minat belajar, *KNN* menghasilkan akurasi **43%**, dan *Naive Bayes* tetap **22%**.

Tabel 4.1 *Classification Report KNN minat belajar*

label	Precision	Recall	F1-Score	Support
Auditori	0.00	0.00	0.00	3
Kinestetik	0.82	0.82	0.82	17
Visual	0.50	0.67	0.57	3
Accuracy			0.70	23
Macro avg	0.44	0.50	0.46	23
Weighted avg	0.67	0.70	0.68	23

Untuk evaluasi kinerja model klasifikasi menunjukkan tingkat akurasi keseluruhan (*overall accuracy*) sebesar **70%**, dengan rata-rata *weighted precision*, *recall*, dan *F1-score* masing-masing mencapai 0.67, 0.70, dan 0.68. Meskipun metrik *weighted average* mengindikasikan performa yang cukup baik secara agregat, nilai *macro average F1-score* yang hanya sebesar 0.46 mengungkap adanya disparitas performa yang signifikan antar kelas akibat ketidakseimbangan distribusi data (*imbalanced data*). Hal ini terlihat dari kemampuan model dalam memprediksi kelas mayoritas seperti Desain Grafis dan Pemrograman (dengan *F1-score* masing-masing 0.60), di sisi lain, model sama sekali gagal mengidentifikasi kelas minoritas seperti Kecerdasan Buatan dan Sistem Operasi (ditandai dengan nilai *precision*, *recall*, dan *F1-score* sebesar 0.00). Kondisi ini mengindikasikan bahwa model cenderung bias terhadap kelas dengan jumlah sampel lebih besar, sehingga diperlukan penanganan khusus terhadap ketidakseimbangan data untuk meningkatkan sensitivitas (*recall*) dan ketepatan (*precision*) model pada seluruh kategori kelas secara merata.

Sedangkan untuk Tingkat akurasi model naïve bayes sebesar **43%** dan macro average *F1-score* 0.28. Model *KNN* memberikan performa tertinggi untuk label Desain Grafis dan Pemograman, dengan *F1-score* masing-masing 0.60. Namun, kinerja model sangat rendah pada kelas-kelas seperti Kecerdasan Buatan dan Sistem Operasi, yang menunjukkan ketidakseimbangan performa antar kelas. Ini juga tergambar dari nilai macro average yang rendah.

Tabel 4.2 Classification report Naïve bayes minat belajar

Label	Precision	Recall	F1-Score	Support
Desain Grafis	0.29	0.25	0.27	8
Jaringan Komputer	0.23	0.50	0.32	6
Kecerdasan Buatan	0.00	0.00	0.00	2
Pemograman	0.00	0.00	0.00	5
Sistem Operasi	0.00	0.00	0.00	2
Accuracy			0.22	23
Macro avg	0.10	0.15	0.12	23
Weighted avg	0.16	0.22	0.18	23

Untuk hasil evaluasi klasifikasi terhadap minat belajar menggunakan algoritma Gaussian Naive Bayes, diperoleh akurasi model sebesar **22%**, yang menunjukkan bahwa hanya sekitar 2 dari 10 prediksi yang berhasil dilakukan dengan benar oleh model. Nilai precision, recall, dan f1-score dari masing-masing kelas juga menunjukkan performa yang cukup rendah. Misalnya, untuk kelas Desain Grafis, precision sebesar 0.29 dan recall sebesar 0.25 menandakan bahwa meskipun ada beberapa prediksi yang benar, jumlah prediksi salah cukup tinggi. Sementara itu, kelas Jaringan Komputer memiliki recall tertinggi sebesar 0.50, yang berarti setengah dari data pada kelas ini berhasil dikenali, meskipun precision-nya hanya 0.23. Namun, untuk kelas Kecerdasan Buatan, Pemograman, dan Sistem Operasi, nilai recall dan precision adalah 0.00, menunjukkan bahwa model tidak berhasil mengklasifikasikan data dari kelas-kelas tersebut sama sekali. Hal ini mencerminkan bahwa model mengalami kesulitan besar dalam membedakan karakteristik antar kelas yang lebih minoritas atau memiliki distribusi yang mirip dengan kelas lain. urasi model sebesar 22% dan macro average F1-score 0.12. Model naïve bayes memberikan performa 0 untuk label Desain Grafis dan Pemograman, dengan F1-score masing-masing 0.60. Namun, kinerja model sangat rendah pada kelas-kelas seperti Kecerdasan Buatan dan Sistem Operasi, yang menunjukkan ketidakseimbangan performa antar kelas. Ini juga tergambar dari nilai macro average yang rendah.

Pada klasifikasi gaya belajar, model Naive Bayes juga menunjukkan hasil yang serupa dengan klasifikasi minat belajar, yakni akurasi sebesar 22%. Dari classification report, terlihat bahwa kelas Auditori dan Visual memiliki recall yang tinggi, masing-masing sebesar 0.67, meskipun jumlah datanya sangat kecil (masing-masing hanya 3 data). Di sisi lain, kelas Kinestetik yang merupakan mayoritas (17data), hanya memiliki recall sebesar 0.06, menunjukkan bahwa model hampir sepenuhnya gagal dalam mengenali pola pada data mayoritas tersebut. Hal menarik yang terlihat adalah precision kelas Kinestetik sebesar 1.00, yang artinya setiap kali model memprediksi "Kinestetik", prediksinya benar. Namun karena jumlah prediksi "Kinestetik" sangat sedikit, hal ini tidak bisa dianggap

sebagai indikator performa yang baik secara keseluruhan. Ketimpangan kinerja antar kelas ini memperjelas bahwa model lebih cenderung memberikan prediksi pada kelas minoritas yang mungkin memiliki fitur lebih “terstandarisasi” atau unik dibandingkan dengan kelas Kinestetik yang lebih bervariasi. Hal ini juga kembali mengindikasikan bahwa asumsi independensi antar fitur yang digunakan oleh Naive Bayes tidak cocok untuk data gaya belajar yang memiliki kemungkinan pola hubungan antar fitur yang kompleks. Sama seperti pada klasifikasi minat belajar, model ini tampaknya tidak cocok diterapkan pada data yang tidak seimbang atau berstruktur rumit.

Temuan ini sejalan dengan penelitian Moerti dan Setiawan [4] serta Yuliansyah et al. [5], yang menyatakan bahwa performa algoritma *machine learning* sangat bergantung pada karakteristik dataset dan model yang digunakan. Meskipun pada penelitian Ramadandi dan Jahring [3] serta Yuliansyah et al. [5] algoritma Naive Bayes menunjukkan akurasi yang sangat tinggi (di atas 90%), pada dataset penelitian ini Naive Bayes justru menunjukkan performa yang rendah (22%). Hal ini mengindikasikan bahwa asumsi dasar Naive Bayes, yaitu independensi antarfitur, tidak terpenuhi secara optimal pada data kuesioner minat dan gaya belajar mahasiswa. Fitur-fitur seperti perangkat yang digunakan, durasi belajar, dan keaktifan organisasi kemungkinan memiliki korelasi atau keterkaitan satu sama lain. Sebaliknya, KNN tidak bergantung pada asumsi distribusi data dan bekerja berdasarkan kedekatan jarak (*distance-based*). Pendekatan ini membuat KNN lebih adaptif dalam menangkap pola kompleks dan non-linear dari fitur-fitur yang saling berkaitan pada dataset ini, sehingga secara komparatif mengungguli Naive Bayes.

Meskipun KNN memberikan hasil yang lebih unggul dibandingkan Naive Bayes, klaim bahwa model ini “lebih efektif” secara mutlak perlu diperhalus, terutama pada klasifikasi minat belajar yang hanya mencapai akurasi 43%. Akurasi 43% ini, meskipun secara komparatif lebih tinggi dari Naive Bayes (22%) dan *random guessing*, masih tergolong rendah untuk diimplementasikan secara langsung sebagai sistem pendukung keputusan. Rendahnya akurasi ini sangat dipengaruhi oleh karakteristik dataset yang mengalami ketidakseimbangan kelas (*imbalanced data*) yang ekstrem. Hal ini terlihat dari dominasi kelas mayoritas seperti gaya belajar Kinestetik (74,1%) dan minat Desain Grafis (39,3%), serta ukuran sampel akhir yang relatif kecil (125 data valid).

Kondisi *imbalanced data* ini, menyebabkan model KNN cenderung bias memprediksi kelas mayoritas dan gagal mengenali kelas minoritas, yang dibuktikan dengan nilai *F1-score* 0.00 pada kelas Kecerdasan Buatan dan Sistem Operasi (Tabel 4.1). Oleh karena itu, hasil penelitian ini tidak serta merta menyatakan KNN sebagai algoritma yang mutlak terbaik, melainkan menunjukkan bahwa KNN lebih *sesuai* (*suitable*) untuk karakteristik data saat ini dibandingkan Naive Bayes. Untuk meningkatkan akurasi klasifikasi minat belajar di masa depan, diperlukan penanganan khusus terhadap ketidakseimbangan data (seperti penerapan *SMOTE* atau *undersampling*), seleksi fitur yang lebih ketat, atau penambahan jumlah sampel data agar model dapat mempelajari distribusi kelas minoritas dengan lebih baik.

5. KESIMPULAN

Penelitian ini telah menerapkan algoritma *KNN* dan *Naive Bayes* untuk melakukan klasifikasi terhadap minat serta gaya belajar mahasiswa. Dataset yang digunakan berasal dari kuesioner 112 mahasiswa Program Studi Pendidikan Teknik Informatika dan Komputer. Berdasarkan hasil pengujian, *KNN* memberikan performa yang lebih tinggi dibandingkan *Naive Bayes*. Pada klasifikasi gaya belajar, *KNN* memperoleh akurasi 70%,

sedangkan *Naive Bayes* menghasilkan 22%. Sementara itu, pada klasifikasi minat belajar, *KNN* memperoleh akurasi 43% dan *Naive Bayes* kembali menghasilkan 22%.

Hasil tersebut menunjukkan bahwa *KNN* lebih konsisten dalam menangani dataset penelitian. Salah satu faktor yang dapat menjelaskan perbedaan tersebut adalah karakteristik data yang tidak sepenuhnya sesuai dengan asumsi independensi fitur pada *Naive Bayes*. Oleh karena itu, berdasarkan hasil eksperimen yang dilakukan, *KNN* dapat dipertimbangkan sebagai metode yang lebih sesuai untuk klasifikasi minat dan gaya belajar pada dataset penelitian ini.

DAFTAR PUSTAKA

- [1] rahayu deny danar dan alvi furwanti Alwie, A. B. Prasetyo, R. Andespa, P. N. Lhokseumawe, and K. Pengantar, "Tugas Akhir Tugas Akhir," *J. Ekon. Vol. 18, Nomor 1 Maret201*, vol. 2, no. 1, pp. 41–49, 2020.
- [2] E. Ayuningrum, A. E. Minarno, and G. W. Wicaksono, "Klasifikasi Penyakit Katarak Pada Mata Manusia Menggunakan Metode Convolutional Neural Network," *J. Repos.*, vol. 4, no. 4, pp. 491–496, 2024, doi: 10.22219/repositor.v4i4.32285.
- [3] S. Ramadandi and J. Jahring, "Klasifikasi Gaya Belajar Mahasiswa Menggunakan Metode Naïve Bayes Classifier," *J. Teknol. dan Inf.*, vol. 10, no. 2, pp. 170–179, 2020, doi: 10.34010/jati.v10i2.
- [4] K. Naive and K. Gaya, "Comparison of Naive Bayes and KNN for Honey-Mumford Learning Style Classification in Interpersonal Skill," vol. 8, no. 2, pp. 153–160, 2025.
- [5] H. Yuliansyah, A. T. Lestari, and A. Yudhana, "Machine Learning Algorithm for Questionnaire-Based Student Learning Style Classification," *Int. J. Adv. Data Inf. Syst.*, vol. 7, no. 1, pp. 470–482, 2026, doi: 10.59395/ijadis.v7i1.1536.
- [6] Nurzaki Alhafiz, "Analisis Profil Gaya Belajar Siswa Untuk Pembelajaran Berdiferensiasi Di Smp Negeri 23 Pekanbaru," *J. Cakrawala Ilm.*, vol. 1, no. 5, pp. 1133–1142, 2022, doi: 10.53625/jcijurnalcakrawalailmiah.v1i5.1203.
- [7] I. Palopo, "Klasifikasi Gaya Belajar Matematika Kelas X Sma Negeri 7 Luwu Klasifikasi Gaya Belajar Matematika Kelas X Sma Negeri 7 Luwu," *Repos. Iain Palopo*, 2023, [Online]. Available: <http://repository.iainpalopo.ac.id/id/eprint/7917/1/Proposal skripsi revisi 9-7 Lengkap.pdf>
- [8] S. Pokhrel, "No TitleEΛENH," *Ayan*, vol. 15, no. 1, pp. 37–48, 2024.
- [9] S. Prasetyo and T. Dewayanto, "Penerapan Machine Learning, Deep Learning, Dan Data Mining Dalam Deteksi Kecurangan Laporan Keuangan-a Systematic Literature Review," *Diponegoro J. Account.*, vol. 13, no. 3, pp. 1–12, 2024, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/accounting>
- [10] D. Pratmanto, A. Widayanto, Y. M. Kristania, and R. Wijianto, "Analisis Perbandingan Algoritma Naive Bayes Dan KNN Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Vidio Di Google Play Store," vol. 4, no. 2, pp. 119–124, 2024.
- [11] P. D. Rinanda, B. Delvika, S. Nurhidayarnis, N. Abror, and A. Hidayat, "Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes pada Ibu Hamil," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 68–75, 2022, doi: 10.57152/malcom.v2i2.432.
- [12] N. C. Ritonga and I. F. Rahma, "Analisis gaya belajar VAK pada pembelajaran daring terhadap minat belajar siswa," *J. Anal.*, vol. 7, no. 1, pp. 76–86, 2021, doi: 10.15575/ja.v7i1.11878.
- [13] F. A. Rohmansyah, B. Bintoro, and I. Santoso, "Analisis Sentimen Terhadap Penerapan Sistem Ganjil Genap Menggunakan Metode K-Nearest Neighbor (Knn)," *iKRAITH-INFORMATIKA*, vol. 7, no. 2, pp. 165–169, 2022.

- [14] J. Saputra, Y. Sa, V. Yoga Pudya Ardhana, and M. Afriansyah, "RESOLUSI : Rekayasa Teknik Informatika dan Informasi Klasifikasi Kematangan Buah Alpukat Mentega Menggunakan Metode K-Nearest Neighbor Berdasarkan Warna Kulit Buah," *Media Online*, vol. 3, no. 5, pp. 347–354, 2023, [Online]. Available: <https://djournals.com/resolusi>
- [15] S. Sati, D. Setiana, and A. N. Amelia, "Implementasi Pembelajaran E-Learning Terhadap Minat Belajar Peserta Didik di Masa Pandemi Covid-19," *J. Basicedu*, vol. 6, no. 1, pp. 51–57, 2021, doi: 10.31004/basicedu.v6i1.1899.
- [16] F. Sholekhah, A. D. Putri, and L. Efrizoni, "Comparison of Naive Bayes and K-Nearest Neighbors Algorithms for Metabolic Syndrome Classification Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbors untuk Klasifikasi Metabolik Sindrom," vol. 4, no. April, pp. 507–514, 2024.
- [17] S. Suherman, M. Purnamasari, A. Rahmatullah, S. Samsuni, and A. Kautsar, "Klasifikasi Gaya Belajar Pada Mahasiswa Menggunakan Metode Algoritma C4.5," *JIKA (Jurnal Inform.)*, vol. 8, no. 1, p. 104, 2024, doi: 10.31000/jika.v8i1.10306.
- [18] S. S. H. Syahril Dwi Prasetyo, "Jurnal KomtekInfo Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan," vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [19] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022, doi: 10.35957/jatisi.v9i2.2048.

Biodata Penulis



Nama Lengkap : Jirjis De Rhinaldie

Jenis Kelamin : Laki-laki

Tempat/Tanggal lahir : Indramayu, 16 September 2000

Agama : Islam

Alamat : Ds. Karanganyar Blok III RT 002/005
Kandanghaur Indramayu Jawa Barat

Negara : Indonesia

Pendidikan:

Kuliah	: Institut Prima Bangsa Cirebon	2021 – 2025
SMA	: Ponpes Daarut Tarbiyah	2016 - 2018
MTS	: Ponpes Daarut Tarbiyah	2013 – 2015
SD	: SDI AL-JANNAH	2007 - 2012