INVENTOR: Jurnal Inovasi dan Tren Pendidikan Teknologi Informasi
https://ejournal.tsb.ac.id/index.php/inventor
Vol. 4, No. 3: Oktober 2026
e-ISSN: 3026-3190

Klasifikasi Tingkat Pemahaman Mahasiswa Berbasis Data Kuis Digital Menggunakan Naïve Bayes

Putri Ramadani[1),*], Nur Aini[2)], Suerni[1)], Atika Sadariah Nasution[1)]

[1)]Institut Teknologi dan Kesehatan Ika Bina
[2)]Institut Teknologi dan Kesehatan Sumatera Utara

*****Correspondence: putri.ramadani98@itkes-ikabina.ac.id

Abstract

Language learning evaluation through digital quiz platforms (Quizizz, Kahoot, Google Forms, and Moodle) generates response data with strong potential for systematically mapping students' levels of understanding. This study aims to design a classification model for students' understanding levels in digital quiz-based language learning by applying the Naïve Bayes algorithm via RapidMiner Studio 10.1. The novelty of this research lies in utilizing digital quiz interaction questionnaire data specifically tailored to measure comprehension within the context of language learning. Involving 173 students from four study programs at Institut Teknologi dan Kesehatan Ika Bina Labuhanbatu selected through total sampling, the data were gathered using a 27-item instrument on a five-point Likert scale. The dataset was subsequently partitioned into 138 training samples (80%) and 35 testing samples (20%). Based on the confusion matrix evaluation, the model achieved an accuracy of 91.43%, with 32 out of 35 testing samples correctly classified. The model categorized understanding levels into two classes: High and Very High. The Very High category recorded a precision of 83.33% and a recall of 100%, whereas the High category attained a precision of 100% and a recall of 85%. These findings demonstrate the effectiveness and potential of the Naïve Bayes algorithm in supporting data-driven educational evaluation.

Keywords: *Classification; Digital Quizzes; Naïve Bayes; Student Understanding; RapidMiner Studio 10.1.*

Abstrak

Evaluasi pembelajaran bahasa melalui platform kuis digital (Quizizz, Kahoot, Google Form, dan Moodle) menghasilkan data respons yang potensial untuk dimanfaatkan dalam memetakan tingkat pemahaman mahasiswa secara sistematis. Penelitian ini bertujuan untuk merancang model klasifikasi tingkat pemahaman mahasiswa pada ranah pembelajaran bahasa berbasis kuis digital dengan menerapkan algoritma Naïve Bayes via RapidMiner Studio 10.1. Kebaruan studi ini terletak pada pemanfaatan data kuesioner interaksi kuis digital yang dikhususkan untuk mengukur pemahaman pada konteks pembelajaran bahasa. Melibatkan 173 mahasiswa dari empat program studi di Institut Teknologi dan Kesehatan Ika Bina Labuhanbatu yang dipilih melalui teknik total sampling, data dikumpulkan menggunakan 27 butir instrumen berskala Likert lima poin. Dataset kemudian dibagi menjadi 138 data latih (80%) dan 35 data uji (20%). Berdasarkan evaluasi confusion matrix, model ini menghasilkan tingkat akurasi sebesar 91,43% dengan rincian 32 dari 35 data uji terklasifikasi secara tepat. Model ini membagi tingkat pemahaman ke dalam dua kategori, yakni Tinggi dan Sangat Tinggi. Kategori Sangat Tinggi mencatatkan precision 83,33% dan recall 100%, sementara kategori Tinggi memperoleh precision 100% dan recall 85%. Temuan ini membuktikan efektivitas serta potensi algoritma Naïve Bayes dalam menunjang evaluasi pendidikan yang berbasis data.

Kata kunci: Klasifikasi; Kuis Digital; Naïve Bayes; Pemahaman Mahasiswa; Rapidminer Studio 10.1

Received: 24 Agu 2026; **Reviewed:** 24 Sep 2026; **Accepted**: 26 Okt 2026; **Available Online:** 31 Okt 2026;
@ 2026 Inventor

PENDAHULUAN

Penguasaan materi dalam pembelajaran bahasa, baik Bahasa Indonesia maupun Bahasa Inggris, memerlukan pemahaman konsep bertahap mulai dari tingkat kosakata hingga implementasi praktis dalam komunikasi. Tingkat pemahaman mahasiswa yang beragam membuat kompetensi akademik menjadi heterogen, sehingga sulit dievaluasi secara komprehensif hanya melalui ujian akhir semester konvensional. Oleh karena itu, tenaga pendidik memerlukan instrumen evaluasi alternatif yang mampu memetakan pemahaman secara terstruktur. Pemanfaatan teknologi digital ini memberikan peluang luas untuk

menciptakan evaluasi pembelajaran yang lebih interaktif dan berorientasi data (Rahmawati et al., 2025; Saputra, 2023).

Perkembangan teknologi dalam dunia pendidikan mendorong meluasnya penggunaan platform penilaian digital—seperti Google Form, Quizizz, Kahoot, dan Moodle—sebagai media evaluasi kontemporer yang efisien sekaligus efektif dalam meningkatkan partisipasi aktif mahasiswa (Saputra, 2023; Sirait et al., 2024). Di samping itu, penggunaan platform interaktif semisal Wordwall turut dikembangkan guna mengoptimalkan efektivitas proses pembelajaran (Wahyuni et al., 2023), yang dipadu dengan inovasi kecerdasan buatan generatif serta elemen gamifikasi untuk mendukung personalisasi belajar (Farida et al., 2026). Meskipun demikian, luaran dari kuis digital ini kerap kali sebatas menghasilkan angka acuan atau nilai akhir semata, padahal data respons yang terkumpul menyimpan wawasan berharga terkait pola pemanfaatan media serta kedalaman pemahaman konseptual peserta didik. Ekstraksi informasi berharga ini dapat dilakukan dengan menerapkan metode Knowledge Discovery in Databases (KDD) yang meliputi alur seleksi data, prapemrosesan, transformasi, ekstraksi data (data mining), hingga tahap evaluasi dan interpretasi hasil.

Pada tahap data mining, algoritma Naïve Bayes dikenal sebagai salah satu metode klasifikasi yang tangguh berlandaskan perhitungan probabilitas dan asumsi independensi bersyarat antarvariabel. Algoritma ini telah diimplementasikan secara luas pada berbagai riset, seperti estimasi risiko mahasiswa mengulang mata kuliah (Irfan et al., 2024), klasifikasi peminatan program studi (Munawirah & Arisha, 2025), prediksi prestasi akademik (Sivi et al., 2023), hingga evaluasi peran asisten virtual pada mata kuliah kecerdasan buatan (Sari et al., 2025). Selain itu, riset evaluasi tingkat kepuasan terhadap metode pengajaran dosen juga memanfaatkan Naïve Bayes berbasis data kuesioner (Ramadani et al., 2024), di samping komparasinya dengan algoritma Random Forest untuk memprediksi capaian akademik (Kurniasari et al., 2024).

Dalam konteks klasifikasi di bidang pendidikan, metode lain seperti Decision Tree turut diterapkan guna memetakan tingkat pemahaman siswa (Lestari & Lestari, 2024), sementara algoritma C4.5 digunakan untuk menganalisis kecenderungan minat studi lanjut (Ayunda et al., 2024) serta penetapan kategori Uang Kuliah Tunggal (Nursafitri & Ramadhan, 2024). Di luar ranah pendidikan, algoritma Naïve Bayes juga terbukti aplikatif pada pengelompokan data konsumsi obat-obatan (Nurmaleni & Masdalipa, 2025), penentuan kelayakan penerima bantuan iuran jaminan kesehatan (Purba, 2025), analisis sentiment publik (Prasetyo et al., 2023), serta pengategorian bidang ilmu publikasi terindeks SINTA (Nurnasikha, 2023). Meskipun penerapannya sangat variatif di berbagai domain, efektivitas dan tingkat keakuratan model klasifikasi ini sangat dipengaruhi oleh distribusi karakteristik data serta penetapan kelas target yang digunakan.

Perbedaan penelitian ini dengan studi terdahulu terletak pada implementasi algoritma Naïve Bayes untuk mengklasifikasikan tingkat pemahaman mahasiswa secara spesifik berdasarkan instrumen kuesioner penggunaan media kuis digital dalam pembelajaran bahasa. Berbeda dari pengembangan Smart Tutor berbasis AI yang berfokus pada keterlibatan dan pemahaman konsep secara langsung (Hajar et al., 2026), penelitian ini memfokuskan analisis pada pola respons kuesioner media kuis digital menggunakan algoritma Naïve Bayes.

Penelitian ini melibatkan 173 mahasiswa Institut Teknologi dan Kesehatan Ika Bina dari empat program studi (D3 Kebidanan, D3 Keperawatan, S1 Sistem Informasi, dan S1 Teknologi Informasi) dengan menggunakan 27 butir instrumen kuesioner terkait platform Quizizz, Kahoot, Google Form, dan Moodle. Pengolahan data dilakukan melalui perangkat lunak RapidMiner Studio 10.1. Hasil riset ini diharapkan mampu memperkaya literatur pemanfaatan Naïve Bayes dalam evaluasi pembelajaran sekaligus menjadi acuan bagi dosen dalam merumuskan strategi pengajaran yang adaptif.

Rumusan masalah penelitian ini adalah: (1) bagaimana pola karakteristik penggunaan kuis digital pada empat program studi di ITKes Ika Bina; (2) bagaimana performa algoritma Naïve Bayes dalam mengklasifikasikan tingkat pemahaman mahasiswa berdasarkan data interaksi kuis digital menggunakan RapidMiner Studio 10.1; dan (3) sejauh mana akurasi model yang dihasilkan layak dijadikan dasar rekomendasi strategi pembelajaran.

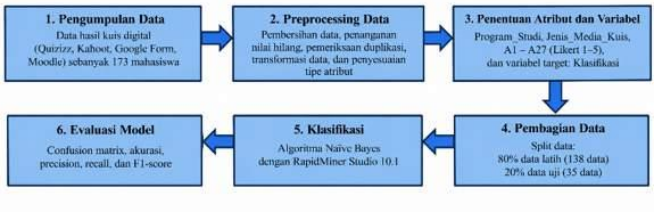
Tujuan penelitian ini adalah mengembangkan dan mengevaluasi model klasifikasi tingkat pemahaman mahasiswa pada pembelajaran bahasa berbasis kuis digital menggunakan algoritma Naïve Bayes yang diimplementasikan pada RapidMiner Studio 10.1, dengan objek penelitian ITKes Ika Bina, Kabupaten Labuhanbatu, Sumatera Utara.

Manfaat penelitian secara teoretis adalah memperkaya literatur mengenai penerapan algoritma Naïve Bayes dalam klasifikasi tingkat pemahaman mahasiswa pada konteks penggunaan kuis digital dalam pembelajaran bahasa. Secara praktis, hasil penelitian ini diharapkan dapat menjadi instrumen pendukung keputusan bagi dosen di Institut Teknologi dan Kesehatan IKA Bina untuk memetakan tingkat pemahaman mahasiswa secara lebih sistematis berdasarkan data yang tersedia. Informasi klasifikasi tersebut juga dapat dimanfaatkan sebagai bahan pertimbangan dalam mengevaluasi proses pembelajaran dan merancang strategi pembelajaran yang lebih adaptif sesuai dengan kebutuhan mahasiswa.

METODE

2.1 Jenis dan Alur Penelitian

enelitian ini menerapkan pendekatan kuantitatif terapan dengan tahapan yang meliputi pengumpulan data, pemeriksaan serta *preprocessing*, pemodelan Naïve Bayes, hingga evaluasi model. Penelitian bertempat di ITKes Ika Bina, Kabupaten Labuhanbatu, Sumatera Utara, pada semester Ganjil tahun akademik 2026/2027. Bagan Metodologi Penelitian dapat dilihat pada Gambar 1.



Gambar 1. Bagan Metodologi Penelitian

2.2 Populasi dan Sampel

Populasi penelitian berjumlah 173 mahasiswa ITKes Ika Bina yang terdistribusi ke dalam program studi D3 Kebidanan (107 orang), D3 Keperawatan (30 orang), S1 Sistem Informasi (19 orang), dan S1 Teknologi Informasi (17 orang). Teknik *total sampling* diterapkan dengan menjadikan seluruh populasi sebagai sampel penelitian.

2.3 Instrumen dan Variabel Penelitian

Dataset bersumber dari kuesioner penggunaan media kuis digital (Google Form, Quizizz, Kahoot, Moodle) yang mencakup atribut Program Studi, Jenis Media Kuis, 27 butir pertanyaan (A1–A27) dengan skala Likert 1–5, serta variabel target Klasifikasi. Uji validitas *Pearson item-total* terkoreksi pada 173 responden mengonfirmasi seluruh 27 butir valid (r hitung > r tabel sekitar 0,149; $p < 0,05$), sementara nilai *Cronbach's Alpha* sebesar 0,915 menunjukkan konsistensi internal yang sangat tinggi (Anshari et al., 2024; Rachmawati & Bayu Pradana, 2025).

2.4 Pengumpulan dan Preprocessing

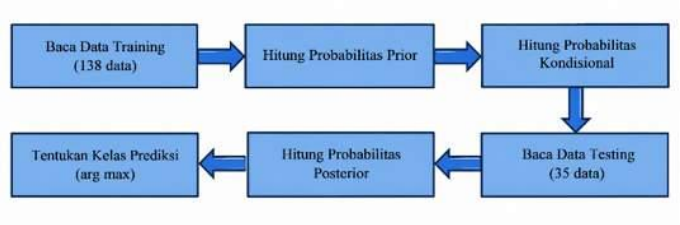
Tabel 1. Data Penelitian (Representasi dari 173 Data)

No	Program_Studi	Jenis_Media_Kuis	A1	A2	A3	…	A27	Klasifikasi
1	S1 Sistem Informasi	Quizizz	2	2	2	…	2	Tinggi
2	D3 Kebidanan	Quizizz	3	3	4	…	3	Sangat Tinggi
3	D3 Keperawatan	Quizizz	3	3	3	…	2	Tinggi
4	D3 Kebidanan	Google Form	2	3	2	…	2	Tinggi
5	D3 Kebidanan	Google Form	2	3	3	…	3	Tinggi
…	…	…	…	…	…	…	…	…
173	D3 Kebidanan	Quizizz	3	3	3	…	3	Sangat Tinggi

Dapat dilihat pada Tabel 1, hasil *preprocessing* dari total 173 responden, kategori klasifikasi terbagi menjadi dua kelompok, yaitu kategori Tinggi sebanyak 113 responden (65,3%) dan Sangat Tinggi sebanyak 60 responden (34,7%).

2.5 Pemodelan Naïve Bayes

Dalam studi kasus Data Mining Metode Naïve Bayes untuk klasifikasi pemahaman mahasiswa, proses segmentasi menggunakan algoritma ini dapat diilustrasikan dalam kerangka metode penelitian. Langkah-langkah penyelesaian untuk menentukan tingkat pemahaman mahasiswa berdasarkan kuesioner penggunaan media kuis digital dengan menggunakan Algoritma Naïve Bayes dapat dilihat pada Gambar 2.



Gambar 2. Proses Metode Naïve Bayes Classfication

Tahapan proses perhitungan menggunakan algoritma Naïve Bayes untuk klasifikasi tingkat pemahaman mahasiswa berdasarkan kuesioner penggunaan media kuis digital adalah sebagai berikut:

A. Baca Data Training
Tahap pertama adalah membaca data *training* sebanyak 138 *dataset*. Data ini akan digunakan sebagai dasar untuk melatih model atau algoritma yang dipilih. Analisis data ini akan mengungkap pola dan tren penting yang akan membantu dalam mengoptimalkan kinerja model pada tahap selanjutnya.

B. Hitung jumlah dan probabilitas
Pada tahap ini, probabilitas setiap kategori atau variabel dihitung untuk memahami distribusi data secara relatif. Langkah ini mencakup perhitungan jumlah dan probabilitas prior, serta menghitung probabilitas training.

1. Hitung jumlah dan probabilitas *prior*

$$P(Ci) = \frac{Si}{s} \quad (2.1)$$

Si adalah jumlah data training dari kategori Ci, dan s adalah jumlah total data training. Menghitung P(Xi|Ci) yang merupakan probabilitas posterior Xi dengan syarat C.

2. Hitung probabilitas training

$$P(Ci) = \frac{P(D|K)*P(K)}{P(D)} \quad (2.2)$$

Dimana:
K = Data dengan *class* yang belum diketahui
D = Hipotesis data merupakan suastu *class* spesifik
P (K|D) = Probabilitas hipotesis D berdasarkan K (*posterior* probabilitas)
P (K) = Probabilitas hipotesis D (*prior* probabilitas)
P (K|D) = Probabilitas K berdasarkan kondisi pada hipotesis D
P (K) = Probabilitas K

C. Baca Data *Testing*
Data testing yang digunakan dalam penelitian ini adalah sebanyak 35 data.

D. Menghitung Probabilitas Data *Testing*
Berdasarkan data training untuk mendapatkan nilai probabilitas dilakukan proses perhitungan menggunakan rumus persamaan (2.3) sebagai berikut :

$$P(X_2/C)P(X_3/C) \ldots P(C/X_1, \ldots X_n) = P(X_1/C) = \pi P^{(Xi|C)} \quad (2.3)$$

E. Menghitung Probabilitas Akhir
Probabilitas akhir merupakan pemaksimalan prediksi dengan rumus persamaan (2.4) sebagai berikut :

$$P(K|D) = P(D|K) * P(K|D) \quad (2.4)$$

F. Probabilitas Akhir
Selanjutnya membandingkan nilai probabilitas dengan menggunakan rumus persamaan (2.5) Sehingga dapat diketahui nilai probabilitas akhir.

$$C = arg\ max\ P(D|K) \quad (2.5)$$

Probabilitas akhir diperoleh dengan menghitung nilai probabilitas akhir kelas menggunakan rumus 2.4. Setelah mendapatkan probabilitas akhir, langkah terakhir adalah normalisasi, yaitu membagikan nilai probabilitas suatu kategori dengan jumlah nilai probabilitas dari semua kategori.

HASIL DAN PEMBAHASAN

Hasil penelitian berdasarkan 173 dataset, sebelum menggunakan metode Naïve Bayes, dilakukan perhitungan awal. Persentase kategori tanpa metode Naïve Bayes adalah 34,68% untuk "Sangat Tinggi", 65,32% untuk "Tinggi", dan 0% untuk "Sedang ", "Rendah", dan "Sangat Rendah". Selanjutnya perhitungan menggunakan metode Naïve Bayes, dataset dibagi menjadi 138 data training dan 35 data testing.

Sebelum penerapan algoritma Naïve Bayes pada 173 dataset, distribusi kategori tanpa model menunjukkan persentase 34,68% untuk "Sangat Tinggi", 65,32% untuk "Tinggi", serta 0% untuk kategori "Sedang", "Rendah", dan "Sangat Rendah". Setelah dataset dibagi menjadi 138 data latih dan 35 data uji, perhitungan probabilitas prior menghasilkan nilai $P(C_{Sangat\ Tinggi})$ = 0,326 dan $P(C_{Tinggi})$ = 0,674.

3.1 Perhitungan Metode Naive Bayes

A. Baca Data Training
Langkah awal adalah membaca data training yang terdiri dari 138 dataset.

B. Hitung Jumlah dan Probabilitas
Langkah selanjutnya adalah menghitung jumlah dan probabilitas untuk setiap kategori dalam dataset tersebut.

1. Perhitungan probabilitas *prior (P(Ci))*
Proses perhitungan probabilitas prior pada kelas Sangat Tinggi, dan Tinggi dengan menggunakan rumus persamaan (2.1) sebagai berikut:

$$P\left(C_{Sangat\ Tinggi}\right) = \frac{45}{138} = 0{,}326$$

$$P\left(C_{Tinggi}\right) = \frac{93}{138} = 0{,}674$$

2. Selanjutnya menghitung probabilitas pada masing-masing kategori menggunakan rumus persamaan (2.2) berikut:
   a. Parameter Kemudahan Pengguna
   Kategori A1 yaitu Saya merasa kuis digital mudah digunakan

$$P\left(5|_{Sangat\ Tinggi}\right) = \frac{0}{45} = 0.000$$

$$P\left(5|_{\ Tinggi}\right) = \frac{0}{93} = 0.000$$

$$P\left(\ 4|_{Sangat\ Tinggi}\right) = \frac{17}{45} = 0.378$$

$$P\left(\ 4|_{\ Tinggi}\right) = \frac{0}{93} = 0.000$$

$$P\left(3|_{Sangat\ Tinggi}\right) = \frac{27}{45} = 0.600$$

$$P\left(3|_{\ Tinggi}\right) = \frac{55}{93} = 0.591$$

$$P\left(2|_{Sangat\ Tinggi}\right) = \frac{0}{45} = 0.000$$

$$P\left(2|_{\ Tinggi}\right) = \frac{35}{93} = 0.376$$

$$P\left(1|_{Sangat\ Tinggi}\right) = \frac{1}{45} = 0.0220$$

$$P\left(1|_{\ Tinggi}\right) = \frac{3}{93} = 0.032$$

Tabel 2. Hasil Perhitungan Probabilitas Kategori A1

Kategori	Kategori Dataset	Sangat Tinggi	Tinggi	P.Sangat Tinggi	P.Tinggi
A1	5	0	0	0.000	0.000
	4	17	0	0.378	0.000
	3	27	55	0.600	0.591
	2	0	35	0.000	0.376
	1	1	3	0.022	0.032

Selanjutnya perhitungan dengan proses yang sama dilakukan berdasarkan kategori yang digunakan sebanyak 27 katagori.

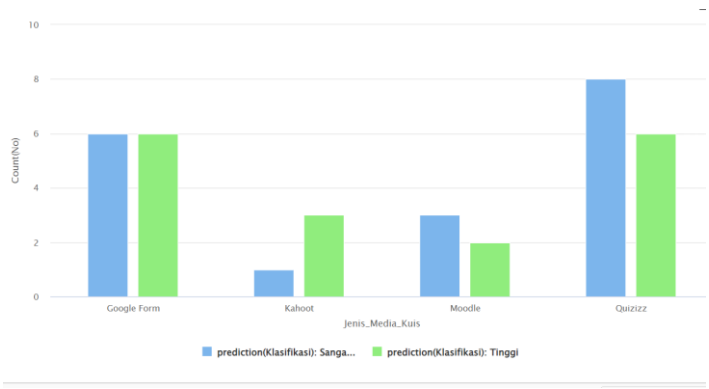
C. Baca data testing
Data testing yang digunakan dalam penelitian ini terdiri dari 35 dataset.

D. Menghitung probabilitas data testing
Berdasarkan data testing, berikut adalah perhitungan nilai untuk data 139 sampai data 173 yang terdiri dari 35 dataset menggunakan rumus persamaan (2.3) sebagai berikut :

Data testing ke-1 memiliki atribut: A1=4, A2=4, A3=5, A4=5, A5=5, A6=5, A7=5, A8=5, A9=4, A10=4, A11=4, A12=5, A13=5, A14=4, A15=3, A16=4, A17=5, A18=5, A19=4, A20=4, A21=3, A22=3, A23=3, A24=3, A25=3, A26=4, dan A27=3.. Selanjutnya, dilakukan pengujian pada semua atribut tersebut.

1) Hasil pengujian "Sangat Tinggi":
= 0.378 * 0.644 * 0.667 * 0.289 * 0.578 * 0.311 * 0.289 * 0.600 * 0.644 * 0.333 * 0.689 * 0.689 * 0.600 * 0.622 * 0.778 * 0.289 * 0.667 * 0.600 * 0.600 * 0.489 * 0.689 * 0.689 * 0.689 * 0.511 * 0.689 * 0.244 * 0.578
= 0.0000000238
2) Hasil pengujian "Tinggi":
= 0.000 * 0.505 * 0.634 * 0.000 * 0.602 * 0.000 * 0.000 * 0.538 * 0.538 * 0.000 * 0.602 * 0.624 * 0.602 * 0.581 * 0.581 * 0.000 * 0.548 * 0.570 * 0.570 * 0.516 * 0.613 * 0.548 * 0.613 * 0.570 * 0.559 * 0.000 * 0.527
= 0

Setelah menyelesaikan perhitungan probabilitas untuk data testing ke-139, proses tersebut dilanjutkan hingga mencapai data testing ke-173, dengan total 35 data testing. Hasil perhitungan ini memberikan nilai probabilitas terkait untuk setiap data testing. Rinciannya dapat ditemukan dalam Gambar 3. Hasil Grafik Visualisasi.



Gambar 3. Hasil Grafik Visualisasi

Berdasarkan Gambar 3, dapat diamati bahwa hasil prediksi tingkat pemahaman mahasiswa berdasarkan jenis media kuis digital terbagi menjadi dua kategori, yaitu Sangat Tinggi yang ditampilkan dengan batang berwarna biru dan Tinggi yang ditampilkan dengan batang berwarna hijau. Pada media Google Form, kedua kategori masing-masing memiliki 6 prediksi. Pada Kahoot, terdapat 1 prediksi Sangat Tinggi dan 3 prediksi Tinggi. Pada Moodle, terdapat 3 prediksi Sangat Tinggi dan 2 prediksi Tinggi. Sementara itu, pada Quizizz, terdapat 8 prediksi Sangat Tinggi dan 6 prediksi Tinggi. Visualisasi tersebut

menggambarkan distribusi hasil klasifikasi tingkat pemahaman mahasiswa berdasarkan jenis media kuis digital menggunakan metode Naïve Bayes.

3.2 Hasil Klasifikasi Naïve Bayes

1. Proses Pengujian Data Dengan *Software* Rapidminer Studio 10.1
   Pada tahap ini, akan dijelaskan tahapan-tahapan proses penggunaan algoritma *Naïve Bayes* pada RapidMiner 10.1 dari data yang telah diimport. Informasi lebih lanjut dapat ditemukan pada Gambar 4.



Gambar 4. Tampilan Melakukan Proses Pengujian Algoritma Naïve Bayes

2. Keluaran sistem (*output*) Rapidminer studio 10.1
   Pada tahap ini, ditampilkan hasil akhir pengolahan data menggunakan metode Naïve Bayes pada perangkat lunak RapidMiner Studio 10.1. Hasil pengujian menunjukkan tingkat akurasi sebesar 91,43%. Berdasarkan confusion matrix pada Gambar 5, terdapat 15 data yang diprediksi sebagai kelas "Sangat Tinggi" dan sesuai dengan kelas aktual, serta 17 data yang diprediksi sebagai kelas "Tinggi" dan sesuai dengan kelas aktual. Selain itu, terdapat 3 data dengan kelas aktual "Tinggi" yang diprediksi sebagai "Sangat Tinggi", sedangkan tidak terdapat data kelas aktual "Sangat Tinggi" yang diprediksi sebagai "Tinggi". Dengan demikian, hasil pengujian menunjukkan bahwa model Naïve Bayes mampu mengklasifikasikan tingkat pemahaman mahasiswa berdasarkan hasil kuis digital dengan tingkat akurasi sebesar 91,43%.

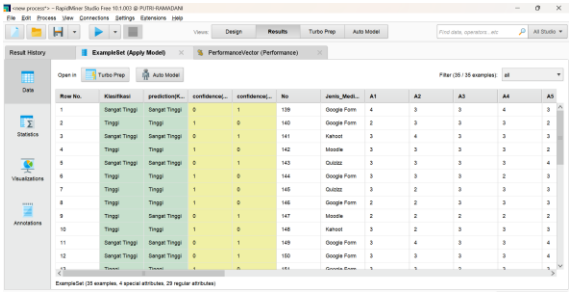
Table View Plot View

accuracy: 91.43%

	true Sangat Tinggi	true Tinggi	class precision
pred. Sangat Tinggi	15	3	83.33%
pred. Tinggi	0	17	100.00%
class recall	100.00%	85.00%	

Gambar 5. Pemodelan Proses Data Testing

3. Hasil Pengujian Probabilitas *Confidence*
   Berdasarkan Gambar 6, hasil penerapan metode Naïve Bayes menggunakan perangkat lunak RapidMiner Studio 10.1 menunjukkan klasifikasi tingkat pemahaman mahasiswa berdasarkan hasil kuis digital ke dalam kategori "Sangat Tinggi" dan "Tinggi". Tabel hasil Apply Model menampilkan kelas aktual, kelas prediksi, serta nilai confidence yang menunjukkan tingkat keyakinan model terhadap hasil klasifikasi setiap data. Hasil tersebut memperlihatkan bahwa model memberikan prediksi untuk setiap data testing berdasarkan atribut yang digunakan, sehingga dapat diketahui kategori tingkat pemahaman mahasiswa yang diprediksi oleh model Naïve Bayes.



Gambar 6. Tampilan Hasil Probabilitas Confidence

3.3 Pembahasan

Berdasarkan metrik evaluasi klasifikasi, keseluruhan 15 subjek aktual pada kategori Sangat Tinggi berhasil teridentifikasi secara tepat (recall 100%), meski presisinya terkoreksi menjadi 83,33% akibat masuknya 3 data positif palsu dari kategori Tinggi. Sebaliknya, kategori Tinggi mencapai presisi absolut (100%) dengan recall sebesar 85%, mengingat 3 dari 20 mahasiswa pada kelompok ini salah diproyeksikan ke kelas Sangat Tinggi. Secara umum, model membuktikan kapabilitas klasifikasi yang mumpuni, namun tetap rentan terhadap galat prediksi yang dipicu oleh kesamaan karakteristik pengisian instrumen pada 27 butir pernyataan. Tingginya derajat kemiripan respon antarmahasiswa antarkelas menjadi faktor pembatas yang mengurangi ketajaman algoritma dalam memisahkan kedua kategori secara definitif.

Dengan capaian akurasi mencapai 91,43%, model Naïve Bayes membuktikan keandalannya melalui klasifikasi benar pada 32 dari total 35 entitas pengujian. Kinerja ini memperlihatkan bahwa informasi laten pada atribut kuesioner memiliki daya pembeda yang cukup kuat untuk mendiferensiasi strata Tinggi dan Sangat Tinggi. Representasi kedua kategori juga telah terserap selama tahap pelatihan melalui keterlibatan 93 data kelas Tinggi serta 45 data kelas Sangat Tinggi, terlepas dari adanya asimetri frekuensi antarspesimen. Walaupun begitu, hubungan langsung antara capaian akurasi dengan kekhasan fitur atau distribusi kelas belum dapat dipastikan secara konklusif tanpa adanya investigasi lebih lanjut. Oleh sebab itu, temuan ini baru menegaskan kapasitas potensial Naïve Bayes dalam mengevaluasi tingkat pemahaman berbasis instrumen kuis digital, bukan jaminan stabilitas luaran pada dataset di luar subjek penelitian.

Keterbatasan penelitian ini mencakup asumsi independensi bersyarat pada Naïve Bayes (di mana aspek kemudahan, interaktivitas, dan motivasi belajar berpotensi saling berkaitan), penggunaan skenario pembagian data tunggal (80:20) dengan 35 data uji, serta cakupan sampel yang terbatas pada 173 mahasiswa dari satu institusi tunggal. Oleh karena itu, riset lanjutan disarankan menerapkan validasi silang (cross-validation), membandingkan performa dengan algoritma lain seperti Decision Tree atau K-Nearest Neighbors (KNN), serta memperluas cakupan institusi.

SIMPULAN

Implementasi algoritma Naïve Bayes memanfaatkan peranti lunak RapidMiner Studio 10.1 pada instrumen data kuesioner kuis digital terbukti mampu memetakan tingkat pemahaman dari 173 mahasiswa di ITKes Ika Bina Labuhanbatu, dengan perolehan akurasi mencapai 91,43%. Melalui pengujian terhadap 35 data sampel uji, sebanyak 32 data berhasil terprediksi dengan presisi ke dalam dua kelompok, yaitu kelas Tinggi (mencakup 113 mahasiswa atau 65,3%) dan kelas Sangat Tinggi (mencakup 60 mahasiswa atau 34,7%). Evaluasi performa pada kategori Sangat Tinggi menghasilkan nilai *precision* 83,33% serta *recall* 100%, sedangkan pada kelompok Tinggi diperoleh *precision* 100% dan *recall* 85%. Implikasi temuan ini menegaskan keandalan data kuis digital sebagai instrumen pendukung evaluasi pembelajaran, kendati penafsirannya tetap wajib diselaraskan secara objektif dengan penilaian akademis utama dari pengajar.

Dalam konteks penerapannya di lingkungan ITKes Ika Bina, luaran dari model klasifikasi ini berfungsi sebagai rujukan pertimbangan bagi para dosen untuk memetakan pemahaman mahasiswa sekaligus menentukan topik materi yang memerlukan pengayaan. Di samping itu, data ini memperkaya evaluasi efektivitas penggunaan platform interaktif seperti Quizizz, Google Form, Kahoot, maupun Moodle sepanjang proses perkuliahan. Kendati demikian, integrasi hasil ini harus diterapkan secara bijak karena prediksi algoritmik tidak menggeser kedudukan asesmen mandiri maupun penilaian langsung dosen terhadap peserta didik.

Sebagai acuan riset mendatang, evaluasi model disarankan menerapkan skema *stratified k-fold cross-validation* demi menjamin konsistensi performa tanpa bergantung pada satu pola pemisahan data saja. Perluasan cakupan sampel yang melibatkan mahasiswa lintas angkatan hingga perguruan tinggi lain juga krusial untuk menguji daya generalisasi model terhadap ragam karakteristik responden yang lebih heterogen. Di samping itu, komparasi performa Naïve Bayes terhadap algoritma klasifikasi lain seperti *Decision Tree* atau *K-Nearest Neighbors* (KNN) dapat diuji menggunakan dataset serta kriteria evaluasi serupa. Optimasi parameter seperti pengaturan *Laplace correction* di RapidMiner Studio 10.1 beserta penyajian metrik evaluasi komprehensif seperti kurva ROC dan nilai AUC jika relevan sangat direkomendasikan guna memperkuat akurasi serta keterulangan (*reproducibility*) penelitian.

DAFTAR PUSTAKA

Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis Validitas dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI. *EDUKATIF: Jurnal Ilmu Pendidikan*, *6*(1), 964–975. https://doi.org/10.31004/edukatif.v6i1.5931

Ayunda, Y. S., Hartama, D., Lubis, M. R., Gunawan, I., & Rafai, M. (2024). Analisis pola minat siswa lulusan SMU/SMK untuk melanjutkan kuliah dengan menggunakan algoritma C4.5. *Jurnal Terapan Informatika Nusantara*, *4*(9), 581–595. https://doi.org/10.47065/tin.v4i9.4880

Farida, M., Rosadi, M. E., Firdaus, M. I., Agustini, D., & Muttaqin, R. (2026). Rancang bangun platform kuis interaktif berbasis AI generatif dan gamifikasi untuk personalisasi pembelajaran siswa. *Smart Comp: Jurnalnya Orang Pintar Komputer*, *15*(1), 88–99. https://doi.org/10.30591/smartcomp.v15i1.9568

Hajar, S., Wahyudin, Muslim, Mulyadin, & Hardiansyah. (2026). Pengembangan Smart Tutor: Asisten Belajar Berbasis Artificial Intelligence untuk Meningkatkan Pemahaman Konsep dan Keterlibatan Siswa. *INVENTOR: Jurnal Inovasi Dan Tren Pendidikan Teknologi Informasi*, *4*(1). https://doi.org/10.37630/inventor.v4i1.5038

Irfan, D., Ramadani, P., Fitriyani, I. N., Damanik, R. S., & Mahenda Hasibuan, Y. I. (2024). Prediksi risiko mahasiswa mengulang mata kuliah menggunakan algoritma Naïve Bayes. *Jurnal Media Informatika*, *5*(2), 271–278. https://doi.org/10.55338/jumin.v5i2.6086

Kurniasari, D., Hidayah, R. N., Notiragayu, N., Warsono, W., & Nisa, R. K. (2024). Classification models for academic performance: A comparative study of Naïve Bayes and Random Forest algorithms in analyzing University of Lampung student grades. *Jurnal Teknik Informatika (JUTIF)*, *5*(5), 1267–1276. https://doi.org/10.52436/1.jutif.2024.5.5.2066

Lestari, D., & Lestari, S. (2024). Penerapan data mining klasifikasi tingkat pemahaman siswa pada kegiatan belajar mengajar dengan metode Decision Tree (Studi kasus SDN Malaka Jaya 11 Duren Sawit). *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, *5*(2), 1260–1268. https://doi.org/10.35870/jimik.v5i2.662

Munawirah, M., & Arisha, A. O. (2025). Implementasi Naïve Bayes untuk klasifikasi peminatan program studi pada penerimaan mahasiswa baru di Fakultas Ilmu Komputer Unika. *Bulletin of Information Technology*, *6*(3), 218–229. https://doi.org/10.47065/bit.v6i3.2142

Nurmaleni, N., & Masdalipa, R. (2025). Klasifikasi data penggunaan obat dengan algoritma Naïve Bayes di Rumah Sakit Umum Daerah (RSUD) Kota Pagar Alam. *Jurnal Media Informatika*, *6*(2), 1169–1184. https://doi.org/10.55338/jumin.v6i2.5470

Nurnasikha. (2023). Klasifikasi bidang ilmu publikasi ilmiah terindeks SINTA menggunakan metode Naïve Bayes. *Transistor: Elektro Dan Informatika*, *5*(3), 147–154. https://doi.org/10.30659/ei.5.3.147-154

Nursafitri, W. N., & Ramadhan, N. S. (2024). Analisis klasifikasi UKT mahasiswa berdasarkan tingkat penghasilan orang tua menggunakan algoritma C4.5. *INVENTOR: Jurnal Inovasi Dan Tren Pendidikan Teknologi Informasi*, *2*(1), 1–9. https://doi.org/10.37630/inventor.v2i1.1248

Prasetyo, S. D., Hilabi, S. S., & Nurapriani, F. (2023). Analisis sentimen relokasi Ibukota Nusantara menggunakan algoritma Naïve Bayes dan KNN. *Jurnal KomtekInfo*, *10*(1), 1–7. https://doi.org/10.35134/komtekinfo.v10i1.330

Purba, D. E. R. (2025). Analisa data mining dengan Naïve Bayes untuk klasifikasi kelayakan penerima bantuan iuran jaminan kesehatan pada Desa Durian. *Jurnal Multimedia Dan Teknologi Informasi*, *7*(2), 11–23. https://doi.org/10.54209/jatilima.v7i02.1279

Rachmawati, D., & Bayu Pradana, A. (2025). Analisis butir soal mata pelajaran ekonomi: Validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. *Jurnal Pendidikan Ekonomi (JUPE)*, *13*(3), 273–284. https://doi.org/10.26740/jupe.v13n3.p273-284

Ramadani, P., Nurcahyo, G. W., & Hendrik, B. (2024). Penerapan metode Naïve Bayes dalam memprediksi kepuasan mahasiswa terhadap cara pengajaran dosen. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer Dan Manajemen)*, *5*(2), 521–528. https://doi.org/10.30645/kesatria.v5i2.361

Saputra, B. (2023). Systematic literature review: Penggunaan Quizizz sebagai evaluasi pembelajaran di Prodi Matematika Universitas Wanita Internasional. *Jurnal Ilmiah Ilmu Pendidikan*, *6*(1), 382–386. https://doi.org/10.54371/jiip.v6i1.1501

Sari, M., Agustini, D., & Farida, M. (2025). Penerapan algoritma Naive Bayes untuk analisa pengaruh asisten virtual terhadap efektivitas pembelajaran mahasiswa Teknik Informatika dalam mata kuliah kecerdasan buatan. *Smart Comp: Jurnalnya Orang Pintar Komputer*, *14*(4), 951–958. https://doi.org/10.30591/smartcomp.v14i4.9847

Sirait, S., Anim, Rahmadani, E., & Syafitri, E. (2024). Penerapan ice breaker game untuk meningkatkan motivasi belajar siswa. *Transformasi: Jurnal Pendidikan Matematika Dan Matematika*, *7*(2), 265–272. https://doi.org/10.36526/tr.v7i2.3277

Sivi, N. A., Dwiatmoko, F., & Khotimah, S. (2023). Prediksi prestasi belajar mahasiswa menggunakan algoritma Naïve Bayes. *Jurnal Ilmiah Teknik Informatika Dan Komunikasi*, *3*(1), 193–201. https://doi.org/10.55606/juitik.v3i1.1822

Wahyuni, D., Fitrianingsih, N., & Fitriati, I. (2023). Pengembangan Media Pembelajaran Interaktif Menggunakan Wordwall pada Mata Pelajaran TIK untuk Meningkatkan Hasil Belajar Siswa. *INVENTOR: Jurnal Inovasi Dan Tren Pendidikan Teknologi Informasi*, *1*(3), 31–36. https://doi.org/10.37630/inventor.v1i3.1195