

# ANALISIS PERBANDINGAN METODE NAÏVE BAYES DENGAN C45 UNTUK MENGUKUR TINGKAT KEPUASAN MAHASISWA

**Kresna Yudha Bayu Saputra**

Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya  
Email: [kresna.20086@mhs.unesa.ac.id](mailto:kresna.20086@mhs.unesa.ac.id)

**Aries Dwi Indriyanti, S.Kom, M.Kom.**

Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya  
Email: [ariesdwi@unesa.ac.id](mailto:ariesdwi@unesa.ac.id)

## Abstrak

Penelitian ini bertujuan untuk membandingkan klasifikasi Data di Kelompok Informatika Universitas Negeri Surabaya memakai dua teknik berbeda, yaitu Naïve Bayes dan C4.5, untuk mengukur tingkat kepuasan mahasiswa terhadap pembelajaran tatap muka. Kepuasan mahasiswa menjadi indikator penting dalam menilai mutu kesempatan belajar serta efektivitas strategi pengajaran. Data dari survei kepuasan mahasiswa dapat dianalisis lebih mendalam menggunakan algoritma C4.5 dan Naïve Bayes. Dalam desain penelitian ini, 150 mahasiswa akan memberikan jawaban mereka pada kuesioner yang mencakup berbagai aspek terkait pendidikan mereka, seperti kehadiran, partisipasi dalam diskusi, fasilitas, kualitas pembelajaran, dan layanan sistem informasi. Kedua metode tersebut digunakan untuk menganalisis data yang terkumpul guna menentukan nilai f1-score, akurasi, presisi, dan recall dari masing-masing metode tersebut. The research results show that the C4.5 algorithm has an accuracy rate of 97.33%, which is superior in handling student satisfaction data classification compared to Naïve Bayes which has an accuracy of 91.33%, especially on datasets with a data sharing ratio of 20:80. These findings provide valuable insights for educational institutions in choosing appropriate analytical methods for evaluating student satisfaction. Thus, the results of this research can be used as a basis for decision making in an effort to improve the quality of offline learning in the university environment.

**Kata Kunci:** Naïve Bayes, C4.5, kepuasan mahasiswa, pembelajaran luring, analisis komparatif

## Abstract

This study aims to compare two data classification methods, Naïve Bayes and C4.5, in measuring student satisfaction with offline learning in the Informatics Cluster at Universitas Negeri Surabaya. Student satisfaction is a crucial indicator for evaluating the standard of instruction and the efficiency of the methods used. The student satisfaction survey data may be analyzed more thoroughly thanks to the application of the C4.5 and Naïve Bayes algorithms. The research technique entails gathering information from 150 students using questionnaires that address a range of topics related to the educational process, including facilities, instruction quality, involvement in discussions, attendance, and information system services. Both algorithms are used to the given data in order to compare the F1-score, accuracy, precision, and recall of each technique. The findings demonstrate that the C4.5 method outperforms Naïve Bayes, which has an accuracy rate of 91.33%, with a rate of 97.33%, especially in datasets with a 20:80 ratio. ratio of data split. These results offer insightful information to help educational institutions select the best analysis technique for assessing student happiness. Therefore, the study's findings can be used as a foundation for decisions made to raise the standard of offline instruction in academic settings.

**Keywords:** Naïve Bayes, C4.5, student satisfaction, offline learning, comparative analysis

## PENDAHULUAN

Mahasiswa sebagai pemangku dalam konteks pembelajaran tatap muka, minat utama seseorang dalam dunia pendidikan sangat berpengaruh dalam menilai seberapa puas mereka dengan pengalaman belajarnya. Tingkat kepuasan yang tinggi bisa menjadi indikasi bahwa metode pengajaran yang digunakan cukup berhasil. (Suharyat et al., 2022). Institusi pendidikan dapat mengambil langkah yang lebih tepat sasaran untuk meningkatkan pengalaman belajar secara luring/offline serta dapat terus dikenal dan juga berkembang dalam sisi akademis maupun jumlah mahasiswa secara umum (Toresa et al., 2023).

Teknik analisis data dan pengambilan keputusan semakin mendapatkan perhatian. Pemanfaatan algoritma cerdas seperti Naïve Bayes dan C4.5 dalam evaluasi pendidikan menunjukkan bagaimana teknologi dapat digunakan untuk meningkatkan pengalaman belajar mahasiswa. Naive Bayes dan C4.5 adalah dua metode analisis data yang populer dan sering digunakan dalam pemodelan analisis prediktif. Metode analisis Naive Bayes dan C4.5, telah digunakan dalam berbagai konteks untuk memahami perilaku dan kecenderungan. Naive Bayes, berdasarkan teorema Bayes, sering digunakan untuk klasifikasi dan prediksi, sedangkan aturan keputusan bisa diambil menggunakan teknik pohon keputusan C4.5. dari data (Fatmawati & Narti, 2022).

Perbandingan algoritma merupakan perbandingan 2 metode atau lebih buat mengenali algoritma mana yang terbaik dari algoritma tersebut (Yunita & Iksari, 2021). Pemilihan algoritma Naïve Bayes dan algoritma C4.5 didasari karena tingkat akurasi dari kedua algoritma ini relatif tinggi. (Yunita & Iksari, 2021). Dilihat dari penelitian terdahulu menurut (Toresa et al., 2023) metode klasifikasi C4.5 lebih baik digunakan dalam mencari pengklasifikasian data daripada naïve bayes.

Penelitian ini bertujuan untuk memberikan kontribusi pada literatur pendidikan dengan menyediakan analisis komparatif mengenai dua metode data mining, yang dapat membantu pendidik dalam mengukur dan meningkatkan kepuasan mahasiswa terhadap pembelajaran luring. Dengan adanya perkembangan ini, Pemilihan metode ini dilatar belakangi oleh keinginan untuk mengetahui kelebihan dan kelemahan masing-masing metode dalam konteks pengukuran kepuasan mahasiswa. Penggunaan metode ini dapat memberikan wawasan yang berharga terkait dengan tingkat kepuasan mahasiswa (Toresa et al., 2023).

Diharapkan hasil dari penelitian ini akan memberikan wawasan mengenai kecocokan dan efektivitas dari kedua teknik tersebut dalam mengukur kebahagiaan siswa serta menentukan apakah pendekatan algoritma C4.5 lebih

unggul dibandingkan dengan algoritma Naïve Bayes dalam hal ketepatan. Sehingga, institusi pendidikan terutama di Universitas Negeri Surabaya dapat membuat keputusan yang lebih baik berdasarkan analisis data ini untuk meningkatkan kualitas pembelajaran luring dan memenuhi harapan mahasiswa secara lebih efektif.

## METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif. Penelitian kuantitatif deskriptif adalah jenis penelitian yang menjelaskan semua fakta dan karakteristik suatu kelompok tertentu secara sistematis, faktual, dan akurat serta memberikan gambaran yang lebih lengkap untuk menentukan seberapa banyak mahasiswa yang puas dalam pembelajaran luring pada data yang telah dikumpulkan saat ini (Hendrika, 2022).

### Populasi dan Sampel

Kelompok yang dikaji dalam penelitian ini terdiri dari mahasiswa jurusan teknik informatika, sebagaimana diidentifikasi oleh penyelidik. Ada beberapa karakteristik yang ditetapkan yakni: mahasiswa aktif Rumpun Informatika angkatan 2023 dan mahasiswa yang pernah merasakan kelas offline dan online selama masa perkuliahan berlangsung.

### Teknik Analisis Data

Langkah ini mencakup verifikasi kuesioner untuk menguji data yang dikumpulkan darinya; tanggapan yang tidak valid akan diperhitungkan.

#### 1. Uji Validitas

Berikut adalah rumus untuk menghitung tingkat validitas menggunakan perhitungan korelasi product moment Karl Pearson, berikut ini:

$$r_{\text{Hitung}} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{\{n \sum x^2 - (\sum x)^2\} \{n \sum y^2 - (\sum y)^2\}}}$$

Rumus 1

Keterangan:

rHitung: koefisien korelasi r pearson

n: jumlah responden

x: skor item

y: skor total

#### 2. Uji Reliabilitas

Mengukur konsistensi sebuah instrumen dalam mengumpulkan data penelitian dikenal sebagai uji reliabilitas. Untuk melakukan Uji Reliabilitas, perlu menggunakan rumus uji Cronbach's Alpha.. Berikut adalah rumus Alpha Cronbach:

$$\propto \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right)$$

Rumus 2

Keterangan:

N: jumlah item

$\sigma_i^2$ : varian setiap item

$\sigma_t^2$ : varian total

(Sumber: Amanda et al., 2019)

### 3. Pembersihan Data

Langkah pengolahan data melibatkan penghapusan data yang tidak konsisten atau berlebihan dari data mentah yang perlu diklasifikasikan. Proses ini mencakup berbagai tahapan, termasuk penghapusan informasi yang berulang dan penghilangan nilai yang hilang seperti pada lampiran. (Toresa et al., 2023)

### 4. Transformasi Data

Dalam penelitian ini, transformasi data didasarkan pada panduan penarikan kesimpulan dari pengguna indeks kepuasan, yang menggunakan skor rata-rata tingkat kepuasan.

Berikut rumus menghitung skor rata-rata untuk setiap pernyataan:

$$X1 = R1Q1 + R1Q2 + R1Q3, \dots$$

$$X2 = R2Q1 + R2Q2 + R2Q3, \dots$$

$$X3 = R3Q1 + R3Q2 + R3Q3, \dots$$

Dst...

Keterangan:

X=Hasil rata-rata responden

R=Responden

Q=Pernyataan pada setiap responden

Data dapat dilabeli dengan pedoman IKP (Tingkat kebahagiaan pengguna) menggunakan rata-rata penilaian dari tingkat kepuasan yang dikategorikan dalam Tabel 1 dengan cara berikut: (Ratnawati, I 2021):

Tabel 1. Index Kepuasan

| No. | Rata-rata IKP | Klasifikasi Kepuasan |
|-----|---------------|----------------------|
| 1.  | 1,00-1,80     | Sangat Tidak Puas    |
| 2.  | 1,81-2,60     | Tidak Puas           |
| 3.  | 2,61-3,40     | Netral               |
| 4.  | 3,41-4,20     | Puas                 |
| 5.  | 4,21-5,00     | Sangat Puas          |

## HASIL DAN PEMBAHASAN

### Hasil Uji Validitas

Uji validitas dengan menggunakan 16 pernyataan. Dengan taraf signifikansi yang digunakan adalah 0.05 atau 5%, tergantung pada hasil uji validitas faktor tersebut.

penelitian ini menunjukkan bahwa hasilnya valid karena hasil uji validitas menyatakan bahwa  $r_{hitung} > r_{tabel}$ .

### Hasil Uji Reliabilitas

Hasil uji reliabilitas pada variabel yang digunakan pada penelitian seperti di Tabel 2.

Tabel 2. Hasil Uji Reliabilitas

#### Reliability Statistics

| Cronbach's |            |
|------------|------------|
| Alpha      | N of Items |
| 0.834      | 16         |

Koefisien alpha Cronbach untuk 16 item dalam uji reliabilitas faktor yang digunakan dalam studi ini adalah 0.834. Data dianggap dapat dipercaya saat menggunakan metode ini jika nilai alpha Cronbach ( $\alpha$ ) lebih besar dari 0.6 mengindikasikan tingkat konsistensi atau keandalan yang baik antara item-item yang ada dalam konstruk atau skala yang diukur dalam penelitian tersebut. Dengan demikian, dapat disimpulkan bahwa variabel tersebut menunjukkan tingkat keandalan yang memadai untuk digunakan dalam analisis dan interpretasi data penelitian.

### Analisis Hasil Pengujian

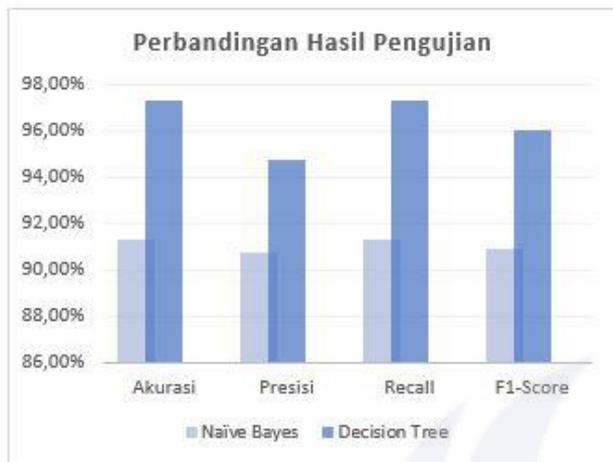
Hasil pengujian algoritma Naïve Bayes dan algoritma Decision Tree C4.5 terdapat pada tabel 3.

Tabel 3. Perbandingan Hasil Pengujian

| Algoritma      | Akurasi | Presisi | Recall | F1-score |
|----------------|---------|---------|--------|----------|
| Naïve Bayes    | 91,33%  | 90,77%  | 91,33% | 90,88%   |
| Algoritma C4.5 | 97,33%  | 94,75%  | 97,33% | 96,02%   |

Gambaran data hasil yang ditampilkan dalam Tabel 4-7, dapat disimpulkan bahwa hasil analisis menggunakan algoritma Decision Tree C4.5 menunjukkan ketepatan yang lebih tinggi, dengan nilai rata-rata 97,33%, skor prediksi 94,75 persen, skor recall 97,33 persen, dan f1-score 96,02 persen. Namun, kinerja algoritma Naïve Bayes juga cukup menonjol dengan tingkat ketepatan 91,33%, prediksi 90,77%, recall 91,33%, dan f1-score 90,88%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma Decision Tree C4.5 berkinerja lebih baik daripada model Naïve Bayes. Perbedaan dalam ketepatan, presisi, recall, dan f1-score antara kinerja kedua algoritma metode Decision Tree C4.5 dengan Naïve Bayes menunjukkan perbedaan sebesar 6%, 3,98%, 6%, dan

5,14% menjelaskan hal ini. Gambar 1 menampilkan perbandingan diagramatis dari hasil uji kedua algoritma tersebut.



Gambar 1. Diagram Perbandingan Hasil Pengujian

## Pembahasan

Hasil penelitian ini terbukti bahwa adanya hubungan antara 5 parameter yang digunakan dengan 16 pernyataan, dengan kepuasan mahasiswa di rumpun informatika Universitas Negeri Surabaya. Hal ini dibuktikan dengan mendapatkan korelasi melalui uji korelasi product moment pearson  $r_{hitung} > r_{tabel}$  dengan taraf signifikansi 5% sebesar  $0,159$  oleh karena itu  $r_{hitung} > r_{tabel}$  artinya ada hubungan yang signifikan antara parameter yang digunakan dengan kepuasan mahasiswa. Hasil dari perhitungan jika ditinjau dari  $r_{tabel}$  koefisien korelasi, maka tingkat hubungan korelasinya tinggi. Hasil penelitian pengolahan data kepuasan mahasiswa di rumpun informatika universitas negeri Surabaya menggunakan metode naïve bayes dan C4.5, bahwa metode C4.5 memiliki akurasi lebih tinggi dari Naïve bayes. Dengan node tertinggi yang dibuat oleh pohon Keputusan C4.5 adalah  $Class = \text{Netral}$ .

## PENUTUP

### Simpulan

Berdasarkan telaah, kesimpulan berikut dapat ditarik mengenai perbandingan antara metode Naïve Bayes dan C4.5:

1. Temuan analisis mengindikasikan bahwa metode Naïve Bayes memiliki nilai akurasi 91,33%, nilai presisi 90,77%, nilai recall 91,33%, dan nilai F1-Score 90,88%, pada dataset yang digunakan. Secara umum, metode ini menunjukkan performa yang cukup baik dalam mengklasifikasikan data tingkat kepuasan mahasiswa meskipun hasil dari akurasi metode Naïve Bayes tidak setinggi algoritma C4.5. Sedangkan hasil analisis metode C4.5 menunjukkan performa yang kompetitif dengan tingkat akurasi 97,33%, nilai presisi

94,75%, nilai recall 97,33%, nilai F1-Score 96,02% pada data yang digunakan, yang menjadikan metode C4.5 lebih efektif untuk klasifikasi tingkat kepuasan mahasiswa pada penelitian ini.

2. Kelebihan utama dari metode Naïve Bayes adalah sangat mudah diimplementasikan dan tidak memerlukan banyak parameter. Metode ini cocok digunakan untuk pemula. Kelemahannya adalah asumsi independensi antar variabel yang kadang tidak realistis dalam situasi nyata. Sedangkan metode C4.5 memiliki kelebihan yaitu hasil yang diperoleh baik pendekatan ini lebih efektif dibandingkan dengan metode Naïve Bayes, terbukti dari nilai ketepatan, presisi, recall, dan f1-score yang lebih tinggi. dalam mengklasifikasikan data dengan lebih baik. Kelemahan pada metode C4.5 ini adalah kompleksitas komputasinya yang lebih tinggi terutama ketika berhadapan dengan dataset yang besar. Secara lebih rinci, hasil analisis menunjukkan bahwa algoritma C4.5 memerlukan lebih banyak waktu untuk memproses dan membangun model dibandingkan dengan Naïve Bayes.

## Saran

Beberapa usulan yang dapat diajukan berdasarkan temuan dan kesimpulan studi adalah sebagai berikut:

1. Dianjurkan untuk mempertimbangkan penggunaan metode hybrid yang mengombinasikan kelebihan dari Naïve Bayes dan C4.5. Hal ini dapat meningkatkan akurasi dan memberikan hasil yang lebih komprehensif dalam pengukuran tingkat kepuasan mahasiswa.
2. Disarankan untuk melakukan penelitian lebih lanjut dengan ukuran sampel yang lebih besar dan variasi metode lain yang dapat dibandingkan seperti KNN, SVM atau metode lainnya. Penelitian lanjutan juga dapat memperluas fokus pada aspek-aspek lain dari pembelajaran yang belum terjangkau oleh penelitian ini.

## DAFTAR PUSTAKA

- Suharyat, Y., Susilawati, T. E., Sikki, M. I., & Nurhayati, S. (2022). Online Learning Quality Evaluation In Higher Education During Covid-19 Pandemic. Society, 10(1), 207–219. <https://doi.org/10.33019/Society.V10i1.406>

- Toresa, D., Hidayat, I., Edriyansyah, E., Muzawi, R., Taslim, T., Lisawita, L., & Yanto, F. (2023). Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Mengukur Tingkat Kepuasan Mahasiswa Dalam Penggunaan Edlink. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(3), 250–256. <https://doi.org/10.47233/Jteksis.V5i3.855>
- Fatmawati, F., & Narti, N. (2022). Perbandingan Algoritma C4.5 Dan Naive Bayes Dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring. *Jtim: Jurnal Teknologi Informasi Dan Multimedia*, 4(1), 1–12. <https://doi.org/10.35746/Jtim.V4i1.196>
- Yunita, D., & Ikasari, I. H. (2021). Perbandingan Metode Klasifikasi C4.5 Dan Naïve Bayes Untuk Mengukur Kepuasan Pelanggan. 6(3), 2622–4615. <https://doi.org/10.32493/Informatika.V6i3.9160>
- Hendrika, D. S. (2022). Gambaran Self Regulated Learning Pada Mahasiswa Di Masa Pandemi Covid-19. *Educativo: Jurnal Pendidikan*, 1(1), 68–74. <https://doi.org/10.56248/Educativo.V1i1.10>
- Amanda, L., Yanuar, F., & Devianto, D. (2019). Uji Validitas Dan Reliabilitas Tingkat Partisipasi Politik Masyarakat Kota Padang. *Jurnal Matematika Unand*, 8(1), 179. <https://doi.org/10.25077/Jmu.8.1.179-188.2019>

