

# Penerapan Algoritma KNN, Naive Bayes Dan C4.5 Dalam Memprediksi Kelulusan Mahasiswa

<sup>1</sup>Embun Fajar Wati, <sup>2</sup>Biktra Rudianto

*Teknik Dan Informatika, Universitas Bina Sarana Informatika<sup>1</sup>*

*Teknik Informatika, Universitas Nusa Mandiri<sup>2</sup>*

*Jl. Kramat Raya No.98, Senen, Jakarta Pusat 10450<sup>1</sup>*

*Jl. Raya Jatiwaringin No.2, RW.13, Cipinang Melayu, Makasar, Jakarta Timur<sup>3</sup>*

embun.efw@bsi.ac.id<sup>1</sup>, biktra.brd@nusamandiri.ac.id<sup>2</sup>

## Abstract

*The selection of new student admissions serves to get qualified students. The quality of students can be measured by the period of education in college. Timely graduates who have the title of quality students. The number of student graduations greatly affects the accreditation assessment for a college. Many factors greatly affect the punctuality of graduation for students such as gender, age, marital status, GPA, and employment status. These variables will be processed with the KNN (K-Nearest Neighbor) algorithm, Naive Baye, C4.5. The preprocessing data uses student data consisting of gender, age, marital status, and employment status as well as GPA. Based on graduation data, 203 students graduated on time and 97 students graduated late. After the transformation is performed, the entire data can be used because there is no empty value. The data changed were age (young : 19 - 24, old : 25 - 50) and GPA (large : 3 - 4, small : 1 - 2.9). The results showed that Naive Bayes had 100.00% accuracy and a 1,000 higher AUC compared to C4.5 and KNN. So the Naive Bayes algorithm has better performance compared to KNN and C4.5.*

**Keyword :** Graduation, Student, KNN, Naive bayes, C4.5

## Abstrak

*Seleksi penerimaan mahasiswa baru berfungsi untuk mendapatkan mahasiswa yang berkualitas. Kualitas mahasiswa dapat diukur dengan masa pendidikan di perguruan tinggi. Lulusan yang tepat waktu yang mempunyai predikat mahasiswa berkualitas. Jumlah kelulusan mahasiswa sangat mempengaruhi penilaian akreditasi bagi suatu perguruan tinggi. Banyak faktor yang sangat berpengaruh terhadap ketepatan waktu kelulusan bagi mahasiswa seperti, jenis kelamin, umur, status pernikahan, IPK, dan status pekerjaan. Variabel-variabel tersebut akan diolah dengan algoritma KNN (K-Nearest Neighbor), Naive Baye, C4.5. Data preprocessing menggunakan data mahasiswa yang terdiri dari jenis kelamin, umur, status pernikahan, dan status pekerjaan juga IPK. Berdasarkan data kelulusan, 203 siswa lulus tepat waktu dan 97 siswa lulus terlambat. Setelah dilakukan transformasi, keseluruhan data dapat digunakan karena tidak ada nilai yang kosong. Data yang diubah adalah umur (muda : 19 - 24, tua : 25 - 50) dan IPK (besar : 3 - 4, kecil : 1 - 2.9). Hasil confusion matrix, menunjukkan bahwa Naive Bayes mempunyai accuracy 100.00% dan AUC 1.000 lebih tinggi dibandingkan dengan C4.5 dan KNN. Sehingga algoritma Naive Bayes mempunyai kinerja yang lebih baik dibandingkan dengan KNN dan C4.5.*

**Keyword :** Kelulusan, Mahasiswa, KNN, Naive bayes, C4.5

## I. PENDAHULUAN

Pendaftaran sebagai calon mahasiswa baru di perguruan tinggi baik negeri maupun swasta semakin meningkat, sehingga harus diadakan seleksi penerimaan mahasiswa baru atau PMB agar mendapatkan mahasiswa yang berkualitas [1]. Implementasi keterampilan akademik yang bersifat profesional yang dapat menciptakan dan mengembangkan IPTEK dan seni bagi mahasiswa dalam bermasyarakat adalah salah satu tujuan dari pendidikan dalam perguruan tinggi [2]. Persaingan antar perguruan tinggi sangat ketat dengan menghasilkan kualitas lulusan mahasiswa yang baik. Kualitas program pendidikan perlu diperbaiki karena merupakan suatu tantangan bagi setiap perguruan tinggi [3]. Kualitas lulusan yang baik dapat dilihat dari masa studi dan waktu kelulusan dengan tepat waktu [4].

Kegiatan akhir pendidikan adalah proses kelulusan yang dialami oleh peserta didik [5]. Institusi membutuhkan penilaian akreditasi yang berproses pada kelulusan mahasiswa di suatu perguruan tinggi [6]. Ketepatan waktu lulus sangat mempengaruhi penilaian akreditasi perguruan tinggi [7]. Perguruan tinggi banyak menyoroti salah satu faktor tentang ketepatan mahasiswa dalam menyelesaikan masa studinya [8]. Masa atau lama studi yaitu rentan waktu yang digunakan mahasiswa untuk penyelesaian pendidikannya [9]. Kelulusan sangat penting dalam evaluasi proses pembelajaran di dalam perguruan tinggi sehingga dapat diperbaiki [10]. Faktor-faktor yang mempengaruhi penyelesaian perkuliahan di perguruan

tinggi, yaitu faktor internal, faktor akademik, faktor tempat tinggal, faktor keluarga, faktor perilaku [11]. Faktor-faktor tersebut berakibat pada masa studi mahasiswa yang melebihi standar kompetensi yaitu untuk S1 selama 4 tahun [12]. Data peserta didik yang didapat dari pendaftaran mahasiswa baru dan data kelulusan yang didapat dari data mahasiswa lama yang sudah menyelesaikan pendidikan ada di dalam institusi [13]. Sedangkan data akademik mahasiswa yang tersimpan pada setiap perguruan tinggi dapat dimanfaatkan untuk diolah menjadi informasi tentang prediksi mahasiswa yang lulus dengan tepat waktu [14]. Data peserta didik yang melimpah seperti data mahasiswa dan data akademik pada universitas dapat dimanfaatkan untuk diolah datanya agar mendapatkan informasi yang tersembunyi seperti prediksi kelulusan mahasiswa [15]. Prediksi variabel target akan didapatkan dari variabel prediktor yang ada di dalam data mahasiswa [16]. Oleh karena itu, diperlukan pengelompokan atau klasifikasi mahasiswa dengan karakteristik lulus tepat waktu dan lulus terlambat, sehingga bisa diberikan tindakan pencegahan untuk mahasiswa yang lulus terlambat [17]. Pengumpulan atau kegiatan untuk mengekstrak atau mengolah data untuk menemukan ilmu pengetahuan baru yang dapat memberikan informasi [18]. Data Mining merupakan suatu kegiatan atau proses pengumpulan suatu data yang kemudian diekstrak atau diolah untuk mendapatkan suatu ilmu pengetahuan yang baru. Pengelompokan dengan data mining akan dilakukan dengan tiga metode yaitu, KNN, Naive Bayes, dan Decision Tree atau yang biasa disebut dengan C4.5. Algoritma KNN berperan untuk mengidentifikasi kasus baru terhadap kasus lama yang memiliki nilai kedekatan mendapatkan nilai kedekatan [19]. KNN juga dapat digunakan dalam pengolahan data angka atau numerik dan menggunakan skema perkiraan parameter perulangan yang tidak rumit, dan bisa digunakan untuk dataset yang besar ukurannya [20]. Efektivitas dan efisiensi dalam metode Naive Bayes dalam pembelajaran induktif dapat berkompetitif dari algoritma klasifikasi lainnya [21]. Di dalam data mining terdapat metode klasifikasi atau algoritma yang sering digunakan untuk memprediksi pola atau model kelulusan mahasiswa yaitu C4.5 [10]. Seperti pada penelitian yang membandingkan ketiga algoritma tersebut dalam pembelajaran daring dan mempunyai hasil yang berbeda, yaitu Naive Bayes 55.49%, KNN 61.47%, C4.5 atau Decision Tree 61.29% [22]. Hasil accuracy tersebut menunjukkan bahwa Naive Bayes adalah algoritma yang terbaik digunakan pada kasus pembelajaran daring. Dengan penggunaan 3 algoritma yaitu Naive Bayes, KNN, dan C4.5 atau Decision Tree diharapkan dapat memprediksi status kelulusan mahasiswa selama perkuliahan, sehingga mahasiswa yang tidak lulus tepat waktu atau terlambat lulus dapat diberikan perhatian lebih oleh prodi dan pembimbing akademik sehingga mahasiswa dapat memperbaiki indeks prestasi agar dapat lulus dengan segera [23].

## II. METODE PENELITIAN

Pada penelitian ini akan digunakan 3 algoritma data mining yaitu KNN (K-Nearest Neighbor), Naive Bayes, dan C4.5 atau yang biasa disebut dengan Decision Tree. Ketiga algoritma tersebut akan dievaluasi dengan membandingkan hasil dari kedua algoritma dengan nilai confusion matrix yaitu nilai accuracy, precision, recall, dan AUC.

### 1. KNN (K-Nearest Neighbor)

KNN (K-Nearest Neighbor) adalah klasifikasi obyek yang sangat dekat jaraknya dan membutuhkan informasi training yang disebut dengan tata cara supervised [9]. KNN mempunyai prinsip kerja antara evaluasi informasi dengan neighbor data pelatihan (k orang sebelah) dengan cara mencari jarak yang paling dekat [24], [25]. Algoritma KNN untuk menghitung jarak yaitu [9] :

$$d_i = \sqrt{\sum_{i=1}^n (x_{ij} - p_j)^2}$$

Keterangan :

$d_i$  = Jarak sampel

$x_{ij}$  = Data sampel pengetahuan

$p_j$  = Data input var ke-j

$n$  = Jumlah sampel

### 2. Naive Bayes

Naive Bayes adalah pencarian nilai probabilitas yang paling baik dan merupakan salah satu dari algoritma jenis klasifikasi [26]. Metode ini dipilih karena mudah diterapkan untuk bekerja secara mandiri, yaitu suatu fitur dalam suatu data tidak berhubungan dengan ada tidaknya fitur lain pada data yang sama [27]. Algoritma Naive Bayes untuk mencari nilai probabilitas yang baik adalah [22] :

$$P(H|X) = \frac{P(H) \cdot P(X|H)}{P(X)}$$

Keterangan [9] :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (Posteriori Probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi hipotesis H
- P(H) : Probabilitas Hipotesis H (Prior Probabilitas)
- P(X) : Probabilitas X

3. C4.5 atau Decision Tree

C4.5 atau Decision Tree yaitu salah satu algoritma dari jenis klasifikasi yang dapat menghasilkan model (pola) pohon keputusan (decision tree) yang mudah dipahami [10]. Algoritma C4.5 atau Decision Tree untuk menghitung information gain adalah [9] :

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

Keterangan :

- S : Himpunan kasus
- A : Atribut
- n : Jumlah patisi S
- Pi : Proporsi dari Si terhadap S
- A|Si| :Jumlah kasus pada partisi ke-i
- |S| : Jumlah kasus dalam S

Entropy (S) yaitu entropy value yang dapat dihitung dengan rumus sebagai berikut [9] :

$$Entropy(S) = \sum_{i=1}^n pi * log_2 pi$$

Keterangan :

- S : Himpunan kasus
- A : Atribut
- n : Jumlah patisi S
- Pi : Proporsi dari Si terhadap S

4. Confusion Matrix

Confusion matrix merupakan informasi yang membandingkan hasil klasifikasi dari sebuah metode. Confusion matrix menghasilkan nilai accuracy, precision, dan recall. Accuracy merupakan jumlah perbandingan total keseluruhan data dengan data yang benar. Precision merupakan tingkat akurasi antara hasil yang diprediksi oleh model dengan data yang diminta. Recall merupakan perbandingan rasio antara keseluruhan data yang benar positif dan data yang diprediksi benar positif [18]. Algoritma untuk perhitungan accuracy, precision dan recall yaitu :

- a. Accuracy = (TP+TN)/(TP+FP+FN+TN)
- b. Precision = TP/(TP+FP)
- c. Recall = TP/(TP+FN)

III. HASIL DAN PEMBAHASAN

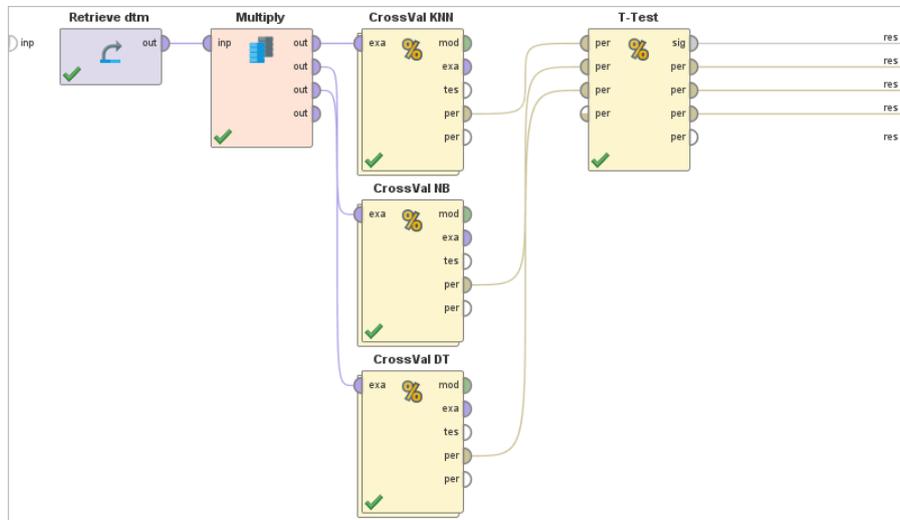
Data yang digunakan dalam penelitian ini sebanyak 300 mahasiswa dari salah satu kampus di Jakarta angkatan 2016-2017. Data mahasiswa terdiri dari 2 kategori yaitu data pribadi (jenis kelamin, umur, status pernikahan, dan status pekerjaan) dan data akademik (IPK). Berdasarkan data kelulusan, 203 siswa lulus tepat waktu dan 97 siswa lulus terlambat. Keseluruhan data dapat digunakan karena tidak ada nilai yang kosong. Data yang diubah adalah umur (muda : 19 - 24, tua : 25 - 50) dan IPK (besar : 3 - 4, kecil : 1 - 2.9). Data yang digunakan untuk pelatihan dan pengujian dengan 10-fold cross-validation yaitu dibagi menjadi 10 bagian dari data atau subset dengan ukuran yang sama. Itu diulang 10 kali dalam proses pelatihan dan pengujian.

Pengolahan data dalam penelitian ini menggunakan software Rapid Miner versi 9.9.000. Untuk variabel input terdiri dari 5 variabel, diantaranya: jenis kelamin, umur, status pernikahan, IPK, dan status pekerjaan. Untuk lebih detailnya dapat dilihat pada tabel 1 berikut.

Tabel 1 Dataset Mahasiswa

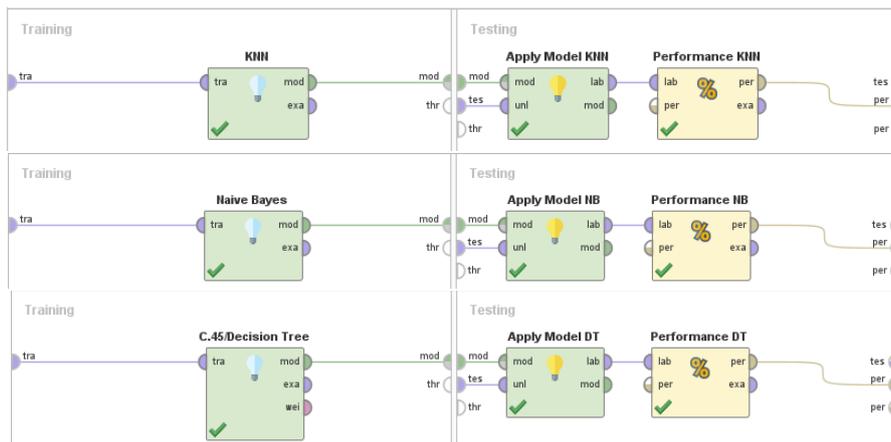
Jenis Kelamin	Umur	Status Pernikahan	IPK	Status Pekerjaan
Laki-Laki	Muda	Menikah	Kecil	Bekerja
Perempuan	Muda	Belum	Kecil	Bekerja
Laki-Laki	Muda	Menikah	Kecil	Tidak
Laki-Laki	Muda	Belum	Kecil	Bekerja

Laki-Laki	Muda	Menikah	Kecil	Tidak
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
Laki-Laki	Tua	Belum	Kecil	Bekerja
Laki-Laki	Muda	Belum	Kecil	Bekerja
Laki-Laki	Muda	Menikah	Kecil	Tidak
Perempuan	Muda	Belum	Kecil	Bekerja
Laki-Laki	Muda	Menikah	Kecil	Tidak



Gambar 1 Usulan Model (Process)

Model proses yang diusulkan berdasarkan dataset mahasiswa pada tabel 1 dengan menggunakan algoritma KNN, Naive Bayes, dan C4.5 atau Decision Tree dapat dilihat pada gambar 1 berikut. Data mahasiswa dilakukan training dan testing yang dapat dilihat pada gambar 2.



Gambar 2 Training dan Testing

Dari hasil uji akhir model didapatkan hasil confusion matrix untuk algoritma KNN pada tabel 2, Naive Bayes pada tabel 3 dan C4.5 atau Decision Tree pada tabel 4.

Tabel 2 Confusion Matrix KNN

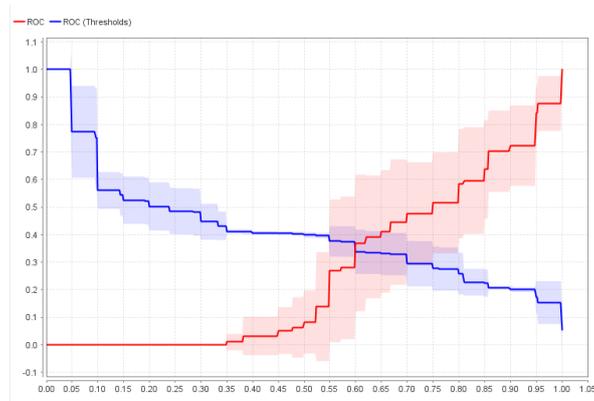
	true Tepat	true Terlambat	class precision
pred. Tepat	169	97	63.53%
pred. Terlambat	34	0	0.00%
class recall	83.25%	0.00%	

Tabel 3 Confusion Matrix Naive Bayes

	true Tepat	true Terlambat	class precision
pred. Tepat	203	0	100.00%
pred. Terlambat	0	97	100.00%
class recall	100.00%	100.00%	

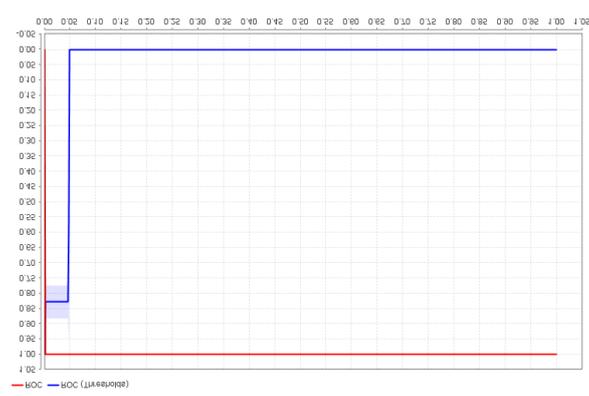
Tabel 4 Confusion Matrix C4.5 atau Decision Tree

	true Tepat	true Terlambat	class precision
pred. Tepat	194	0	100.00%
pred. Terlambat	9	97	91.51%
class recall	95.57%	100.00%	



Gambar 3 Kurva ROC KNN

Hasil kurva ROC dari algoritma KNN dapat dilihat pada gambar 3 yang menunjukkan AUC 0.258. Lebih rendah dibandingkan dengan AUC Naive Bayes dan Decision Tree.



Gambar 4 Kurva ROC Naive Bayes

Hasil kurva ROC dari algoritma Naive Bayes dapat dilihat pada gambar 4 yang menunjukkan AUC 1.000. Paling tinggi dibandingkan dengan AUC KNN dan Decision Tree. Sedangkan hasil kurva ROC dari algoritma Decision Tree dapat dilihat pada gambar 5 yang menunjukkan AUC 0.975. Menempati urutan kedua setelah hasil AUC Naive Bayes.



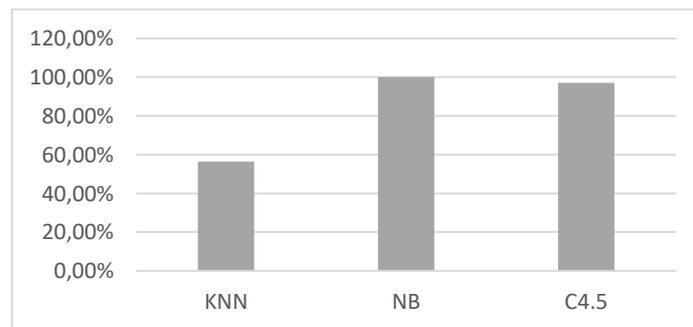
Gambar 5 Kurva ROC C4.5 atau Decision Tree

Hasil confusion matrix ditunjukkan dengan cara membandingkan nilai accuracy, precision, recall, dan AUC dari kurva ROC pada ketiga algoritma (tabel 5).

Tabel 5 Hasil Confusion Matrix

Algorithms	Accuracy	Precision	Recall	AUC
KNN	56.33%	63.53%	83.53%	0.258
NB	100.00%	100.00%	100.00%	1.000
C4.5	97.00%	100.00%	95.57%	0.978

Pada table 5 hasil confusion matrix, dapat disimpulkan bahwa algoritma Naive Bayes mempunyai accuracy 100.00% dan AUC 1.000 lebih tinggi dibandingkan dengan algoritma C4.5 yang menempati urutan kedua dengan accuracy 97.00% dan AUC 0.975. Sedangkan urutan ketiga yaitu algoritma KNN dengan accuracy 56.33% dan AUC 0.258.



Gambar 6 Perbandingan Accuracy Hasil Confusion Matrix

Algoritma Naive Bayes mempunyai kinerja yang lebih baik dibandingkan dengan KNN dan C4.5 atau Decision Tree dan dapat dilihat pada gambar 6. Akurasi dan nilai AUC pada algoritma KNN sangat rendah sehingga mempunyai kinerja yang kurang baik dalam pengolahan data kelulusan mahasiswa.

#### IV. KESIMPULAN

Kesimpulan dari penelitian ini adalah klasifikasi data kelulusan mahasiswa dapat dengan mudah dilakukan dengan ketiga algoritma yaitu KNN, Naive Bayes, dan C4.5 atau Decision Tree yang divalidasi dengan metode 10-fold cross-validation sehingga menghasilkan accuracy tertinggi dari algoritma Naive Bayes yaitu 100.00% . Pada presentase terlihat algoritma Naive Bayes lebih unggul dibandingkan dengan algoritma C4.5 atau Decision Tree dan KNN. Hasil Recall tertinggi yaitu 100%, sedangkan precision dengan nilai tertinggi yaitu 100.00% yang diraih oleh algoritma Naive Bayes. Decision Tree juga mempunyai hasil nilai AUC yang sangat baik yaitu 0.978. Sedangkan KNN mempunyai hasil nilai AUC yang sangat rendah yaitu 0.258 yang berarti bahwa kinerja KNN sangat buruk atau gagal dalam pengolahan data kelulusan mahasiswa. Pada penelitian selanjutnya dengan kasus yang serupa, diharapkan dapat dikembangkan dengan perbandingan beberapa algoritma klasifikasi lainnya selain KNN, Naive Bayes, maupun C4.5 atau Decision Tree.

REFERENSI

- [1] S. P. Nabila, N. Ulinuha, and A. Yusuf, “MODEL PREDIKSI KELULUSAN TEPAT WAKTU DENGAN METODE FUZZY C-MEANS DAN K-NEAREST NEIGHBORS MENGGUNAKAN DATA REGISTRASI MAHASISWA,” *J. Ilm. NERO*, vol. 6, no. 1, pp. 38–46, 2021.
- [2] I. Maulidah, T. T. W, and Q. A’yun, “Perbandingan Algoritma K-Nearest Neighbor (K-Nn) Dan Id3 Untuk Prediksi Kelulusan Mahasiswa Fakultas Teknik Universitas Muhammadiyah Jember,” *J. Smart Teknol.*, vol. 3, no. 4, pp. 374–381, 2022.
- [3] R. H. Sukarna and Y. Ansori, “IMPLEMENTASI DATA MINING MENGGUNAKAN METODE NAIVE BAYES DENGAN FEATURE SELECTION UNTUK PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU,” *J. Ilm. Sains dan Teknol.*, vol. 6, no. 1, pp. 50–61, 2022.
- [4] M. N. YATIMAH, “Implementasi Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa STIMIK ESQ Menggunakan Decision Tree C4.5,” *JUMANJI*, vol. 5, no. 2, pp. 89–98, 2021.
- [5] F. A. Ma’ruf, A. Pratama, I. Sholihin, A. R. Rinaldi, and Faturrhohman, “Penerapan Model Prediksi Menggunakan Algoritma C.45 Untuk Prediksi Kelulusan Siswa SMK Wahidin,” *J. DATA Sci. Inform.*, vol. 1, no. 1, pp. 16–20, 2021.
- [6] L. Y. L. Gaol, M. Safii, and D. Suhendro, “Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5,” *BRAHMANA J. Penerapan Kecerdasan Buatan*, vol. 2, no. 2, pp. 97–106, 2021.
- [7] E. P. Rohmawan, “PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE DECISION TREE DAN ARTIFICIAL NEURAL NETWORK,” *J. Ilm. Matrik*, vol. 20, no. 1, pp. 21–30, 2018.
- [8] S. D. Anggita and Ikamah, “IMPLEMENTASI PSO UNTUK OPTIMASI BOBOT ATRIBUT PADA ALGORITMA C4.5 DALAM PREDIKSI KELULUSAN MAHASISWA,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 6, no. 2, pp. 416–423, 2021.
- [9] Hozairi, Anwar, and S. Alim, “IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES,” *J. Ilm. NERO*, vol. 6, no. 2, pp. 133–144, 2021.
- [10] A. Rohman and S. Mujiyono, “Permodelan Prediksi Predikat Kelulusan Mahasiswa Menggunakan Decision Tree C4.5,” *J. Prodi Tek. Inform. UNW Multimatrix*, vol. 3, no. 2, pp. 1–5, 2021.
- [11] N. M. Suhaimi, “Review on Predicting Students’ Graduation Time Using Machine Learning Algorithms,” *I.J. Mod. Educ. Comput. Sci.*, vol. 7, pp. 1–13, 2019.
- [12] D. A. Wulandari, B. N. Sari, and T. N. Padilah, “Prediction of Student Graduation Accuracy Using C45 Algorithm (Case Study: Fasilkom Unsika),” *SYSTEMATICS*, vol. 4, no. 1, pp. 372–381, 2022.
- [13] D. Nugraheny and A. S. Honggowibowo, “Analysis of the Validity of Determination of Graduation Predicate Based on Student Individual Data at Institut Teknologi Dirgantara Adisutjipto (ITDA),” *J. Multidisiplin Madani*, vol. 2, no. 3, pp. 1067–1082, 2022.
- [14] N. Ramsari and A. R. Firmansyah, “SISTEM PENDUKUNG KEPUTUSAN UNTUK MENINGKATKAN MUTU KINERJA PROGRAM STUDI MELALUI PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES BERBASIS FRAMEWORK LARAVEL (STUDI KASUS : FAKULTAS ILMU KOMPUTER DAN INFORMATIKA UNIVERSITAS NURTANIO BANDUNG),” *J. Inform. Teknol. dan Sains FORMATEKS*, vol. 1, no. 1, pp. 6–18, 2022.
- [15] Andik, Syarli, and C. R. Sari, “DATA MINING KLASIFIKASI KELULUSAN MAHASISWA MENGGUNAKAN METODE NAÏVE BAYES,” *J. Pegguruang Conf. Ser.*, vol. 4, no. 1, pp. 423–428, 2022.
- [16] S. Widaningsih, “PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI DAN WAKTU KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA DENGAN ALGORITMA C4.5, NAÏVE BAYES, KNN, DAN SVM,” *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019.
- [17] I. Fadil, R. Goejantoro, and S. Prangga, “Aplikasi K-Nearest Neighbor Dengan Fungsi Jarak Gower Dalam Klasifikasi Kelulusan Mahasiswa (Studi Kasus : Mahasiswa Program Studi Statistika, Jurusan Matematika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Mulawarman),” *J. EKSPONENSIAL*, vol. 13, no. 1, pp. 57–62, 2022.
- [18] R. A. Iswanto, J. Sahertian, and M. A. D. Widyadara, “Pengembangan Sistem Prediksi Kelulusan Mahasiswa Berdasarkan Nilai Disiplin, Hasil Belajar, Aktivitas Sosial Ekonomi, dan Aktivitas Organisasi Menggunakan Algoritma Naïve Bayes,” in *PROSIDING SEMINAR NASIONAL TEKNOLOGI DAN SAINS*, 2022, pp. 349–358.
- [19] L. A. R. Hakim, A. A. Rizal, and D. Ratnasari, “APLIKASI PREDIKSI KELULUSAN MAHASISWA BERBASIS K-NEAREST NEIGHBOR (K-NN),” *J. Teknol. Inf. dan Multimed.*, vol. 1, no. 1, pp. 30–36, 2019.
- [20] E. S. Susanto, Kusri, and H. Al Fatta, “PREDIKSI KELULUSAN MAHASISWA MAGISTER TEKNIK INFORMATIKA UNIVERSITAS AMIKOM YOGYAKARTA MENGGUNAKAN METODE K-NEAREST NEIGHBOR,” *J. Teknol. Inf.*, vol. 13, no. 2, pp. 67–72, 2018.
- [21] J. MANULLANG and J. F. R. PANGGABEAN, “ANALISIS KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *J. Sains dan Teknol. ISTP*, vol. 16, no. 2, pp. 174–179, 2022.
- [22] T. W. Putra, A. Triayudi, and Andrianingsih, “Analisis Sentimen Pembelajaran Daring menggunakan Metode Naïve Bayes, KNN, dan Decision Tree,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 1, pp. 20–26, 2022.
- [23] A. Y. Saputra and Y. Primadasa, “Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan

Algoritma K-Nearest Neighbour,” *Techno.COM*, vol. 17, no. 4, pp. 395–403, 2018.

- [24] I. A. A. Amra and A. Y. A. Maghari, “Students performance prediction using KNN and Naïve Bayesian,” in *8th International Conference on Information Technology (ICIT)*, 2017, pp. 909–913.
- [25] S. H. Rukmawan, F. R. Aszhari, Z. Rustam, and J. Pandelaki, “Cerebral Infarction Classification Using the KNearest Neighbor and Naive Bayes Classifier,” *J. Phys. Conf. Ser.*, vol. 1752, no. 1, p. 012045, 2019.
- [26] M. Alghobiri, “A Comparative Analysis of Classification Algorithms on Diverse Datasets,” *Eng. Technol. Appl. Sci. Res.*, vol. 8, no. 2, pp. 2790–2795, 2018.
- [27] E. F. Wati and B. Sudrajat, “Application of Naive Bayes Method For Diagnosis of Pregnancy Disease,” *Int. J. Inf. Syst. Technol.*, vol. 6, no. 1, pp. 93–100, 2022.