



**MATH
VISION**

JURNAL MATEMATIKA

VOL. 8. No. 1 Maret 2026

Diterbitkan Oleh :
PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM (FMIPA)
UNIVERSITAS PGRI RONGGOLAWE TUBAN



Editorial Team

Editor in Chief:

Ahmad Zaenal Arifin ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas PGRI Ronggolawe

Editor:

Ahmad Hanif Asyhar, ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Islam Negeri Sunan Ampel Surabaya, Indonesia

Kresna Oktafianto ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas PGRI Ronggolawe

Sintia Safrianti ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Prof. Dr. Hazairin SH, Bengkulu

Fitroh Resmi ([Scholar](#) | [Sinta](#)), Politeknik Perkapalan Negeri Surabaya

Eriska Fitri Kurniawati ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas PGRI Ronggolawe

Nia Nurfitri, ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas PGRI Ronggolawe

Desain Grafis

Kuntum Febriyantiningrum, Universitas Negeri Yogyakarta

Web Admin

Sriwulan, Universitas PGRI Ronggolawe

Sekretariat

Isyuari Isti Widya Rani Igati, Universitas PGRI Ronggolawe

Akreditasi



Reviewer

Lilik Muzdalifah ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Jenderal Soedirman

Dian Candra Rini Novitasari, ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Islam Negeri Sunan Ampel Surabaya, Indonesia

Tahiyatul Asfihani ([Scopus](#) | [Scholar](#) | [Sinta](#)), Institut Teknologi Sepuluh Nopember (ITS) Surabaya, Indonesia

Mohammad Iqbal, ([Scopus](#) | [Scholar](#) | [Sinta](#)), Institut Teknologi Sepuluh Nopember (ITS) Surabaya, Indonesia

Muhammad Jakfar, ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Negeri Surabaya, Indonesia

Dila Puspita, ([Scopus](#) | [Scholar](#) | [Sinta](#)), Institut Teknologi Bandung, Indonesia

Indira Anggriani, ([Scopus](#) | [Scholar](#) | [Sinta](#)), Institut Teknologi Kalimantan, Indonesia

Mohammad Tafrikan, ([Scopus](#) | [Scholar](#) | [Sinta](#)), UIN WALISONGO, Indonesia

Uswatun Khasanah, ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Bilfath, Indonesia

Rahmawati Erma Standsyah, ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Negeri Surabaya, Indonesia

Taufan Talib ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Pattimura, Indonesia

Yopi Andri Lesnussa ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Pattimura, Indonesia

Asmianto ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Negeri Malang, Indonesia

Zainullah Zuhri ([Scopus](#) | [Scholar](#) | [Sinta](#)), Universitas Nahdlatul Ulama Pasuruan, Indonesia

Akreditasi



Vol. 8 No. 1 (2026)

DOI: <https://doi.org/10.55719/mv.v8i1>

Published: 2026-03-31

Articles

PELABELAN SISI PRIMA GANJIL PADA BEBERAPA OPERASI GRAF SIKLUS

Hafif Komarullah

1-8

 pdf

ANALISIS PREDIKSI NILAI EKSPOR KOMODITAS ALAS KAKI INDONESIA KE AMERIKA DENGAN PENDEKATAN REGRESI DERET FOURIER NONPARAMETRIK

Utsna Rosalin Maulidya, M. Fariz Fadillah Mardianto, Elly Pusporani, Dita Amelia

9-16

 pdf

CLASSIFICATION OF STUDENT CAREER CHOICES USING THE C4.5 ALGORITHM

Daviana Widya Maurora, Adhi Surya Nugraha

17-27

 pdf

MAPPING CAREER CHOICE OF MATHEMATICS EDUCATION STUDENTS AT SANATA DHARMA UNIVERSITY USING C4.5 DECISION TREE ALGORITHM

Daviana Widya Maurora¹, Adhi Surya Nugraha^{2*}

^{1,2} Program Studi Pendidikan Matematika
Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sanata Dharma
Jl. Paingan, Maguwoharjo, Depok, Sleman, D.I. Yogyakarta, 55282

Email Penulis Korespondensi: * a.s.nugraha@usd.ac.id

ABSTRAK

Article History:

Receive 27 January 2026

Revised 4 February 2026

Accepted 16 March 2026

Available
online date month year

Keywords:

Classification;
Career;
Algorithm;
C4.5;

Career choice is an important decision that students must take to plan steps to develop their expertise and skills to face the world of work. Data mining can be used to facilitate decision-making regarding students' career plans. This study aims to build a classification model for career choices of Mathematics Education students at Sanata Dharma University using the C4.5 algorithm and determine the accuracy level of the classification model built in mapping career choices of Mathematics Education students at Sanata Dharma University. This research is an applied research using the C4.5 algorithm and is carried out based on the CRISP-DM data mining stages which include business/research understanding, data understanding, data preparation, modeling, evaluation, and deployment. Classification is carried out based on the attributes of GPA, total points, gender, average grade of mathematics courses, and average grade of education courses. The classification results show an initial accuracy of 15.52% and an increase of 36.61%, to 52.13% after oversampling.



Artikel ini adalah artikel akses terbuka yang didistribusikan berdasarkan syarat dan ketentuan [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).


Cara mengutip artikel ini:

Daviana Widya Maurora, Adhi Surya Nugraha., "MAPPING CAREER CHOICE OF MATHEMATICS EDUCATION STUDENTS AT SANATA DHARMA UNIVERSITY USING C4.5 DECISION TREE ALGORITHM," *MathVision: Jurnal Matematika.*, vol. 08, iss. 01, pp. 17-27, 2026.

KONTAK:

Penulis Korespondensi (Primary Contact),  a.s.nugraha@usd.ac.id

 Universitas Sanata Dharma

 Artikelnya dapat diakses di sini. <https://doi.org/10.55719/mv.v8i1.2204>

Copyright © 2026 Author(s)

Journal homepage: <https://journal.unirow.ac.id/index.php/mv/>

Journal e-mail: jmathvision@gmail.com

Research Article · Open Access

1. PENDAHULUAN

Pemilihan karier merupakan salah satu keputusan penting yang harus dibuat mahasiswa dalam mempersiapkan masa depan. Berdasarkan data Badan Pusat Statistik (BPS) tahun 2025, tingkat pengangguran terbuka lulusan universitas terus mengalami peningkatan, yaitu sebesar 4,80% pada 2022, 5,18% pada 2023, dan 5,25% pada tahun 2024. Hal ini menunjukkan bahwa menjadi lulusan perguruan tinggi tidak menjamin seseorang untuk mencapai kesuksesan karier, sehingga diperlukan adanya perencanaan karier yang lebih matang dan terarah. Dengan mengetahui karier yang diinginkan, mahasiswa dapat merencanakan langkah-langkah yang harus diambil untuk mengembangkan keahlian dan keterampilan yang dapat membantu mereka mencapai kesuksesan di dunia kerja [1]. Pemilihan karier menjadi bagian keputusan strategis pada fase transisi mahasiswa menuju dunia kerja karena berpengaruh pada arah pengembangan kompetensi, pilihan pengalaman belajar, hingga kesiapan memasuki pasar kerja. Tantangan pada fase ini semakin kompleks dikarenakan kebutuhan di lapangan yang berubah dengan cepat dan membutuhkan kesiapan serasi kompetensi individu, pengalaman, serta keterampilan yang relevan. Dalam pendidikan tinggi, konteks *graduate employability* menjadi perhatian penting dikarenakan keterlambatan dan ketidaktepatan keputusan karier dapat berdampak pada rendahnya kesiapan kerja dan resiko pengangguran lulusan [2], [3].

Pemilihan karier seseorang biasanya dipengaruhi oleh beberapa faktor. Faktor-faktor tersebut dapat dianalisis, diolah, dan dimanfaatkan untuk memetakan pilihan karier mahasiswa dengan bantuan data mining. Data mining merupakan kegiatan pengumpulan dan pemakaian data historis untuk menemukan pola atau hubungan dalam dataset yang besar [4]. Data mining dapat digunakan dalam berbagai bidang kehidupan, termasuk pendidikan. Sejalan dengan hal tersebut, pendidikan tinggi semakin di dorong untuk memanfaatkan pendekatan berbasis data untuk dapat memahami pola dan faktor yang memengaruhi hasil akademik maupun kesiapan kerja mahasiswa. Di konteks Indonesia, analisis berbasis data juga telah digunakan untuk memetakan profil kelompok numerasi yang dikaji dari data pendidikan berskala besar [5]. Lebih lanjut, kajian dan praktik *Educational Data Mining (EDM)* menunjukkan bahwa data pendidikan (akademik, demografis, aktivitas, dan rekam jejak pembelajarannya) juga dapat didalami dan digali untuk menemukan pola yang mendukung pengambilan keputusan yang lebih tepat oleh mahasiswa ataupun institusi [6], [7].

Algoritma data mining yang digunakan dalam penelitian ini adalah algoritma C4.5. Algoritma C4.5 menjadi salah satu algoritma yang populer karena kemampuannya yang dapat menangani data numerik dan menghasilkan model yang mudah diinterpretasikan [8]. Penelitian yang memanfaatkan algoritma C4.5 untuk klasifikasi pilihan karier mahasiswa juga sudah pernah dilakukan sebelumnya, seperti penelitian yang dilakukan oleh Asroni [9]. Algoritma C4.5 merupakan salah satu pendekatan *tree-based* yang relevan untuk klasifikasi. Algoritma ini menghasilkan struktur pohon keputusan berbasis pemilihan atribut dan aturan klasifikasi. Dalam berbagai penelitian pendidikan, teknik klasifikasi diterapkan untuk memetakan kategori hasil (seperti performa akademik, status *employability*, ataupun arah pilihan karir) berdasarkan kombinasi atribut mahasiswa [3], [4]. Lebih lanjut lagi, penelitian berbasis *campus big data* juga menunjukkan bahwa data mahasiswa dapat digunakan untuk memprediksi kecenderungan pilihan karier, sekaligus menafsirkan kontribusi faktor-faktor penentu melalui pendekatan interpretatif [10].

Meskipun penelitian terkait EDM dan prediksi karier berkembang, namun masih diperlukan kajian yang lebih beragam pada konteks program studi, terutama pada mahasiswa kependidikan. Hal ini dikarenakan karakter kompetensi, pola pengalaman belajar, orientasi karier dapat berbeda dibandingkan rumpun non-kependidikan. Berdasarkan uraian di atas, penelitian ini bertujuan untuk membangun model klasifikasi pilihan karier mahasiswa Pendidikan Matematika Universitas Sanata Dharma menggunakan algoritma C4.5 dan mengevaluasi akurasi model yang dihasilkan. Sejalan dengan tujuan tersebut, pertanyaan penelitian ini meliputi: (1) bagaimana model klasifikasi pilihan karier mahasiswa dibangun menggunakan algoritma C4.5, (2) bagaimana performa model pada data awal, dan (3) bagaimana perubahan performa model setelah dilakukan oversampling. Hasil penelitian ini diharapkan dapat menjadi dasar awal dalam pemetaan pilihan karier mahasiswa secara lebih sistematis dan berbasis data.

2. METODE

Penelitian ini merupakan penelitian terapan dengan memanfaatkan algoritma C4.5 untuk mengklasifikasi pilihan karier mahasiswa Pendidikan Matematika Universitas Sanata Dharma. Algoritma C4.5 merupakan salah satu algoritma *data mining* yang digunakan untuk menemukan hubungan antara variabel *input* dengan variabel target dalam bentuk pohon keputusan [11]. Algoritma C4.5 menggunakan

parameter *Gain Ratio* untuk menentukan atribut terbaik. Perhitungan *Gain Ratio* dilakukan untuk setiap atribut, dan atribut dengan nilai *Gain Ratio* tertinggi akan dipilih menjadi atribut terbaik untuk dijadikan node. Nilai *Gain Ratio* dapat dihitung menggunakan persamaan berikut [12].

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \quad (1)$$

Gain Ratio merupakan modifikasi dari *Gain* atau *Information Gain* yang bertujuan untuk mengurangi bias terhadap suatu atribut, dengan memperhitungkan jumlah cabang yang akan dihasilkan sebelum melakukan split. *Information Gain* sendiri digunakan untuk mengukur seberapa baik suatu atribut dalam memisahkan data sesuai dengan klasifikasi targetnya. Namun, *Information Gain* cenderung lebih mengutamakan pengujian yang menghasilkan banyak keluaran/cabang sehingga atribut yang dipilih menjadi atribut *splitting* adalah atribut memiliki banyak nilai [13]. Perhitungan nilai *Information Gain* dapat dilakukan menggunakan persamaan berikut [12].

$$Gain(A) = Entropy(p) - \left(\sum_{i=1}^k \frac{n_i}{n} Entropy(i) \right) \quad (2)$$

Keterangan:

p = *parent node* yang akan dilakukan percabangan

n_i = banyaknya data pada nilai ke- i dalam atribut A

n = banyaknya data pada atribut A

Sebelum menghitung *Information Gain* suatu atribut, perlu diketahui nilai *Entropy* dari *parent node* dan setiap kemungkinan cabang yang dapat terbentuk dari *node* tersebut. *Entropy* merupakan ukuran heterogenitas atau ketidakpastian (*impurity*) dari suatu kumpulan data. Perhitungan *Entropy* pada *node p* dilakukan dengan persamaan berikut [12].

$$Entropy(p) = - \sum_{i=1}^k p(i|p) \log_2 p(i|p) \quad (3)$$

Keterangan:

p = *parent node* yang akan dilakukan percabangan

$p(j | t)$ = proporsi kelas ke- i pada *node t*

Nilai *Information Gain* yang diperoleh harus dibagi dengan *Split Info* untuk mendapatkan nilai *Gain Ratio*. *Split Info* merupakan teknik pengelompokkan data yang dilakukan dengan membagi data menjadi dua atau lebih bagian berdasarkan nilai dari suatu atribut. *Split Info* dapat dihitung menggunakan persamaan berikut [12].

$$SplitInfo(A) = - \sum_{i=1}^k \frac{n_i}{n} \log_2 \frac{n_i}{n} \quad (4)$$

Keterangan:

n_i = banyaknya data pada nilai ke- i dalam atribut A

n = banyaknya data pada atribut A

Pengambilan data dilakukan melalui Biro Administrasi Perencanaan dan Sistem Informasi (BAPSI) Universitas Sanata Dharma. Subjek dalam penelitian ini adalah mahasiswa Pendidikan Matematika Universitas Sanata Dharma angkatan 2011 sampai 2020. Variabel penelitian yang digunakan pada penelitian ini adalah variabel bebas dan variabel terikat. Variabel bebas dari penelitian ini yaitu data mahasiswa Pendidikan Matematika Universitas Sanata Dharma yang meliputi data IPK, total poin, jenis kelamin, nilai

rata-rata mata kuliah kependidikan, dan nilai rata-rata mata kuliah matematika. Variabel terikat dari penelitian ini yaitu pilihan karier mahasiswa Pendidikan Matematika Universitas Sanata Dharma. Pemilihan atribut IPK, total poin, jenis kelamin, nilai rata-rata mata kuliah kependidikan, dan nilai rata-rata mata kuliah matematika didasarkan pada pertimbangan bahwa pilihan karier mahasiswa dipengaruhi oleh faktor akademik, demografis, dan pengalaman pengembangan diri. IPK merepresentasikan performa akademik umum, nilai mata kuliah matematika dan kependidikan merepresentasikan kompetensi inti mahasiswa Pendidikan Matematika, jenis kelamin digunakan sebagai atribut demografis, sedangkan total poin mencerminkan partisipasi mahasiswa dalam kegiatan kemahasiswaan dan pengembangan soft skills sesuai dengan perhitungan poin yang ada di universitas. Oleh karena itu, kelima atribut tersebut dipandang relevan untuk memetakan pilihan karier mahasiswa

Penelitian ini dilakukan dengan mengikuti tahapan-tahapan *data mining CRISP-DM (Cross-Industry Standard Process for Data Mining)* yang meliputi enam tahap, yaitu *business/research understanding, data understanding, data preparation, modeling, evaluation, dan deployment* [14], [15]. Model ini berfokus pada pemahaman proses bisnis dan karakteristik data sehingga dapat menghasilkan interpretasi output yang lebih spesifik serta merekomendasikan solusi yang relevan [16].

3. HASIL DAN PEMBAHASAN

Klasifikasi pilihan karier mahasiswa Pendidikan Matematika Universitas Sanata Dharma dilakukan menggunakan algoritma C4.5 dengan mengikuti tahapan data mining CRISP-DM meliputi *business/research understanding, data understanding, data preparation, modeling, evaluation, dan deployment*.

3.1. Business/Research Understanding

Tahap ini bertujuan untuk memahami tujuan penelitian secara menyeluruh dan merumuskan masalah yang akan diselesaikan. Penelitian ini bertujuan untuk membangun model klasifikasi pilihan karier mahasiswa berdasarkan data historis mahasiswa Pendidikan Matematika Universitas Sanata Dharma menggunakan algoritma C4.5, serta mengetahui tingkat akurasi model klasifikasi yang dibangun dalam memetakan pilihan karier mahasiswa Pendidikan Matematika Universitas Sanata Dharma.

3.2. Data Understanding

Data yang digunakan merupakan data sekunder yang diperoleh dari Biro Administrasi Perencanaan dan Sistem Informasi (BAPSI) Universitas Sanata Dharma. Data yang diberikan berupa data informasi alumni mahasiswa Pendidikan Matematika Universitas Sanata Dharma angkatan 2011 hingga 2020 sebanyak 901 mahasiswa yang diberikan dalam lima file excel, meliputi file data IPK, file data IPS, file data pilihan karier, file data daftar poin, dan file data detail nilai.

Hasil eksplorasi awal menunjukkan bahwa kualitas data masih rendah karena banyaknya data yang mengandung nilai kosong, terutama pada file data pilihan karier yang merupakan data inti dalam penelitian ini. Selain itu, ditemukan juga adanya entri data yang tidak valid dan data dengan variasi kategori yang cukup tinggi, sehingga diperlukan penanganan lebih lanjut pada tahap selanjutnya

3.3. Data Preparation

Sebelum digunakan pada tahap modeling, data harus terlebih dahulu dipersiapkan menjadi data bersih. Tahapan ini dilakukan menggunakan Microsoft Excel dengan langkah-langkah sebagai berikut.

1. Transformasi dan pembersihan data

Pada tahap ini dilakukan pengkategorian pilihan karier mahasiswa yang terdiri dari enam kelas, yaitu guru SD, guru SMP, guru SMA/K, pendidik/pengajar (mencakup pekerjaan-pekerjaan di bidang pendidikan atau yang bekerja di sekolah, tetapi tidak ada keterangan jenjang SD/SMP/SMA), lainnya (mencakup pekerjaan-pekerjaan di luar bidang pendidikan, seperti wiraswasta, perbankan, atau pegawai swasta selain di bidang pendidikan), dan melanjutkan pendidikan. Setelah transformasi data, dilakukan pembersihan data untuk baris-baris yang tidak dapat disimpulkan pilihannya, seperti baris yang memiliki nilai kosong pada kedua kolom status pekerjaan dan nama tempat kerja. Proses transformasi data juga dilakukan untuk menghitung total poin dan nilai rata-rata mata kuliah kependidikan serta matematika untuk setiap mahasiswa.

2. Seleksi dan integrasi data

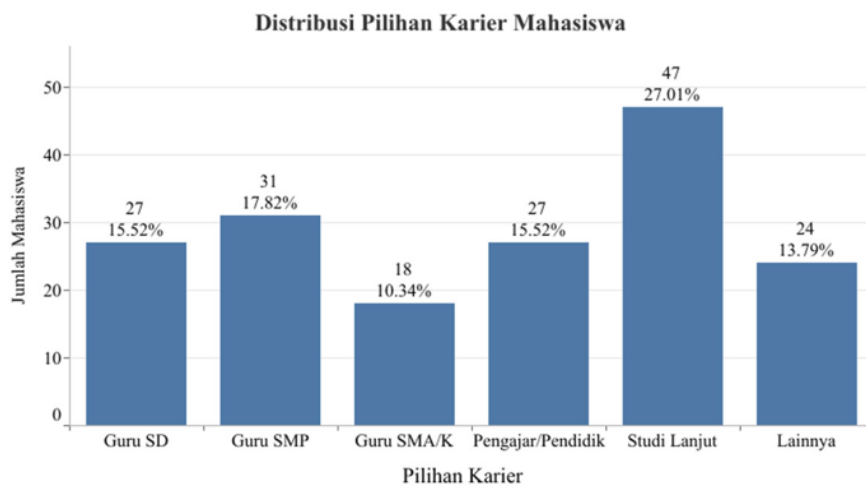
Pada tahap ini, dilakukan pemilihan atribut yang akan digunakan untuk proses modeling. Terdapat enam atribut yang digunakan pada penelitian ini, yaitu IPK, total poin, jenis kelamin, nilai rata-rata mata kuliah kependidikan, nilai rata-rata mata kuliah matematika, dan pilihan karier mahasiswa (atribut target). Keenam atribut tersebut kemudian diintegrasikan menjadi satu file excel. Penggabungan data dilakukan menggunakan kolom kunci NIM, yang berfungsi sebagai penghubung antar data dari file yang berbeda.

Nilai rata-rata mata kuliah kependidikan dan matematika yang semula berbentuk skor numerik dikonversi ke huruf mutu berdasarkan Peraturan Akademik Program Sarjana Universitas Sanata Dharma. Adapun pedoman konversi yang digunakan adalah: $85 \leq x \leq 100 = A$, $80 \leq x < 85 = A-$, $75 \leq x < 80 = B+$, $70 \leq x < 75 = B$, $65 \leq x < 70 = B-$, $60 \leq x < 65 = C+$, $56 \leq x < 60 = C$, $40 \leq x < 56 = D$, dan $0 \leq x < 40 = E$. Konversi ini dilakukan agar data konsisten dengan sistem penilaian institusi dan memudahkan interpretasi hasil klasifikasi. Hasil pengolahan data pada tahap data preparation menghasilkan data bersih yang terdiri dari enam atribut, dengan rincian pada Tabel 1.

Dari 901 data awal, hanya tersisa 174 data bersih atau sekitar 19,31% dari total keseluruhan data. Jumlah tersebut disebabkan karena banyaknya data kosong, khususnya pada atribut pilihan karier, serta adanya data yang tidak sesuai dengan aturan akademik, seperti pada atribut total poin. Untuk menjaga kualitas model yang dibangun, hanya data dengan informasi lengkap yang akan digunakan, sehingga data yang memiliki nilai kosong di suatu atribut tertentu akan dihapus. Data bersih ini kemudian dieksplorasi untuk melihat distribusi dari pilihan karier mahasiswa.

Tabel 1. Rincian Atribut Data Bersih

	Nama Atribut	Jenis Data	Nilai
Input	IPK	Data Numerik	$2,56 \leq \text{IPK} \leq 3,94$
	Jenis Kelamin	Data Kategorikal	"L" dan "P"
	Total Poin	Data Numerik	$10 \leq \text{Total Poin} \leq 47$
	Nilai Rata-rata Mata kuliah Kependidikan	Data Kategorikal	"A", "A-", "B+", dan "B"
Output (Target)	Nilai Rata-rata Mata Kuliah Matematika	Data Kategorikal	"A", "A-", "B+", "B", "B-", "C+", dan "C"
	Pilihan karier	Data Kategorikal	"Guru SD", "Guru SMP", "Guru SMA/K", "Pengajar/Pendidik", "Lainnya", dan "Melanjutkan Pendidikan"



Gambar 1. Distribusi Pilihan Karier Mahasiswa

Berdasarkan grafik distribusi pilihan karier pada Gambar 1, terlihat bahwa mayoritas mahasiswa memilih untuk melanjutkan pendidikan dengan persentase sebesar 27,01%. Namun, jika dikategorikan lebih lanjut ke dalam tiga kategori besar, yaitu pendidik (mencakup guru SD, SMP, SMA/K, dan pengajar/pendidik), melanjutkan pendidikan, dan lainnya, maka mayoritas mahasiswa melanjutkan karier sebagai pendidik dengan total persentase sebesar 59,20%. Hal ini menunjukkan kecenderungan mahasiswa melanjutkan karier tetap dalam ranah pendidikan. Sementara itu, di antara kelompok pendidik yang memiliki

keterangan jenjang pendidikan, jumlah terbanyak berada pada jenjang SMP, SD, dan yang paling sedikit adalah jenjang SMA/K.

3.4. Modeling

Data bersih yang diperoleh dari tahap data *preparation* kemudian diolah pada tahap modeling untuk membangun model klasifikasi yang mampu memetakan pilihan karier mahasiswa berdasarkan atribut-atribut yang tersedia. Pada penelitian ini, pemodelan dilakukan menggunakan perangkat lunak WEKA versi 3.8.6 melalui antarmuka Explorer. Algoritma klasifikasi yang digunakan adalah C4.5 melalui implementasi classifier J48 pada WEKA, karena mampu menangani atribut numerik dan kategorikal serta menghasilkan model berbentuk pohon keputusan yang mudah diinterpretasikan. Proses pemodelan menggunakan pengaturan bawaan (*default setting*) J48, yaitu menggunakan *pruning* (*post-pruning*) pada pohon keputusan (*pruned tree*), dengan *confidence factor* (*CF*) = 0,25 dan *minimum number of instances per leaf* (*minNumObj*) = 2. Evaluasi model dilakukan menggunakan k-fold cross validation dengan nilai k = 10, yaitu data dibagi menjadi 10 bagian. Pada setiap iterasi, 9 bagian digunakan sebagai data latih dan 1 bagian sebagai data uji, lalu proses diulang hingga seluruh bagian pernah menjadi data uji. Skema ini digunakan agar estimasi performa model lebih stabil pada data yang tersedia. Hasil pohon keputusan dapat dilihat seperti pada Gambar 2.

Model pohon keputusan yang dihasilkan memiliki 102 node, terdiri atas 1 root node, 41 internal node, dan 60 leaf node. Atribut yang menjadi root node adalah rata-rata nilai mata kuliah kependidikan, yang menunjukkan bahwa atribut tersebut memiliki daya pisah paling besar terhadap kelas target pada data awal. Secara substantif, hasil ini relevan dengan konteks mahasiswa Pendidikan Matematika, karena capaian pada mata kuliah kependidikan mencerminkan kesiapan pedagogik yang berkaitan dengan orientasi profesi di bidang pendidikan. Temuan ini berimplikasi bahwa data capaian kependidikan dapat dimanfaatkan sebagai pemetaan awal kecenderungan karier mahasiswa, meskipun tidak dapat dimaknai sebagai hubungan sebab-akibat. Hasil ini sejalan dengan penelitian Asroni (2018) dan Wang dkk. (2022) yang menunjukkan bahwa atribut akademik dapat digunakan untuk memetakan pilihan karier mahasiswa melalui model yang interpretatif. Selain itu, pohon keputusan juga menghasilkan aturan klasifikasi yang transparan, misalnya mahasiswa dengan Nilai Kependidikan = B dan Poin > 14 cenderung diklasifikasikan sebagai Guru SD, sedangkan mahasiswa dengan Nilai Kependidikan = A dan Nilai Matematika = C cenderung diklasifikasikan sebagai Melanjutkan Pendidikan. Hal ini menunjukkan bahwa model membentuk keputusan berdasarkan kombinasi beberapa atribut, bukan satu faktor tunggal.

(ROS). Teknik ini menjadi pilihan yang populer dan efektif karena mudah digunakan dan mampu mempertahankan informasi asli pada data [18]. Teknik ROS bekerja dengan cara menduplikasi sampel data dari kelas minoritas secara acak dan menambahkannya ke dalam dataset hingga distribusi kelas seimbang [19].

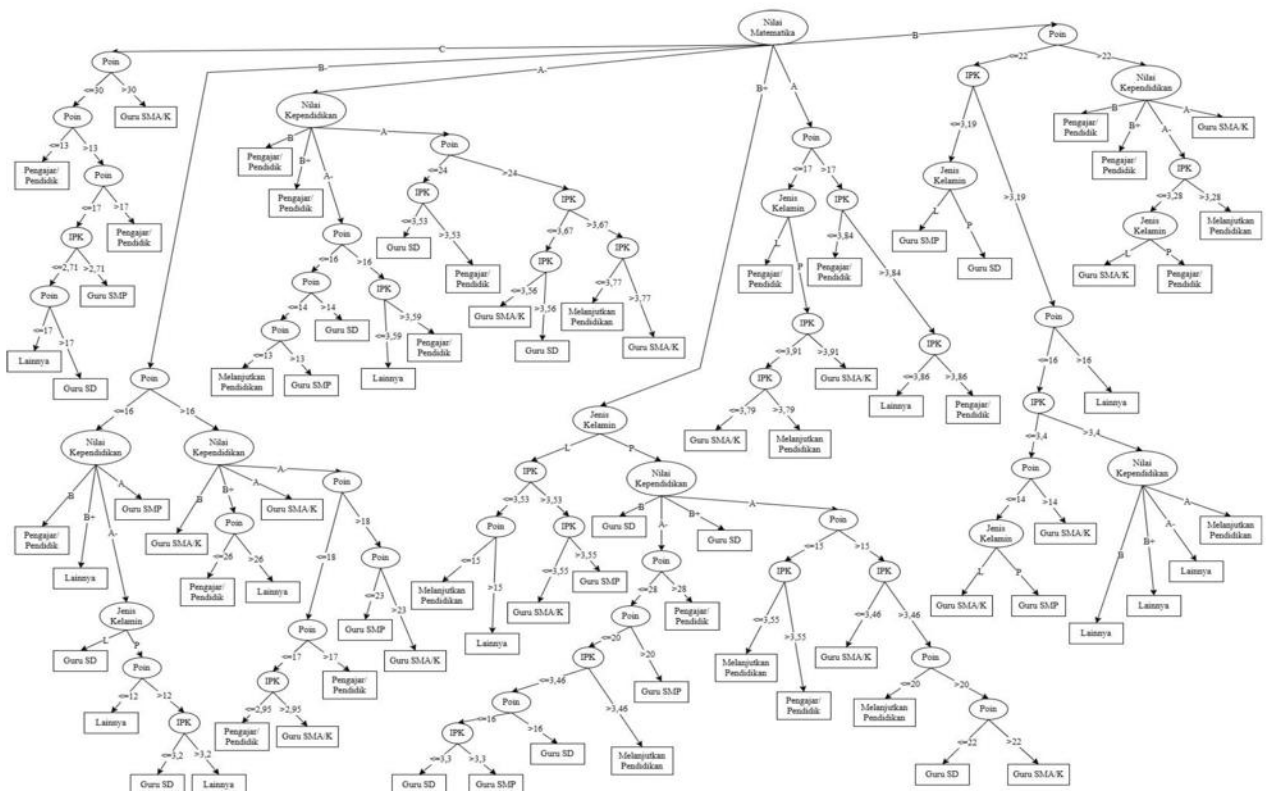
Proses ROS dilakukan dengan bantuan google colab. Setelah dilakukan ROS, distribusi kelas target menjadi seimbang dengan seluruh kelas memiliki 47 baris data. Dataset hasil ini kemudian dilakukan proses modeling menggunakan algoritma C4.5 dengan bantuan perangkat lunak WEKA, dan dievaluasi dengan proses yang sama seperti data sebelumnya. Hasil evaluasi model dapat dilihat dari *confusion matrix* seperti pada Gambar 4

```

=== Confusion Matrix ===
      a  b  c  d  e  f  <-- classified as
 25  6  5  5  4  2 | a = Pengajar/Pendidik
  5 34  0  1  4  3 | b = Lainnya
  3  1 29  4  5  5 | c = Guru SMA/K
  7  6  4 20  6  4 | d = Guru SMP
 11  6 12  6  6  6 | e = Melanjutkan Pendidikan
  2  1  2  4  5 33 | f = Guru SD
    
```

Gambar 4. Confusion Matrix Model Klasifikasi Data Oversampling

Dari hasil *confusion matrix* tersebut terlihat terjadinya peningkatan hasil evaluasi model dari data hasil *oversampling* yang menunjukkan tingkat akurasi model sebesar 52,13%. Setelah diterapkan teknik *oversampling*, performa model meningkat cukup signifikan. Nilai *precision* per kelas meningkat menjadi 20,00%–62,96%, *recall* per kelas menjadi 12,77%–72,34%, dan F1-score per kelas menjadi 15,58%–67,33%. Secara keseluruhan, *macro precision* naik menjadi 49,69%, *macro recall* menjadi 52,13%, dan *macro F1-score* menjadi 50,58%. Temuan ini menunjukkan bahwa *oversampling* mampu memperbaiki kemampuan model dalam mengenali kelas-kelas minoritas dan meningkatkan keseimbangan performa klasifikasi antarkelas. Namun demikian, kelas Melanjutkan Pendidikan masih menunjukkan nilai *precision*, *recall*, dan *F1-score* yang relatif rendah, sehingga kelas ini masih menjadi kategori yang paling sulit diprediksi secara konsisten oleh model. Proses modeling ini menghasilkan pohon keputusan seperti pada Gambar 5 berikut.



Gambar 5. Pohon Keputusan Hasil Klasifikasi Data *Oversampling*

Evaluasi menggunakan *precision*, *recall*, dan *F1-score* menunjukkan bahwa *oversampling* secara umum meningkatkan performa model secara substansial. Hal ini terlihat dari kenaikan rata-rata *macro precision* dari 12,97% menjadi 49,69%, *macro recall* dari 13,26% menjadi 52,13%, dan *macro F1-score* dari 13,04% menjadi 50,58%. Peningkatan paling besar terjadi pada kelas Guru SD, Lainnya, Pengajar/Pendidik, dan Guru SMA/K. Sebagai contoh, *F1-score* kelas Guru SD meningkat dari 4,00% menjadi 66,00%, sedangkan kelas Lainnya meningkat dari 11,76% menjadi 67,33%. Kelas Guru SMP juga mengalami peningkatan, meskipun *recall*-nya masih relatif lebih rendah dibanding beberapa kelas lain. Sebaliknya, kelas Melanjutkan Pendidikan justru mengalami penurunan performa, dengan *F1-score* turun dari 29,70% menjadi 15,58%, sehingga tetap menjadi kelas yang paling sulit diprediksi. Hasil ini menunjukkan bahwa *oversampling* efektif memperbaiki performa model pada sebagian besar kelas, tetapi belum mampu mengatasi sepenuhnya kesulitan pemisahan pada kelas tertentu yang kemungkinan memiliki karakteristik atribut yang beririsan dengan kelas lain.

3.6. Deployment

Pada tahap ini, dilakukan pengambilan keputusan sebagai tindak lanjut dari hasil evaluasi model. Hasil klasifikasi menunjukkan adanya peningkatan akurasi model menjadi 52,13% setelah dilakukan proses *oversampling*. Meskipun terjadi peningkatan, nilai akurasi ini masih tergolong rendah untuk digunakan secara nyata dalam klasifikasi pilihan karier mahasiswa. Namun, model klasifikasi ini tetap memiliki potensi untuk dimanfaatkan oleh pihak program studi sebagai gambaran awal kecenderungan pilihan karier mahasiswa dan juga sebagai bahan pertimbangan dalam memberikan arahan atau rekomendasi awal pilihan karier yang dapat didiskusikan lebih lanjut dalam proses pembinaan karier mahasiswa.

Berdasarkan hasil penelitian, rekomendasi untuk penelitian selanjutnya sebagai berikut. Peningkatan jumlah data yang digunakan dalam proses modeling, baik itu dengan meningkatkan kualitas database universitas terkait data karier mahasiswa maupun dengan melakukan survei tambahan untuk melengkapi kekosongan data. Ketersediaan data yang lebih lengkap diharapkan dapat mendukung pembangunan model klasifikasi yang lebih baik dan akurat. Penggunaan atribut lain yang berpengaruh terhadap pilihan karier mahasiswa, khususnya atribut-atribut yang termasuk dalam faktor eksternal pemilihan karier, seperti pengaruh keluarga dan lingkungan. Eksplorasi terhadap atribut lainnya diharapkan dapat menghasilkan model klasifikasi dengan tingkat akurasi yang lebih tinggi. Penerapan algoritma data mining lain untuk mengklasifikasikan pilihan karier mahasiswa. Salah satu algoritma yang dapat digunakan yaitu *Random Forest* yang merupakan turunan dari algoritma *decision tree* dan dapat mengatasi masalah kemungkinan terjadinya *overfitting* pada *decision tree* [20]. Pengembangan model klasifikasi berbasis antarmuka grafis (GUI) agar model klasifikasi dapat digunakan secara lebih mudah dan praktis oleh pengguna.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma C4.5 dapat digunakan untuk membangun model klasifikasi pilihan karier mahasiswa Pendidikan Matematika Universitas Sanata Dharma secara interpretatif. Temuan utama penelitian ini adalah bahwa atribut yang paling dominan dalam model berubah dari nilai mata kuliah kependidikan pada data awal menjadi nilai mata kuliah matematika setelah *oversampling*, yang menunjukkan bahwa distribusi data sangat memengaruhi struktur pohon keputusan. *Oversampling* terbukti meningkatkan performa model secara substansial pada sebagian besar kelas, tetapi belum mampu memperbaiki semua kelas secara merata, terutama pada kategori Melanjutkan Pendidikan yang tetap menjadi kelas paling sulit diprediksi. Oleh karena itu, model ini lebih tepat diposisikan sebagai alat pemetaan awal berbasis data untuk mendukung layanan bimbingan karier, bukan sebagai alat prediksi final. Penelitian lanjutan perlu menambah atribut yang lebih relevan dengan orientasi karier dan mengeksplorasi algoritma lain agar diperoleh model yang lebih kuat.

Kontribusi Penulis

Penulis Pertama : Konseptualisasi, metodologi, penulisan-draf awal, perangkat lunak, validasi.

Penulis Kedua : Konseptualisasi, Kurasi data, Analisis formal, Perolehan dana, Investigasi,

Metodologi, Sumber daya, Pengawasan, Penulisan - draf asli, Penulisan - tinjauan & penyuntingan.

Funding Statement

"Penelitian ini tidak menerima hibah khusus dari badan pendanaan mana pun di sektor publik, komersial, atau nirlaba."

Deklarasi

Para penulis menyatakan bahwa mereka tidak memiliki konflik kepentingan terkait studi yang dilaporkan..

Deklarasi AI Generatif dan Teknologi yang Dibantu AI

Penulis menyatakan bahwa tidak ada AI generatif atau teknologi berbantuan AI yang digunakan dalam persiapan naskah ini, termasuk untuk penulisan, penyuntingan, analisis data, atau pembuatan tabel dan gambar.

REFERENSI

- [1] D. Ozora, L. Suharti, and H. Sirine, "Potret Perencanaan Karir pada Mahasiswa (Studi terhadap Mahasiswa di Sebuah Perguruan Tinggi di Jawa Tengah)," in *Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers UNISBANK KE-2 Tahun 2016*, Semarang, 2016, pp. 623–632.
- [2] D. K. Dake, "Information technology graduates employability prediction model in a low-income country using tree-based machine learning classifiers," *Discover Global Society*, vol. 3, no. 1, Dec. 2025, doi: 10.1007/s44282-025-00304-3.
- [3] L. S. Chen, T. T. Huynh-Cam, V. C. Nguyen, T. C. Lu, and D. K. Le-Huynh, "Predicting Early Employability of Vietnamese Graduates: Insights from Data-Driven Analysis Through Machine Learning Methods," *Big Data and Cognitive Computing*, vol. 9, no. 5, May 2025, doi: 10.3390/bdcc9050134.
- [4] I. H. Sarker, "Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective," *SN Comput. Sci.*, vol. 2, no. 5, p. 377, Sep. 2021, doi: 10.1007/s42979-021-00765-8.
- [5] A. S. Nugraha and M. A. Rudhito, "Unveiling Numeracy Competency Domains of High School Students in Indonesia: A Clustering Analysis Approach," *JMPM: Jurnal Matematika dan Pendidikan Matematika*, vol. 10, no. 2, pp. 161–179, Oct. 2025, doi: 10.26594/jmpm.v10i2.5972.
- [6] E. Kalita *et al.*, "Educational data mining: a 10-year review," Dec. 01, 2025, *Springer Science and Business Media B.V.* doi: 10.1007/s10791-025-09589-z.
- [7] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40561-022-00192-z.
- [8] F. Rezkika, B. N. Sari, and A. S. Y. Irawan, "Klasifikasi Masa Tunggu Alumni untuk Mendapatkan Pekerjaan Menggunakan Algoritma C4.5 (Studi Kasus: Fasilkom Unsika)," *Progresif: Jurnal Ilmiah Komputer*, vol. 17, no. 2, pp. 95–106, 2021.
- [9] A. Asroni, B. Masajeng Respati, and S. Riyadi, "Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta," *Semesta Teknika*, vol. 21, no. 2, 2018, doi: 10.18196/st.212222.
- [10] Y. Wang, L. Yang, J. Wu, Z. Song, and L. Shi, "Mining Campus Big Data: Prediction of Career Choice Using Interpretable Machine Learning Method," *Mathematics*, vol. 10, no. 8, Apr. 2022, doi: 10.3390/math10081289.
- [11] Amrul Hadiyanoor, Siti Cholifah, Husnul Ma'ad Junaidi, and Irfan Febrian, "Using C4.5 Decision Tree to Determine the Majors of Students in SMAN 4 Banjarmasin to Reduce the Cause of DropOut from School," *IIAI Letters on Informatics and Interdisciplinary Research*, vol. 5, p. 1, 2024, doi: 10.52731/liir.v005.209.
- [12] S. Diwandari, E. I. Sela, B. E. Syahputra, N. A. Parama, and A. Septiarini, "The Utility of Decision Tree and Analytics Hierarchy Process in Prioritizing of Social Aid Distribution due to Covid-19 Pandemic in Indonesia," *Journal of ICT Research and Applications*, vol. 17, no. 1, pp. 82–98, Apr. 2023, doi: 10.5614/itbj.ict.res.appl.2023.17.1.6.
- [13] P. Rim and E. Liu, "Optimizing the C4.5 Decision Tree Algorithm using MSD-Splitting," *IJACSA International Journal of Advanced Computer Science and Applications*, vol. 11, no. 10, pp. 41–47, 2020.

- [14] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. Van Erven, “Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil,” *J. Bus. Res.*, vol. 94, pp. 335–343, Jan. 2019, doi: 10.1016/j.jbusres.2018.02.012.
- [15] C. Schröer, F. Kruse, and J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534.
- [16] Y. A. Singgalen, “Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT),” *Jurnal Media Informatika Budidharma*, vol. 7, no. 3, p. 1551, Jul. 2023, doi: 10.30865/mib.v7i3.6461.
- [17] R. A. Akmal and A. Kurniasih, “Penerapan Algoritma Klasifikasi untuk Menangani Data Tidak Seimbang pada Peningkatan Kualitas Siswa,” in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, Jakarta, 2023, pp. 560–566.
- [18] I. K. Ananda, A. Z. Fanani, D. Setiawan, and D. F. Wicaksono, “Penerapan Random Oversampling dan Algoritma Boosting untuk Memprediksi Kualitas Buah Jeruk,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 282–289, Jun. 2024, doi: 10.29408/edumatic.v8i1.25836.
- [19] U. Hasanah, A. M. Soleh, and K. Sadik, “Effect of Random Under sampling, Oversampling, and SMOTE on the Performance of Cardiovascular Disease Prediction Models,” *Jurnal Matematika, Statistika dan Komputasi*, vol. 21, no. 1, pp. 88–102, Sep. 2024.
- [20] A. Ferdita Nugraha, R. Faticha, A. Aziza, and Y. Pristyanto, “Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing,” *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan*, vol. 7, no. 1, pp. 39–44, 2022.