

# Klasifikasi Tingkat Kesiapan Kerja Mahasiswa Universitas Negeri Surabaya Menggunakan Algoritma C5.0

Grace Cynthia<sup>1</sup>, Anita Qoiriah<sup>2</sup>

<sup>1,2</sup> Jurusan Teknik Informatika, Universitas Negeri Surabaya

[grace.20108@mhs.unesa.ac.id](mailto:grace.20108@mhs.unesa.ac.id)

[anitaqoiriah@unesa.ac.id](mailto:anitaqoiriah@unesa.ac.id)

**Abstrak**— Era globalisasi dan kompleksitas dunia kerja menuntut kesiapan kerja bagi mahasiswa demi menjamin keberhasilan mereka untuk menghindari terjadinya pengangguran sarjana. Pada periode wisuda 109, Universitas Negeri Surabaya meluluskan 1.380 mahasiswa. Jumlah lulusan yang banyak tersebut penting untuk diimbangi dengan tingkat kesiapan kerja yang mumpuni. Maka, analisis terhadap tingkat kesiapan kerja bagi mahasiswa diperlukan untuk mengungkap hal-hal yang mempengaruhi kesiapan kerja mahasiswa. Analisis dilakukan dengan melakukan proses klasifikasi. Penelitian ini menggunakan data dari 114 responden yang merupakan mahasiswa tingkat akhir di Universitas Negeri Surabaya yang tengah berada pada semester 6 dan 8. Responden mengisi survei yang terdiri atas 16 atribut, di antaranya fleksibel, perencanaan, prioritas, komorganisir, tangkas, openmind, adaptasitek, manfaattek, pelatihan, multitask, demonstrasi, kebersiapan, inovatif, dan analisis, komefektif, dan studi. Data yang diperoleh melalui survei kemudian melalui *preprocessing* menggunakan metode SMOTE (*Synthetic Minor-Oversampling Technique*) yang dilanjutkan dengan seleksi fitur menggunakan teknik *Information Gain* untuk mengurangi jumlah atribut pada dataset. Dataset yang dihasilkan kemudian diklasifikasikan menggunakan algoritma C5.0 *decision tree* untuk membagi dataset ke dalam kategori siap kerja dan tidak siap kerja. Proses klasifikasi dilakukan dalam tiga skema pembagian data latih dan data uji, yaitu 70:30, 80:20, dan 90:10. Evaluasi dilakukan menggunakan *confusion matrix* dengan menggunakan beberapa metrik evaluasi seperti akurasi, presisi, dan *recall*. Hasil evaluasi menampilkan bahwa algoritma C5.0 mampu mengklasifikasikan kesiapan kerja mahasiswa tingkat akhir dengan tingkat akurasi tinggi 93% (skema 80:20), presisi 100% (skema 80:20), dan *recall* 92% (skema 70:30). Model klasifikasi ini membantu mahasiswa mengidentifikasi aspek-aspek yang mempengaruhi kesiapan kerja dan mempersiapkan diri untuk memasuki dunia kerja.

**Kata Kunci**— *Klasifikasi, Kesiapan Kerja, Mahasiswa, Universitas Negeri Surabaya, Decision Tree, Algoritma C5.0, Seleksi Fitur, SMOTE*

## I. PENDAHULUAN

Dalam era perkembangan global yang semakin pesat dan di tengah kompleksitas dunia kerja yang terus berkembang, kesiapan kerja bagi mahasiswa menjadi suatu keharusan yang penting untuk dilakukan demi menjamin keberhasilan mereka menghadapi tantangan masa depan dan menghindari terjadinya pengangguran sarjana. Pengangguran sarjana disebabkan karena banyak mahasiswa merasa bingung tentang apa yang akan mereka kerjakan dalam hidupnya setelah tamat dari perguruan tinggi [1]. Hal ini merepresentasikan ketidaksiapan mereka dalam melakukan persiapan kerja yang baik.

Data terbaru dari Badan Pusat Statistik (BPS) berdasarkan Survei Angkatan Kerja Nasional (Sakernas) pada bulan Agustus 2023 menunjukkan bahwa Indonesia, dengan jumlah populasi masyarakat sebanyak 278.696.200 jiwa, telah mencapai tingkat angkatan kerja sebanyak 147,71 juta orang [2]. Hal ini mengalami peningkatan sebanyak 3,99 juta orang dibandingkan dengan periode yang sama pada tahun 2022. Hal ini mencerminkan intensitas tinggi masyarakat dalam mencari peluang kerja.

Pada bulan Februari 2024, Universitas Negeri Surabaya meluluskan 1.380 mahasiswa yang berasal dari jenjang Ahli Madya, Sarjana Terapan, Sarjana, Magister, dan Doktor dalam Wisuda Periode ke-109 [3]. Hal ini mengalami penurunan dibandingkan periode sebelumnya, di mana Wisuda Periode ke-108 pada bulan Oktober 2023 meluluskan 1.501 mahasiswa [4]. Meskipun mengalami penurunan wisudawan sekitar 8,06 persen, jumlah lulusan Universitas Negeri Surabaya setiap periodenya tetap tinggi, sehingga setiap individu harus bersaing di lapangan untuk memperoleh pekerjaan setelah lulus. Persaingan untuk mendapatkan kerja ini semakin ketat setiap waktunya, karena ketersediaan lapangan pekerjaan tidak seimbang dengan keberadaan para pencari kerja. Maka, individu yang memiliki kesiapan kerja berkesempatan besar untuk memenangkan perebutan pekerjaan yang diinginkan [5].

Tingkat pengangguran yang masih tinggi menunjukkan perlunya peningkatan signifikan dalam kesiapan kerja bagi mahasiswa yang akan bersaing di pasar kerja yang kompetitif. Pada pendidikan tinggi, mahasiswa semester akhir pada tingkat sarjana adalah para calon lulusan yang akan melanjutkan ke dunia kerja. Mahasiswa dituntut untuk mampu mengimbangi mutu dan kualitas yang diperlukan oleh perusahaan [6].

Oleh karena itu, analisis terhadap tingkat kesiapan kerja bagi mahasiswa diperlukan untuk mengungkap hal-hal yang mempengaruhi kesiapan kerja mahasiswa. Analisis dilakukan dengan melakukan proses klasifikasi. Klasifikasi berupaya untuk memprediksi kelas tujuan dengan presisi tertinggi. Algoritma klasifikasi mencari hubungan antara atribut masukan dan atribut keluaran untuk membangun sebuah model dalam proses training [7].

Dalam penelitian ini, klasifikasi dibangun dengan *decision tree*. Aspek kesiapan kerja dalam penelitian ini diambil dari delapan kompetensi kesiapan karier dari NACE. Beberapa aspek tersebut, di antaranya adalah *Career and Self Development, Communication, Critical Thinking, Equity and Inclusion, Leadership, Professionalism, Teamwork, dan*

Technology [8]. Klasifikasi ini dapat membantu mahasiswa untuk dapat memahami hal-hal yang memengaruhi kesiapan kerja. Dengan demikian, mahasiswa dapat meningkatkan kemampuan dirinya dalam bersiap untuk memasuki dunia kerja.

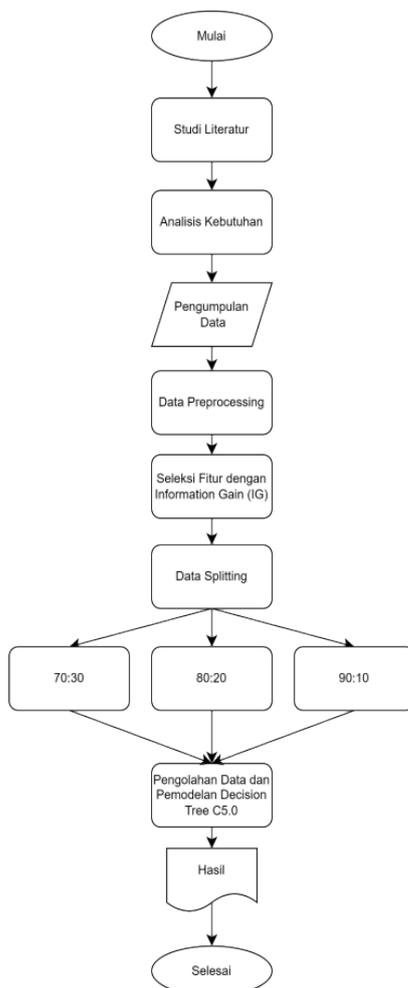
Algoritma C5.0 adalah sebuah algoritma *decision tree* (pohon keputusan). Algoritma ini sesuai untuk kumpulan data besar dan merupakan pengembangan dari algoritma C4.5 yang telah ada sebelumnya, menjadikan C5.0 lebih akurat. [9].

Berjalan bersama dengan algoritma C5.0, penelitian ini menggunakan teknik *data preprocessing* yang meliputi SMOTE (*Synthetic Minority Oversampling Technique*) untuk menangani masalah ketidakseimbangan data dan seleksi fitur Information Gain guna memilih fitur-fitur yang relevan terhadap penelitian.

Diharapkan, penelitian ini dapat menghasilkan hasil evaluasi yang baik dalam mengklasifikasikan tingkat kesiapan kerja mahasiswa.

## II. METODE PENELITIAN

Gbr 1 berikut menampilkan *flowchart* metode penelitian.



Gbr 1. Flowchart Metode Penelitian

### A. Data Mining

Data mining adalah bidang ilmu yang digunakan untuk menangani masalah pengambilan informasi dari *database* yang besar dengan menggabungkan teknik dari statistik, pembelajaran mesin, visualisasi data, pengenalan pola, dan *database* [10]. Terdapat beberapa fungsi dari data mining, di antaranya: (a) deskripsi, (b) klasifikasi, (c) *clustering*, (d) asosiasi, (e) *sequencing*, (f) *forecasting*, (g) prediksi [11]. Setiap fungsi memiliki tujuan utama untuk menciptakan efisiensi pengumpulan data besar guna menghasilkan informasi yang berguna.

Dalam tahapan proses KDD (*Knowledge Discovery in Database*) di dalam *data mining*, terdapat beberapa urutan proses, di antaranya *selection*, *preprocessing*, *transformation*, *data mining*, *evaluation*, dan *knowledge* [12]. Proses ini, secara berurutan, bertugas untuk menyeleksi, melakukan pra-proses, transformasi, dan penambahan data yang akan melalui tahapan evaluasi untuk menghasilkan informasi yang bermanfaat berdasarkan hasil analisis yang telah dicapai.

### B. SMOTE (*Synthetic Minority Oversampling Technique*)

*Synthetic Minority Oversampling Technique* (SMOTE) adalah metode *oversampling* yang paling sering digunakan dan pendekatan yang populer untuk menangani permasalahan machine learning mengenai ketidakseimbangan data [13].

Metode ini bekerja dengan membentuk sampel baru yang bersumber dari kelas minoritas untuk membuat proporsi data menjadi seimbang [14]. Proses SMOTE dimulai dengan memilih setiap sampel dari kelas minoritas secara berurutan sebagai sampel dasar untuk sintesis sampel baru, di mana proses ini diulang sebanyak  $n$  kali. Setelahnya, interpolasi linier dilakukan antara sampel dasar dan setiap sampel tambahan untuk menghasilkan  $n$  sampel sintetis [15].

### C. Seleksi Fitur

Seleksi Fitur adalah sebuah proses untuk memilih fitur yang relevan dan menghapus yang fitur yang tidak relevan dan berlebihan yang bertujuan untuk mendapatkan subset 17 dengan kinerja terbaik [16]. Terdapat beberapa metode seleksi fitur, di antaranya *Filter*, *Wrapper*, dan *Embedded*. Dalam penelitian ini, metode seleksi fitur yang digunakan adalah seleksi fitur berdasarkan pemeringkatan nilai *information gain* yang merupakan bagian dari metode *filter*.

Pemeringkatan *information gain* digunakan untuk mengevaluasi nilai setiap variabel dengan mengukur perolehan entropi sehubungan dengan hasilnya [17]. Dalam proses seleksi fitur, pemilihan jumlah atribut yang diseleksi dilihat dari nilai setiap atribut beserta nilai *threshold* (ambang batas) yang dapat ditentukan sendiri oleh *user* [18]. Oleh karena itu, penulis akan menggunakan  $\text{threshold} > 0.05$  dalam seleksi fitur.

### D. Decision Tree

*Decision tree* merupakan salah satu metode yang sering digunakan dalam berbagai bidang, seperti pemrosesan gambar, pembelajaran mesin, dan identifikasi pola. Secara esensial, *decision tree* berfungsi untuk melakukan pengkategorian atau pengelompokan data.

Struktur dari *decision tree* terdiri atas node dan cabang-cabang yang menggambarkan setiap pohon. Tiap node merepresentasikan fitur dalam kategori yang akan diklasifikasikan, sementara setiap subset mendefinisikan nilai yang dapat diambil oleh node tersebut[19]. Keunggulan *decision tree* terletak pada proses analisis yang sederhana dan tingkat presisi yang tinggi terhadap berbagai jenis data, sehingga metode ini telah berhasil diimplementasikan secara luas dalam berbagai bidang [20].

Sebuah *decision tree* terbentuk dari beberapa elemen, di antaranya *Root Node* (Simpul Akar), *Branches* (Ranting), dan *Leaf Node* (Simpul Daun). Klasifikasi menggunakan *decision tree* dilakukan oleh *routing* dari simpul akar hingga simpul daun. Beberapa contoh algoritma *decision tree*, di antaranya adalah ID3, C4.5, C5.0, CART, CHAID, dan *Random Forest*.

E. Algoritma C5.0

C5.0 adalah sebuah algoritma *decision tree* yang merupakan pengembangan dari C4.5. Algoritma C4.5 memiliki kelemahan pada terjadinya overlapping terutama saat data yang dikelola sangat banyak [21]. Pada dasarnya, C4.5 juga merupakan hasil pengembangan dari ID3, sehingga keilmuan yang digunakan terus berkembang dan mengalami kemiripan. C5.0 menyediakan seleksi fitur, penggunaan teknik cross validation, serta fasilitas pengurangan *error pruning* [22]. Algoritma C5.0 memiliki beberapa keuntungan, di antaranya tingkat akurasi yang dicapai lebih tinggi, lebih cepat dalam pengambilan keputusan, serta menggunakan memori yang jauh lebih rendah dari sebelumnya [21]. Tabel I di bawah ini berisi daftar perbedaan antara algoritma C4.5 dengan algoritma C5.0 [22, 23].

TABEL I  
 PERBEDAAN ALGORITMA C4.5 DAN C5.0

Kategori	Algoritma C4.5	Algoritma C5.0
Tingkat Pruning Error	Lebih tinggi	Lebih rendah
Akurasi	Lebih rendah	Lebih tinggi
Kecepatan	Pembuatan <i>rules tree</i> memerlukan 8 jam atau lebih	Pembuatan <i>rules tree</i> memerlukan kurang dari 3 menit
Memori	Pembuatan <i>rules tree</i> memerlukan memori lebih dari 3 GB	Pembuatan <i>rules tree</i> memerlukan memori kurang dari 200 MB
Tahapan Penghitungan	<i>Entropy - Information Gain</i>	<i>Entropy - Information Gain - Split Info - Gain Ratio</i>
Penentuan Node Akar	Dihasilkan dari nilai <i>Information Gain</i> tertinggi	Dihasilkan dari nilai <i>Gain Ratio</i> tertinggi

Pada algoritma C5.0, pembentukan node akar dan cabang unti setiap node di dalam *decision tree* dilakukan menggunakan nilai *gain ratio* tertinggi. Node akar dihasilkan dari penghitungan nilai *entropy*, *information gain*, *split info*,

lalu *gain ratio* [23]. Tahapan dalam membuat *decision tree* dalam algoritma C5.0 adalah:

1. Menentukan variabel atau atribut yang akan digunakan.
2. Mempersiapkan data latih (*training*) yang diambil dari survei dan dikategorikan dalam beberapa kelas tertentu sesuai atributnya.
3. Menentukan *node* akar dengan menghitung nilai *gain ratio* dari setiap atribut dan mencari nilai *gain ratio* tertinggi [23]. Rumus (1) berikut ini menunjukkan cara menghitung *entropy*.

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

S adalah himpunan kasus, n adalah jumlah partisi dalam S, dan  $p_i$  adalah proporsi  $S_i$  terhadap S. Nilai entropy memiliki rentang dari 0 hingga 1.

4. Menghitung *information gain*  
 Rumus (2) berikut ini menunjukkan cara menghitung *information gain*.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S) \dots\dots\dots(2)$$

S adalah himpunan kasus, A adalah atribut yang diproses, n adalah jumlah partisi di dalam atribut A,  $S_i$  adalah jumlah kasus partisi partisi atribut A,  $|S_i|$  adalah jumlah kasus partisi i, dan |S| adalah jumlah kasus dalam S.

5. Menghitung *split info*  
 Rumus (3) berikut ini menunjukkan cara menghitung *split info*.

$$Split Info(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{S} \dots\dots\dots(3)$$

S adalah himpunan kasus, A adalah atribut yang diproses,  $S_i$  adalah proporsi  $S_i$  terhadap S, dan S adalah jumlah kasus dalam S.

6. Menghitung *gain ratio*  
 Rumus (4) berikut ini menunjukkan cara menghitung *gain ratio*.

$$Gain Ratio(S, A) = \frac{Gain(S,A)}{Split Info(S,A)} \dots\dots\dots(4)$$

Proses perhitungan *gain ratio* diulangi hingga setiap cabang mempunyai kelas masing-masing.

F. Teknik Kategorisasi Kesiapan Kerja

Kategorisasi kesiapan kerja dibagi ke dalam dua jenjang, siap kerja dan tidak siap kerja. Skala yang digunakan dalam pengukuran kesesuaian pernyataan dari setiap atribut kompetensi kerja yang ada adalah skala likert interval 1-4.

Tabel II menampilkan penghitungan kategorisasi yang diperoleh dari rumus [24]:

- 4 = Sangat Sesuai (SS)
- 3 = Sesuai (S)
- 2 = Tidak Sesuai (TS)
- 1 = Sangat Tidak Sesuai (STS)

TABEL III  
KATEGORISASI KESIAPAN KERJA

Rumus	Kategori
$X \geq \mu$	Siap Kerja
$X < \mu$	Tidak Siap Kerja

X adalah skor mentah sampel dan  $\mu$  adalah rata-rata distribusi dalam populasi.

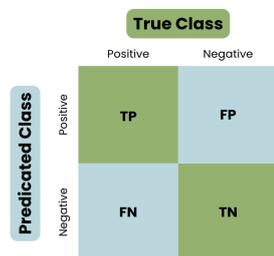
G. Python

Python adalah sebuah bahasa pemrograman tingkat tinggi yang ditafsirkan, berorientasi objek, dan bahasa dengan semantik dinamis. Python dapat berjalan dalam platform apapun yang sudah memiliki Python interpreter di dalamnya, seperti sistem operasi Windows, macOS, dan Linux. Python memiliki beberapa manfaat [25]. Manfaat tersebut, di antaranya:

1. Memiliki kode yang lebih pendek dibandingkan dengan bahasa pemrograman lain.
2. Mendukung beberapa paradigma pemrograman.
3. Memiliki dukungan komunitas yang besar.
4. Dilengkapi dengan sejumlah besar libraries sebagai bagian dari penginstal standar.
5. Pengetikan yang dinamis membantu mengurangi stres pada para coder baru, di mana tipe data diatur secara otomatis.
6. Memiliki libraries pihak ketiga yang luas untuk berbagai bagian kerja yang beragam.

H. Evaluasi

Uji akurasi menggunakan *confusion matrix* digunakan untuk penyelesaian masalah klasifikasi. Uji ini menunjukkan berapa banyak prediksi yang terbukti benar dan salah pada setiap kelas yang ada. Contoh *confusion matrix* dapat dilihat pada Gbr 2 [26].



Gbr 2. Confusion Matrix

1. True Positive (TP) : prediksi positif, terbukti positif
2. True Negative (TN) : prediksi negatif, terbukti negatif

3. False Positive (FP) : prediksi positif, terbukti negatif
4. False Negative (FN) : prediksi negatif, terbukti Positif

Terdapat beberapa metrik yang digunakan untuk mengevaluasi performa pengklasifikasi. Metrik tersebut, di antaranya akurasi, presisi, dan *recall*. Penghitungan metrik-metrik tersebut dijelaskan sebagai rumus (5), (6), dan (7) [27].

1. Menghitung akurasi  

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \dots\dots\dots(5)$$

2. Menghitung presisi  

$$Presisi = \frac{TP}{TP+FP} \times 100\% \dots\dots\dots(6)$$

3. Menghitung *recall*  

$$Recall = \frac{TP}{TP+FN} \times 100\% \dots\dots\dots(7)$$

III. HASIL DAN PEMBAHASAN

A. Persiapan Data

*Dataset* diperoleh dari hasil survei yang disebarakan kepada 114 orang mahasiswa semester akhir. Tabel III memaparkan 16 butir atribut survei berdasarkan 8 kompetensi kesiapan karier [8].

TABEL IIIIIII  
ATRIBUT SURVEI KESIAPAN KERJA MAHASISWA

Nomor	Pernyataan Survei	Atribut	Kompetensi
1.	Saya dapat membangun rencana dan tujuan pasca-studi yang saya lakukan	A11	Karier dan Pengembangan diri
2.	Saya secara sukarela berpartisipasi dalam pendidikan lanjutan, pelatihan, dan atau kegiatan lain yang menunjang karier saya	A12	
3.	Saya berkomunikasi secara jelas dan terorganisir sehingga orang lain dapat memahami maksud saya secara efektif	A21	Komunikasi
4.	Saya dapat mendemonstrasikan sesuatu secara verbal (lisan/tertulis) dan non-verbal (bahasa tubuh)	A22	
5.	Saya mengumpulkan dan menganalisis informasi dari beragam sumber dan individu untuk memahami suatu masalah secara sepenuhnya	A31	Berpikir Kritis
6.	Saya melakukan banyak tugas dengan	A32	

	baik dalam lingkungan yang serba cepat		
7.	Saya berpikiran terbuka terhadap ide yang beragam dan cara berpikir baru	A41	Ekuitas dan Inklusi
8.	Saya menunjukkan fleksibilitas diri dengan beradaptasi pada lingkungan yang beragam	A42	
9.	Saya merencanakan, memulai, mengelola, menyelesaikan, dan mengevaluasi proyek yang saya kerjakan	A51	Kepemimpinan
10.	Saya menggunakan pemikiran inovatif untuk melampaui cara berpikir tradisional	A52	
11.	Saya selalu hadir dan bersiap dalam setiap kegiatan yang saya jalani	A61	Profesionalisme
12.	Saya memprioritaskan dan menyelesaikan tugas untuk mencapai tujuan organisasi	A62	
13.	Saya mendengarkan orang lain, memahami maksud mereka, dan menanyakan hal yang tepat dengan baik tanpa menyela	A71	Kerja Tim
14.	Saya berkompromi dengan orang lain dan bekerja dengan tangkas	A72	
15.	Saya memanfaatkan teknologi untuk meningkatkan efisiensi dan produktivitas kerja	A81	Teknologi
16.	Saya cepat beradaptasi dengan teknologi baru atau asing	A82	

### B. Data Preprocessing

Tahapan ini meliputi penghapusan atribut yang tidak digunakan, penggantian nama atribut, *handling missing values*, penghapusan duplikasi data, penjumlahan baris data, serta penentuan kolom variabel serta kolom target. Setelah itu, dilakukan pelabelan terhadap data yang telah *melalui preprocessing*. Tabel IV menampilkan perubahan nama atribut survei untuk mempermudah pemahaman penulis terhadap setiap atribut survei yang ada.

TABEL IVV  
PERUBAHAN NAMA ATRIBUT

Nomor	Pernyataan Survei	Atribut Awal	Atribut Baru
1.	Saya dapat membangun rencana dan tujuan pasca-studi yang saya lakukan	A11	studi

2.	Saya secara sukarela berpartisipasi dalam pendidikan lanjutan, pelatihan, dan atau kegiatan lain yang menunjang karier saya	A12	pelatihan
3.	Saya berkomunikasi secara jelas dan terorganisir sehingga orang lain dapat memahami maksud saya secara efektif	A21	komorganisir
4.	Saya dapat mendemonstrasikan sesuatu secara verbal (lisan/tertulis) dan non-verbal (bahasa tubuh)	A22	demonstrasi
5.	Saya mengumpulkan dan menganalisis informasi dari beragam sumber dan individu untuk memahami suatu masalah secara sepenuhnya	A31	amnalisis
6.	Saya melakukan banyak tugas dengan baik dalam lingkungan yang serba cepat	A32	multitask
7.	Saya berpikiran terbuka terhadap ide yang beragam dan cara berpikir baru	A41	openmind
8.	Saya menunjukkan fleksibilitas diri dengan beradaptasi pada lingkungan yang beragam	A42	fleksibel
9.	Saya merencanakan, memulai, mengelola, menyelesaikan, dan mengevaluasi proyek yang saya kerjakan	A51	perencanaan
10.	Saya menggunakan pemikiran inovatif untuk melampaui cara berpikir tradisional	A52	inovatif
11.	Saya selalu hadir dan bersiap dalam setiap kegiatan yang saya jalani	A61	kebersiapan
12.	Saya memprioritaskan dan menyelesaikan tugas untuk mencapai tujuan organisasi	A62	prioritas
13.	Saya mendengarkan orang lain, memahami maksud mereka, dan menanyakan hal yang tepat dengan baik tanpa menyela	A71	komefektif
14.	Saya berkompromi dengan orang lain dan bekerja dengan tangkas	A72	tangkas
15.	Saya memanfaatkan	A81	manfaattek

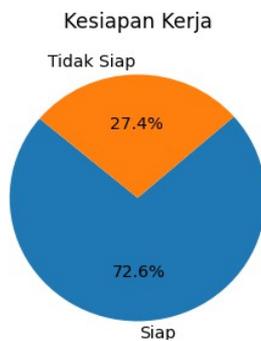
	teknologi untuk meningkatkan efisiensi dan produktivitas kerja		
16.	Saya cepat beradaptasi dengan teknologi baru atau asing	A82	adaptasitek

Setelah *preprocessing* dilakukan, didapatkan dataset seperti yang ditampilkan pada Gbr 3.

	stud	pelatih an ganisir	komor demonstrasi	analisis	multitask	openmind	fleksibel	perencanaan	inovatif	kebersihan	prioritas	komefektif	tangkas	manfaat	adaptasitek	kesiapan_kerja
0	3	1	1	1	2	3	2	2	1	2	3	3	3	3	4	2
1	2	1	1	1	2	4	4	3	3	3	4	1	3	3	3	2
2	4	3	4	2	4	1	3	2	2	3	1	3	1	3	4	4
3	4	2	3	3	4	4	4	4	4	4	4	4	4	1	4	4
4	2	2	2	1	2	4	2	1	2	4	1	2	4	4	4	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
108	2	2	1	3	2	3	3	1	3	3	2	4	2	3	3	3
109	1	2	2	1	1	3	2	4	2	2	1	4	4	3	3	3
110	1	3	3	3	3	3	3	3	1	3	4	4	4	3	3	2
111	1	2	1	1	1	3	2	2	2	3	4	1	2	3	4	3
112	2	4	1	3	1	3	2	3	3	2	2	2	4	2	3	3

Gbr 3. Dataset Setelah Preprocessing

Berdasarkan *dataset* di atas, dihasilkan *pie chart* kesiapan kerja yang ditampilkan pada Gbr 4.



Gbr 4. Pie Chart Kesiapan Kerja

Terdapat ketidakseimbangan data pada Gbr 4 di atas. 27.4% responden masuk ke dalam kategori 'Tidak Siap' kerja dan 72.6% masuk kategori 'Siap' kerja. Oleh karena itu, *oversampling* dengan SMOTE (*Synthetic Minority Oversampling Technique*) untuk menyeimbangkan jumlah data pada kolom target. Gbr 5 di bawah ini menampilkan kode yang digunakan untuk menjalankan SMOTE.

```
import pandas as pd
from imblearn.over_sampling import SMOTE
import matplotlib.pyplot as plt

# DataFrame yang Anda miliki
df = preprocessed_data

# Menghitung jumlah responden yang 'Siap' dan 'Tidak Siap' Kerja
counts = df['kesiapan_kerja'].value_counts()
print("Distribusi sebelum oversampling:")
print(counts)

# Menyiapkan fitur (X) dan label (y)
X = df.drop(columns=['kesiapan_kerja'])
y = df['kesiapan_kerja']

# Melakukan oversampling menggunakan SMOTE
smote = SMOTE(random_state=42)
X_res, y_res = smote.fit_resample(X, y)

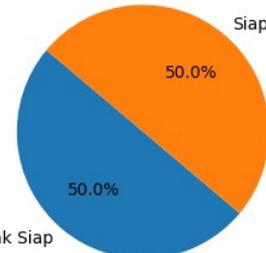
# Membuat DataFrame hasil oversampling
df_resampled = pd.concat([X_res, y_res], axis=1)

# Menghitung kembali jumlah responden setelah oversampling
counts_resampled = y_res.value_counts()
print("\nDistribusi setelah oversampling:")
print(counts_resampled)
```

Gbr 5. Source Code SMOTE

Setelah SMOTE dijalankan, dibuatlah sebuah *pie chart* yang menampilkan persentase distribusi *preprocessed data* setelah *oversampling*. *Pie chart* ditampilkan pada Gbr 6.

Kesiapan Kerja (Setelah Oversampling)



Gbr 6. Pie Chart Kesiapan Kerja Setelah Oversampling

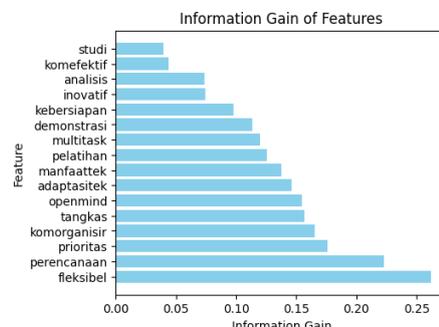
Setelah proses *oversampling* selesai, jumlah baris data pada dataset meningkat menjadi 164 data dari 114 data. Gbr 6 menunjukkan hasil pelabelan keseluruhan data, di mana 50% responden dikategorikan sebagai 'Tidak Siap' kerja dan 50% dikategorikan sebagai 'Siap' kerja. Dengan proporsi ini, *dataset* siap digunakan dalam penelitian. *Dataset* yang telah melalui proses *oversampling* ditampilkan pada Gbr 7.

	stud	pelatih an ganisir	komor demonstrasi	analisis	multitask	openmind	fleksibel	perencanaan	inovatif	kebersihan	prioritas	komefektif	tangkas	manfaat	adaptasitek	kesiapan_kerja
0	3	1	1	1	2	3	2	2	1	2	3	3	3	3	4	2
1	2	1	1	1	2	4	4	3	3	3	4	1	3	3	3	2
2	4	3	4	2	4	1	3	2	2	3	1	3	1	3	4	4
3	4	2	3	3	4	4	4	4	4	4	4	4	4	1	4	4
4	2	2	2	1	2	4	2	1	2	4	1	2	4	4	4	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
159	3	2	1	2	2	3	2	2	3	3	2	2	2	2	2	1
160	2	1	2	2	1	2	3	1	1	3	2	3	3	2	3	3
161	3	2	3	1	3	1	3	2	2	2	1	3	2	3	2	1
162	2	1	1	1	2	3	2	1	1	2	2	2	3	3	4	2
163	1	2	1	3	2	2	3	2	1	2	2	2	3	2	3	2

Gbr 7. Dataset Setelah Oversampling

C. Seleksi Fitur

Pada tahapan ini, seleksi fitur dilakukan terhadap *dataset* hasil *oversampling* untuk mengurangi jumlah atribut survei yang digunakan sebagai variabel penelitian. Metode seleksi fitur yang digunakan adalah seleksi fitur *information gain*. Penghitungan *information gain* dilakukan menggunakan *source code*. Gbr 8 di bawah ini menampilkan nilai *information gain* dari 16 atribut survei dalam bentuk diagram batang.



Gbr 8. Diagram Batang Nilai *Information Gain* Setiap Atribut

Dari Gbr 8 di atas, diketahui bahwa atribut *fleksibel* memiliki nilai *information gain* tertinggi dan atribut *studi* memiliki nilai *information gain* terendah. Secara rinci, Tabel V menjabarkan nilai *information gain* setiap atribut.

TABEL V  
NILAI INFORMATION GAIN SETIAP ATRIBUT

Atribut Terpilih	Information Gain
fleksibel	0.2624064809716422
perencanaan	0.2232287453696591
prioritas	0.17575310925068965
komorganisir	0.16515980589090384
tangkas	0.15672540332823526
openmind	0.1543855546212918
adaptasitek	0.14601833595133362
manfaattek	0.137774215498062
pelatihan	0.12590296686272318
multitask	0.11957577938417274
demonstrasi	0.11352768846978445
kebersiapan	0.09761833595689495
inovatif	0.07416758832049375
analisis	0.07382398333911033
komefektif	0.04367914914077109
studi	0.03923570956239564

Dengan informasi di atas,  $threshold > 0.05$  diterapkan untuk melakukan seleksi terhadap keseluruhan fitur yang ada. Dari seleksi fitur yang telah dilakukan, terpilih 14 atribut yang sesuai dengan threshold yang ditentukan. Atribut-atribut yang terpilih ditampilkan pada Tabel VI.

TABEL VI  
ATRIBUT TERPILIH

Atribut Terpilih	Information Gain
fleksibel	0.2624064809716422
perencanaan	0.2232287453696591
prioritas	0.17575310925068965
komorganisir	0.16515980589090384
tangkas	0.15672540332823526
openmind	0.1543855546212918
adaptasitek	0.14601833595133362
manfaattek	0.137774215498062
pelatihan	0.12590296686272318
multitask	0.11957577938417274
demonstrasi	0.11352768846978445
kebersiapan	0.09761833595689495
inovatif	0.07416758832049375
analisis	0.07382398333911033

D. Penghitungan *Entropy*, *Information Gain*, *Split Info*, dan *Gain Ratio* Setiap Atribut

*Entropy*, *Information Gain*, *Split Info*, dan *Gain Ratio* dihitung menggunakan rumus (1), (2), (3), dan (4) dalam proses komputasi menggunakan *source code*. Hasil penghitungan elemen-elemen algoritma C5.0 ditampilkan pada Tabel VII.

TABEL VII  
OUTPUT PENGHITUNGAN *ENTROPY*, *INFORMATION GAIN*, *SPLIT INFO*, DAN *GAIN RATIO* SETIAP ATRIBUT

Atribut Terpilih	Entropy	Information Gain	Split Info	Gain Ratio
fleksibel	1.815396517 523369	0.262406480971 6422	1.815396517 523369	0.174054 4499310 8252
perencanaan	1.783488547 1098787	0.223228745369 6591	1.78348 85471098787	0.195615 3823296 8054
prioritas	1.699432842 4474663	0.175753109250 68965	1.699432842 4474663	0.114773 9195164 7964
komorganisir	1.794070169 1322296	0.165159805890 90384	1.79407 01691322296	0.048656 7109279 5721
tangkas	1.728361601 5908798	0.156725403328 23526	1.728361601 5908798	0.077241 9787259 5487
openmind	1.723391701 9692635	0.15438555462 12918	1.72339 17019692635	0.148864 0290858 6654
adaptasitek	1.835054824 4334444	0.146018335951 33362	1.835054824 4334444	0.115081 8737486 2565
manfaattek	1.705634314 8957458	0.13777421549 8062	1.705634314 8957458	0.112699 61 9880440 36
pelatihan	1.838553894 560433	0.125902966862 72318	1.838553894 560433	0.090360 3325132 6987
multitask	1.582371130 9109365	0.119575779841 7274	1.582371130 9109365	0.075426 2813475 389
demonstrasi	1.782112704 258005	0.113527688469 78445	1.782112704 258005	0.098540 6989694 5392
kebersiapan	1.894024298 9747536	0.097618335956 89495	1.894024298 9747536	0.091635 6887750 5933
inovatif	1.611432734 8245556	0.074167588320 49375	1.611432734 8245556	0.038950 4852701 96136
analisis	1.818337624 3592146	0.073823983339 11033	1.818337624 3592146	0.075752 7121337 9803

Dari tabel di atas, diketahui atribut *perencanaan* memiliki *gain ratio* tertinggi. Maka, atribut ini akan menjadi *root node* dalam *decision tree* yang akan dibangun.

E. Skema Pembagian Data

Skema pembagian data yang akan digunakan sebagai data latih (*training*) dan data uji (*testing*) ditampilkan pada Tabel VIII.

TABEL VIII  
SKEMA PENGUJIAN DATA

Skema ke-n	Persentase Data Latih	Persentase Data Uji	Keterangan
1	70%	30%	70% dari jumlah dataset akan menjadi data latih dan 30% menjadi data uji



Prediksi	1 (Negative)	17	2
----------	--------------	----	---

Berdasarkan Tabel IX di atas, diketahui bahwa jumlah data pada skema splitting 70:30 adalah 114 data training dan 50 data testing. Selanjutnya, penghitungan secara manual dilakukan untuk mengukur akurasi, presisi, dan recall dari klasifikasi dengan skema 70:30.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$= \frac{26 + 17}{26 + 17 + 5 + 2} \times 100\% = 86\%$$

$$Presisi = \frac{TP}{TP + FP} \times 100\%$$

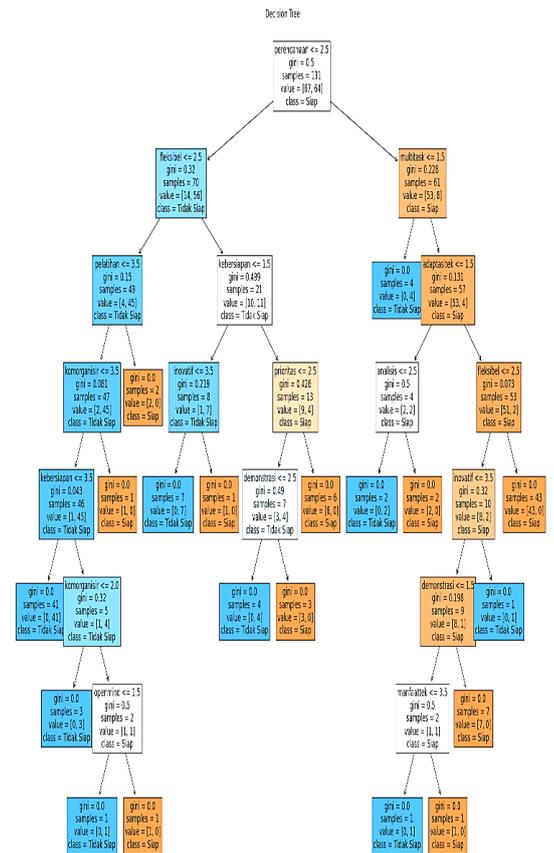
$$= \frac{26}{26 + 5} \times 100\% = 83.87\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$= \frac{26}{26 + 2} \times 100\% = 92.85\%$$

2. Skema 80:20

Gbr 10 berikut merupakan hasil *decision tree* untuk skema 80:20.



Gbr 10. *Decision Tree* Skema 80:20

Pada Gbr 10, atribut *perencanaan* menjadi *root node* dalam *tree* ini. *perencanaan* <= 2.5 memiliki arti bahwa responden dengan skor kurang dari atau sama dengan 2.5 jika True akan bercabang ke atribut *multitask* dan jika False akan bercabang ke atribut *fleksibel*. Percabangan ini dilanjutkan hingga didapatkan nilai gini 0.0. Pada skema ini, 33 sampel data uji terbagi menjadi 18 orang ‘Siap’ kerja dan 15 orang ‘Tidak Siap’ kerja.

Dengan *decision tree* yang telah terbentuk, *rules* untuk skema 80:20 dibentuk dengan daftar sebagai berikut:

1. If perencanaan <= 2.5, fleksibel <= 2.5, adaptasitek <= 3.5, pelatihan <= 3.5, kebersiapan <= 3.5, then predicted class is Tidak Siap
2. If perencanaan <= 2.5, fleksibel <= 2.5, adaptasitek <= 3.5, pelatihan <= 3.5, kebersiapan > 3.5, pelatihan <= 1.5, then predicted class is Tidak Siap
3. If perencanaan <= 2.5, fleksibel <= 2.5, adaptasitek <= 3.5, pelatihan <= 3.5, kebersiapan > 3.5, pelatihan > 1.5, komorganisir <= 2.0, then predicted class is Tidak Siap

4. If perencanaan  $\leq 2.5$ , fleksibel  $\leq 2.5$ , adaptasitek  $\leq 3.5$ , pelatihan  $\leq 3.5$ , kebersiapan  $> 3.5$ , pelatihan  $> 1.5$ , komorganisir  $> 2.0$ , then predicted class is Siap
5. If perencanaan  $\leq 2.5$ , fleksibel  $\leq 2.5$ , adaptasitek  $\leq 3.5$ , pelatihan  $> 3.5$ , then predicted class is Siap
6. If perencanaan  $\leq 2.5$ , fleksibel  $\leq 2.5$ , adaptasitek  $> 3.5$ , then predicted class is Siap
7. If perencanaan  $\leq 2.5$ , fleksibel  $> 2.5$ , kebersiapan  $\leq 1.5$ , inovatif  $\leq 3.5$ , then predicted class is Tidak Siap
8. If perencanaan  $\leq 2.5$ , fleksibel  $> 2.5$ , kebersiapan  $\leq 1.5$ , inovatif  $> 3.5$ , then predicted class is Siap
9. If perencanaan  $\leq 2.5$ , fleksibel  $> 2.5$ , kebersiapan  $> 1.5$ , prioritas  $\leq 2.5$ , demonstrasi  $\leq 2.5$ , then predicted class is Tidak Siap
10. If perencanaan  $\leq 2.5$ , fleksibel  $> 2.5$ , kebersiapan  $> 1.5$ , prioritas  $\leq 2.5$ , demonstrasi  $> 2.5$ , then predicted class is Siap
11. If perencanaan  $\leq 2.5$ , fleksibel  $> 2.5$ , kebersiapan  $> 1.5$ , prioritas  $> 2.5$ , then predicted class is Siap
12. If perencanaan  $> 2.5$ , multitask  $\leq 1.5$ , then predicted class is Tidak Siap
13. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $\leq 1.5$ , analisis  $\leq 2.5$ , then predicted class is Tidak Siap
14. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $\leq 1.5$ , analisis  $> 2.5$ , then predicted class is Siap
15. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $> 1.5$ , fleksibel  $\leq 2.5$ , inovatif  $\leq 3.5$ , demonstrasi  $\leq 1.5$ , kebersiapan  $\leq 2.5$ , then predicted class is Tidak Siap
16. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $> 1.5$ , fleksibel  $\leq 2.5$ , inovatif  $\leq 3.5$ , demonstrasi  $\leq 1.5$ , kebersiapan  $> 2.5$ , then predicted class is Siap
17. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $> 1.5$ , fleksibel  $\leq 2.5$ , inovatif  $\leq 3.5$ , demonstrasi  $> 1.5$ , then predicted class is Siap
18. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $> 1.5$ , fleksibel  $\leq 2.5$ , inovatif  $> 3.5$ , then predicted class is Tidak Siap
19. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $> 1.5$ , fleksibel  $> 2.5$ , then predicted class is Siap

Tabel *confusion matrix* dari klasifikasi dengan skema 80:20 dapat dilihat pada Tabel X.

TABEL X  
CONFUSION MATRIX (80:20)

Confusion Matrix		Aktual	
		0 (True)	1 (False)
Prediksi	0 (Positive)	16	0
	1 (Negative)	15	2

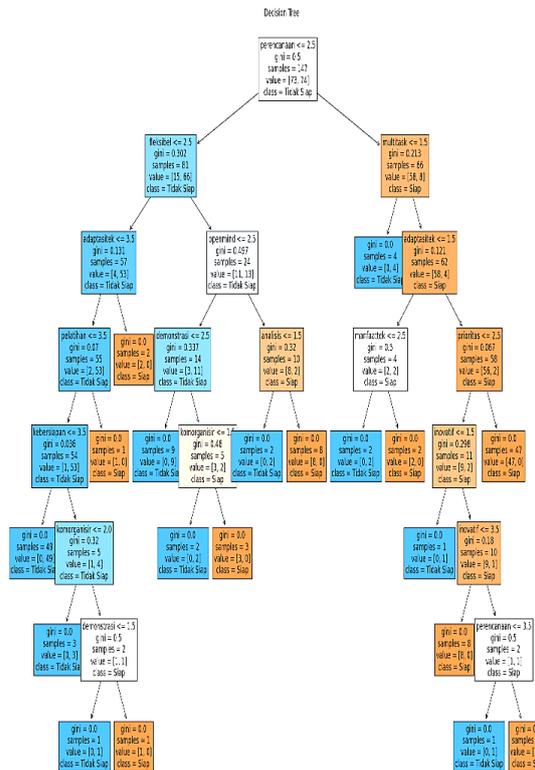
Berdasarkan Tabel X di atas, diketahui bahwa jumlah data pada skema splitting 80:20 adalah 131 data training dan 33 data testing. Selanjutnya, penghitungan secara manual dilakukan untuk mengukur akurasi, presisi, dan recall dari klasifikasi dengan skema 80:20.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{16 + 15}{16 + 15 + 0 + 2} \times 100\% \\
 &= 93.93\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Presisi} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{16}{16 + 0} \times 100\% = 100\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{16}{16 + 2} \times 100\% = 88.88\%
 \end{aligned}$$

3. Skema 90:10  
Gbr 11 berikut merupakan hasil *decision tree* untuk skema 90:10.



Gbr 11. Decision Tree Skema 90:10

Pada Gbr 11, atribut perencanaan tetap menjadi root node dalam tree ini. perencanaan  $\leq 2.5$  memiliki arti bahwa responden dengan skor kurang dari atau sama dengan 2.5 jika True akan bercabang ke atribut multitask dan jika False akan bercabang ke atribut fleksibel. Percabangan ini dilanjutkan hingga didapatkan nilai gini 0.0. Pada skema ini, 17 sampel data uji terbagi menjadi 8 orang ‘Siap’ kerja dan 9 orang ‘Tidak Siap’ kerja.

Dengan *decision tree* yang telah terbentuk, rules untuk skema 90:10 dibentuk dengan daftar sebagai berikut:

1. If perencanaan  $\leq 2.5$ , fleksibel  $\leq 2.5$ , adaptasitek  $\leq 3.5$ , pelatihan  $\leq 3.5$ , kebersiapan  $\leq 3.5$ , then predicted class is Tidak Siap
2. If perencanaan  $\leq 2.5$ , fleksibel  $\leq 2.5$ , adaptasitek  $\leq 3.5$ , pelatihan  $\leq 3.5$ , kebersiapan  $> 3.5$ , pelatihan  $\leq 1.5$ , then predicted class is Tidak Siap
3. If perencanaan  $\leq 2.5$ , fleksibel  $\leq 2.5$ , adaptasitek  $\leq 3.5$ , pelatihan  $\leq 3.5$ , kebersiapan  $> 3.5$ , pelatihan  $> 1.5$ , demonstrasi  $\leq 1.5$ , then predicted class is Tidak Siap
4. If perencanaan  $\leq 2.5$ , fleksibel  $\leq 2.5$ , adaptasitek  $\leq 3.5$ , pelatihan  $\leq 3.5$ , kebersiapan  $> 3.5$ , pelatihan  $> 1.5$ ,

demonstrasi  $> 1.5$ , then predicted class is Siap

5. If perencanaan  $\leq 2.5$ , fleksibel  $\leq 2.5$ , adaptasitek  $\leq 3.5$ , pelatihan  $> 3.5$ , then predicted class is Siap
6. If perencanaan  $\leq 2.5$ , fleksibel  $\leq 2.5$ , adaptasitek  $> 3.5$ , then predicted class is Siap
7. If perencanaan  $\leq 2.5$ , fleksibel  $> 2.5$ , openmind  $\leq 2.5$ , demonstrasi  $\leq 2.5$ , then predicted class is Tidak Siap
8. If perencanaan  $\leq 2.5$ , fleksibel  $> 2.5$ , openmind  $\leq 2.5$ , demonstrasi  $> 2.5$ , komorganisir  $\leq 1.5$ , then predicted class is Tidak Siap
9. If perencanaan  $\leq 2.5$ , fleksibel  $> 2.5$ , openmind  $\leq 2.5$ , demonstrasi  $> 2.5$ , komorganisir  $> 1.5$ , then predicted class is Siap
10. If perencanaan  $\leq 2.5$ , fleksibel  $> 2.5$ , openmind  $> 2.5$ , analisis  $\leq 1.5$ , then predicted class is Tidak Siap
11. If perencanaan  $\leq 2.5$ , fleksibel  $> 2.5$ , openmind  $> 2.5$ , analisis  $> 1.5$ , then predicted class is Siap
12. If perencanaan  $> 2.5$ , multitask  $\leq 1.5$ , then predicted class is Tidak Siap
13. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $\leq 1.5$ , analisis  $\leq 2.5$ , then predicted class is Tidak Siap
14. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $\leq 1.5$ , analisis  $> 2.5$ , then predicted class is Siap
15. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $> 1.5$ , prioritas  $\leq 2.5$ , inovatif  $\leq 1.5$ , then predicted class is Tidak Siap
16. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $> 1.5$ , prioritas  $\leq 2.5$ , inovatif  $> 1.5$ , inovatif  $\leq 3.5$ , then predicted class is Siap
17. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $> 1.5$ , prioritas  $\leq 2.5$ , inovatif  $> 1.5$ , inovatif  $> 3.5$ , tangkas  $\leq 2.5$ , then predicted class is Tidak Siap
18. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $> 1.5$ , prioritas  $\leq 2.5$ , inovatif  $> 1.5$ , inovatif  $> 3.5$ , tangkas  $> 2.5$ , then predicted class is Siap
19. If perencanaan  $> 2.5$ , multitask  $> 1.5$ , adaptasitek  $> 1.5$ , prioritas  $> 2.5$ , then predicted class is Siap

Tabel *confusion matrix* dari klasifikasi dengan skema 90:10 dapat dilihat pada Tabel XI.

TABEL XI  
CONFUSION MATRIX (90:10)

Confusion Matrix		Aktual	
		0 (True)	1 (False)
Prediksi	0 (Positive)	7	1
	1 (Negative)	8	1

Berdasarkan Tabel XI di atas, diketahui bahwa jumlah data pada skema splitting 90:10 adalah 147 data training dan 17 data testing. Selanjutnya, penghitungan secara manual dilakukan untuk mengukur akurasi, presisi, dan *recall* dari klasifikasi dengan skema 90:10.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{7 + 8}{7 + 8 + 1 + 1} \times 100\% \\
 &= 88.23\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Presisi} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{7}{7 + 1} \times 100\% = 87.5\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{7}{7 + 1} \times 100\% = 87.5\%
 \end{aligned}$$

### G. Pembahasan

Hasil evaluasi semua skema secara singkat ditampilkan pada Tabel XII.

TABEL XII  
HASIL EVALUASI SEMUA SKEMA PEMBAGIAN DATA

No.	Skema Pengujian	Akurasi	Presisi	Recall
1.	70:30	86.00%	83.87%	92.85%
2.	80:20	93.93%	100%	88.88%
3.	90:10	88.23%	87.50%	87.50%

Dari Tabel XII, terlihat bahwa algoritma *decision tree* C5.0 dengan skema pengujian 70:30 menghasilkan akurasi sebesar 86%, sedangkan skema pengujian 80:20 dan 90:10, secara berturut-turut, memiliki akurasi 93.93% dan 88.23%. Hal ini menunjukkan bahwa skema 80:20 memiliki tingkat akurasi tertinggi, di mana 33 sampel data uji terbagi ke dalam 18 orang 'Siap' kerja dan 15 orang 'Tidak Siap' kerja.

Dalam hal presisi, skema pengujian 70:30, 80:20, dan 90:10 masing-masing menghasilkan presisi sebesar

83.87%, 100%, dan 87.50%. Ini berarti pada skema 80:20, model tidak melakukan kesalahan dalam mengidentifikasi individu yang 'Siap' kerja, menunjukkan bahwa model ini sangat andal dalam mengenali positif yang benar.

Selanjutnya, *recall* dari setiap skema turut dihitung. Skema pengujian 70:30 menghasilkan *recall* 92.85%, skema pengujian 80:20 menghasilkan *recall* 88.88%, dan skema pengujian 90:10 menghasilkan *recall* 87.50%. Hasil ini menunjukkan bahwa skema 70:30 memiliki kemampuan terbaik dalam mendeteksi individu yang 'Siap' kerja di antara semua individu yang sebenarnya 'Siap' kerja.

Adanya pembagian skema pengujian, yaitu skema 70:30, 80:20, dan 90:10, adalah untuk memastikan bahwa model yang dibangun dapat diuji dengan proporsi data yang berbeda-beda, memberikan gambaran yang lebih komprehensif tentang kinerja model. Dengan menguji model pada skema data yang berbeda, penulis dapat memastikan bahwa model tidak hanya bekerja baik pada satu proporsi data saja, sehingga mengurangi risiko *overfitting* (terlalu menyesuaikan dengan data latih) dan *underfitting* (tidak cukup menyesuaikan dengan data). Oleh karena itu, penggunaan berbagai skema pengujian memberikan kepercayaan lebih pada validitas dan generalisasi model yang dibangun, memastikan bahwa hasil yang diperoleh dapat diaplikasikan secara lebih luas pada data yang serupa di masa mendatang.

Dengan mempertimbangkan akurasi, presisi, dan *recall* yang diperoleh, algoritma C5.0 terbukti efektif sebagai algoritma klasifikasi yang memberikan hasil akurat. Selain itu, C5.0 juga dapat menangani missing values dan menghasilkan model yang lebih sederhana dan mudah diinterpretasikan. Algoritma C5.0 juga memiliki kemampuan untuk menghasilkan aturan klasifikasi dalam bentuk yang mudah dimengerti, sehingga memudahkan pengguna dalam mengidentifikasi pola dan hubungan dalam data. Tak hanya itu, dapat diputuskan bahwa di dalam skema 80:20, dengan akurasi model tertinggi, sebanyak 7 dari 19 data testing terbukti benar "Siap Kerja" dan 8 dari 17 data testing terbukti benar "Tidak Siap Kerja". Artinya, melalui pengujian ini, lebih banyak mahasiswa tingkat akhir Universitas Negeri Surabaya yang termasuk "Tidak Siap Kerja" dibandingkan dengan "Siap Kerja" berdasarkan atribut pengujian yang ada. Artinya, mahasiswa perlu meningkatkan kompetensi dirinya agar mampu bersaing dalam dunia kerja.

### IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah diuraikan, kesimpulan yang diperoleh dari penelitian berjudul "Klasifikasi Tingkat Kesiapan Kerja Mahasiswa Universitas Negeri Surabaya Menggunakan Algoritma C5.0" adalah:

- Langkah awal dalam penelitian ini melibatkan pengumpulan dan persiapan data mahasiswa serta faktor-faktor yang memengaruhi kesiapan kerja mereka, termasuk penghapusan atribut yang tidak digunakan, penggantian nama atribut, pemeriksaan missing values, penghapusan duplikasi, penjumlahan data, dan penentuan

kolom variabel serta target. Setelah itu, dilakukan seleksi fitur untuk menentukan variabel yang paling relevan dalam memprediksi kesiapan kerja. Dengan mempersempit fokus pada variabel yang signifikan, akurasi prediksi meningkat. Melalui algoritma C5.0, dibangun *decision tree* yang kompleksitasnya sesuai dengan masalah, memberikan wawasan tentang faktor-faktor yang mempengaruhi kesiapan kerja mahasiswa. Evaluasi dan penyetulan model dilakukan secara berkelanjutan untuk memastikan keakuratan dan keandalan prediksi, sementara pengujian menggunakan data uji independen mengonfirmasi keefektifan model. Dengan demikian, penerapan metode klasifikasi data dengan algoritma C5.0 dalam kerangka Data Mining membantu mengklasifikasikan tingkat kesiapan kerja mahasiswa secara akurat dan efektif.

2. Penggunaan algoritma C5.0 pada penelitian ini berhasil mencapai hasil evaluasi yang memuaskan. Dengan skema 70:30, diperoleh akurasi 86%, presisi 83.87%, dan *recall* 92.85%. Di lain sisi, skema 80:20 menghasilkan akurasi 93.93%, presisi 100%, dan *recall* 88.88%. Terakhir, skema 90:10 menghasilkan akurasi 88.23%, presisi 87.50%, dan *recall* 87.50%. Dengan begitu, akurasi dan presisi tertinggi dimiliki oleh skema 80:20 dan *recall* terbaik dimiliki oleh skema 70:30.

#### UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Tuhan Yesus Kristus atas rahmat yang diberikan-Nya sehingga penulis dapat menyelesaikan artikel ini hingga akhir dengan baik. Selanjutnya, penulis turut mengucapkan terima kasih kepada ibu, adik, serta seluruh keluarga yang turut mendoakan dan memberikan semangat. Berikutnya, penulis mengucapkan terima kasih kepada dosen pembimbing yang telah membimbing penulis dengan sabar dan memberi saran kepenulisan yang selalu membangun. Tak lupa, terima kasih penulis ucapkan kepada semua pihak yang telah memberikan doa dan semangat selama penelitian ini berlangsung. Penulis berharap penelitian ini dapat bermanfaat bagi pengembangan ilmu pengetahuan dan dapat menjadi referensi bagi penelitian selanjutnya.

#### REFERENSI

- [1] Masturina, D., "Pengaruh Kompetensi Diri Dan Kepercayaan Diri Terhadap Perencanaan Karir" *PSIKOBORNEO: Jurnal Ilmiah Psikologi*, vol. 6(2), hal. 198-205, 2018.
- [2] (2023) Badan Pusat Statistik (BPS). [Online], <https://www.bps.go.id/id/pressrelease/2023/11/06/2002/tingkat-pengangguran-terbuka--tpt--sebesar-5-32-persen-dan-rata-rata-upah-buruh-sebesar-3-18-juta-rupiah-per-bulan.html>, tanggal akses: 12 November 2023.
- [3] (2024) Redaksi UNESA. [Online], <https://www.unesa.ac.id/unesa-kukuhkan-1380-wisudawan-gubernur-jatim-sebut-unesa-center-of-gravity-bidang-sport-science#:~:text=Unesa.ac.id.%2C1%2F2%2F2024>, tanggal akses: 1 Mei 2024.
- [4] (2023) Redaksi UNESA. [Online], <https://www.unesa.ac.id/unesa-kukuhkan-1501-wisudawan-rector-lulusan-harus-adaptif-dan-punya-kontribusi>, Tanggal akses: 1 Mei 2024.
- [5] Muspawi, M. dan Lestari, A., "Membangun Kesiapan Kerja Calon Tenaga Kerja" *Jurnal Literasiologi*, vol. 4 (1), hal. 112-117, 2020.
- [6] Agusta, Y. "Hubungan Antara Orientasi Masa Depan dan Daya Juang Terhadap Kesiapan Kerja pada Mahasiswa Tingkat Akhir Fakultas Ilmu Sosial dan Ilmu Politik di Universitas Mulawarman" *Jurnal Psikologi*, vol. 3(1), hal. 369-381, 2014.
- [7] Jijo B.T. dan Abdulazeez, A.M., "Classification Based on Decision Tree Algorithm for Machine Learning" *Journal of Applied Science and Technology Trends*, vol. 2(1), 20-28, 2021.
- [8] (2021) National Association of Colleges and Employers (NACE). [Online], <https://www.naceweb.org/career-readiness/competencies/career-readiness-defined>, tanggal akses: 1 Oktober 2023.
- [9] Asri, S.D., Helmi, H., dan Huda, N.M., "Implementasi Algoritma C5.0 pada Klasifikasi Data Sosial Masyarakat (Studi Kasus: Kelayakan Penerimaan BLT di Kelurahan Condong Kota Singkawang)" *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, vol. 12(3), hal. 259-268, 2023.
- [10] Indah Werdiningsih, Barry Nuqoba, & Muhammadun, *Data Mining menggunakan Android, Weka, dan SPSS*. Surabaya: Airlangga University Press, 2020.
- [11] Mustika, et al., *Data Mining dan Aplikasinya*. Bandung: Penerbit Widina, 2021.
- [12] (2021) Redaksi SIS Binus. [Online], <https://sis.binus.ac.id/2021/09/30/proses-data-mining-kdd/>, tanggal akses: 3 November 2023
- [13] Fernandez, A., Garcia, S., Herrera, F., dan Chawla, N.V "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary" *Journal of Artificial Intelligence Research*, vol. 61, hal. 863-905, 2018.
- [14] Siringoringo, R. "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor" *Journal Information System Development (ISD)*, vol. 3(1), hal. 44-49, 2018.
- [15] Ridwan, R., Hermaliani, E.H., dan Ernawati, M. "Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian" *Computer Science (CO-SCIENCE)*, vol. 4(1), 80-88, 2024.
- [16] Bouchlaghe, Y., Akhiat, Y., dan Amjad, S., "Feature Selection: A Review and Comparative Study" *E3S Web of Conferences*, vol. 351 (01046), 2022.
- [17] Liu, S., Yang, S., Xing, A., et al., "Machine learning-based long-term outcome prediction in patients undergoing percutaneous coronary intervention. Cardiovascular Diagnosis and Therapy" *Cardiovascular Diagnostic & Therapy*, vol. 11(3), hal. 736- 743, 2021.
- [18] Karabulut, E.M., Özel, S.A., dan İbrikiç, T., "A comparative study on the effect of feature selection on classification accuracy" *Procedia Technology*, vol. 1, hal. 323-327, 2012.
- [19] Zhou, H.F., Zhang, J.W., Zhou, Y.Q., et al., "A feature selection algorithm of decision tree based on feature weight" *Expert Systems with Applications*, vol. 164, 2021.
- [20] Rahman, A., "Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes" *Jurnal SAINTEKOM (Sains, Teknologi, Komputer, dan Manajemen)*, vol. 13(1), hal. 23-31, 2023.
- [21] Benediktus, N., dan Oetama, R.S., "Algoritma Klasifikasi Decision Tree C5.0 untuk Memprediksi Performa Akademik Siswa" *ULTIMATICS: Jurnal Teknik Informatika*, vol. 12, 2020.
- [22] Pandya, R. dan Pandya, J., "C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning" *International Journal of Computer Applications*, vol. 117 (16), hal. 18-21, 2015.
- [23] Abidin, Z., Nurhana, E., Permata, P., dan Ulum, F., "ANALISIS PERBANDINGAN ALGORITMA DECISION TREE C4.5 DAN C5.0 PADA DATA KARYAWAN BERPOTENSI PROMOSI JABATAN" *JURNAL TEKNOINFO*, vol.17 (2), hal. 567-582, 2023.
- [24] Audina, F.S. dan Putra, R.E., "Klasifikasi Tingkat Kepuasan Wali Murid Terhadap Hasil Belajar Anak Menggunakan Algoritma C5.0" *Journal of Informatics and Computer Science (JINACS)*, vol. 4(4), 2023.
- [25] Cutting, V. dan Stephen N., "A Review on using Python as a Preferred Programming Language for Beginners International" *International Research Journal of Engineering and Technology (IRJET)*, vol. 8(8), hal. 4258-4263, 2021.
- [26] Nisha Arya Ahmed. (2023) Machine Learning page on Datacamp Website [Online], <https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>, tanggal akses: 4 November 2023.

- [27] Jiawei Han & Micheline Kamber, *Data Mining: Concepts and Techniques*, Massachusetts, USA: Morgan Kaufmann Publishers, 2006.