



## Klasifikasi Kesiapan Kerja Lulusan Universitas Dengan Pendekatan *Data Mining* Menggunakan KNN

Nur Achmad Fauzi<sup>1</sup>, Betha Nurina Sari<sup>2</sup>

<sup>1,2</sup>Program Studi Informatika, Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang

---

### Abstract

Received: 02 Januari 2026

Revised: 16 Januari 2026

Accepted: 28 Januari 2026

*Work readiness is an essential indicator for assessing the quality of higher education graduates. This study aims to classify the work readiness of Informatics alumni at the Faculty of Computer Science, Universitas Singaperbangsa Karawang (UNSIKA), using the k-Nearest Neighbors (kNN) algorithm. The dataset was obtained from tracer studies of graduates from 2018 to 2024. The research follows the Knowledge Discovery in Databases (KDD) methodology, which includes five stages: data selection, preprocessing, transformation, data mining, and evaluation. To improve model performance, feature selection was performed using ANOVA, and data normalization was applied using the Standard Scaler. The classification model was tested under eight scenarios, with variations in the number of features, values of k, and training-testing data splits. The best result was achieved without feature selection, using a train-test ratio of 80:20 and k = 21, yielding 60.82% accuracy, 63.43% precision, 86.74% recall, and a 73.28% F1-score. The findings demonstrate that the kNN algorithm is suitable for predicting graduate work readiness, and that proper preprocessing and feature handling significantly affect classification performance.*

**Keywords:** *k-Nearest Neighbors, work readiness, tracer study, classification, machine learning*

(\*) Corresponding Author: [nur.achmad18078@student.unsika.ac.id](mailto:nur.achmad18078@student.unsika.ac.id)

**How to Cite:** Fauzi, N., & Sari, B. (2026). Klasifikasi Kesiapan Kerja Lulusan Universitas Dengan Pendekatan Data Mining Menggunakan KNN. *Jurnal Ilmiah Wahana Pendidikan*, 12(2.C), 207-215. Retrieved from <https://jurnal.peneliti.net/index.php/JIWP/article/view/12487>.

---

## PENDAHULUAN

*Tracer study* merupakan suatu mekanisme yang diselenggarakan oleh institusi pendidikan guna mengumpulkan data terkait alumni setelah menyelesaikan masa pendidikannya. Informasi yang dikumpulkan mencakup berbagai aspek hasil pendidikan yang telah diperoleh selama menempuh perkuliahan. Tujuan utama pelaksanaan *tracer study* adalah untuk meningkatkan mutu pendidikan serta menjadi dasar pengembangan institusi secara berkelanjutan, mengingat perguruan tinggi memiliki tanggung jawab dalam mempersiapkan lulusannya agar mampu berkarier dan beradaptasi di dunia kerja (Aprilliah & Supratman, 2022).

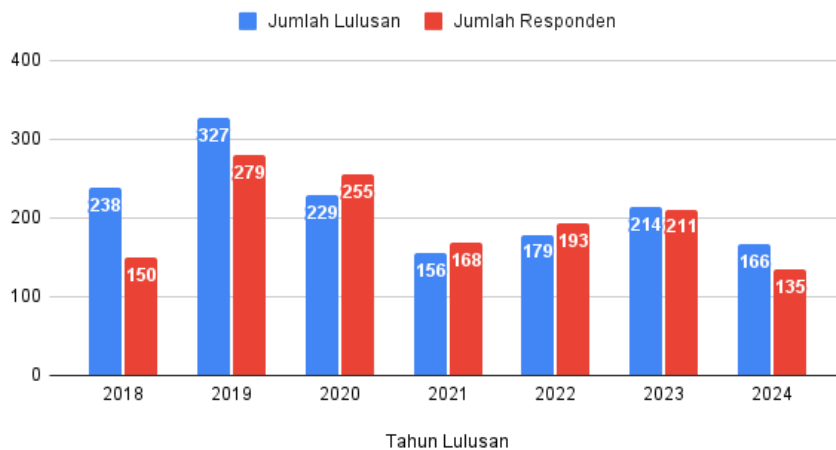
Kesiapan lulusan dalam memasuki dunia kerja menjadi indikator penting dalam menilai sejauh mana suatu institusi pendidikan mampu memberikan kontribusi terhadap pengembangan kemampuan peserta didiknya, baik dari segi kompetensi teknis (*hard skill*) maupun kesiapan mental. Salah satu contohnya adalah kesiapan mahasiswa setelah menyelesaikan pendidikan di perguruan tinggi untuk terjun ke dunia profesional. Data mengenai kesiapan kerja lulusan umumnya diperoleh melalui pelaksanaan *tracer study*, yaitu survei terhadap alumni yang dilakukan secara terstruktur, baik melalui platform

daring seperti Google Form maupun melalui formulir fisik yang dibagikan langsung kepada para lulusan (Ayaturrahman & Rahayu, 2023).

Di Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang (Fasilkom UNSIKA), *tracer study* terhadap alumni telah dilaksanakan sejak tahun 2008 dan terus diperbarui hingga saat penelitian ini dilakukan. Melalui *tracer study* tersebut, fakultas mengumpulkan informasi dari berbagai aspek untuk menilai dan mengklasifikasikan kesiapan kerja alumni, seperti Indeks Prestasi Kumulatif (IPK), kemampuan berdiskusi, penguasaan bahasa Inggris, serta status pekerjaan saat ini.

Namun karena pelaksanaan *tracer study* dilakukan secara *online*, terdapat kemungkinan bahwa tidak semua alumni berpartisipasi dalam pengisian kuesioner, yang dapat memengaruhi representasi data. Gambar 1 menyajikan distribusi jumlah lulusan dan responden *tracer study* dari tahun 2018 hingga 2024 sebagai ilustrasi tingkat partisipasi alumni dalam survei tersebut.

Data Tracer Study FASILKOM UNSIKA (2018 - 2024)



**Gambar 1.** Data *Tracer Study* FASILKOM UNSIKA

Secara umum, jumlah responden dalam *tracer study* seharusnya tidak melebihi jumlah lulusan pada tahun yang bersangkutan. Namun, berdasarkan data yang ditampilkan pada Gambar 1.1, terjadi anomali pada tahun 2022, di mana jumlah responden tercatat sebanyak 193, sedangkan jumlah lulusan hanya 179 orang. Ketidaksiwaan ini disebabkan oleh beberapa alumni yang mengisi formulir lebih dari satu kali, serta adanya kesalahan *input*, seperti kekeliruan dalam mencantumkan tahun kelulusan yang justru diisi dengan tahun masuk kuliah (tahun angkatan).

Pengumpulan data *tracer study* di lingkungan Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang (FASILKOM UNSIKA) telah dimulai sejak tahun 2021 melalui formulir *online* yang disebarluaskan melalui media sosial. Sayangnya, meskipun data telah dikumpulkan, pemanfaatannya masih tergolong minim. Oleh karena itu, penelitian ini bertujuan untuk mengklasifikasikan tingkat kesiapan kerja alumni berdasarkan data *tracer study* FASILKOM UNSIKA, dengan harapan dapat mengidentifikasi faktor-faktor penting yang memengaruhi kesiapan kerja serta memprediksi kesiapan lulusan di masa mendatang. Salah satu pendekatan yang digunakan adalah pemanfaatan *machine learning* dengan metode klasifikasi menggunakan data atribut yang tersedia dalam *tracer study* tersebut.

Beberapa penelitian sebelumnya telah menunjukkan efektivitas pendekatan serupa. Sebagai contoh, penelitian oleh (Rahmayani et al., 2023) yang berjudul “Penerapan Algoritma C4.5 Dengan *Feature Forward Selection* Untuk Analisis Capaian Indikator Kinerja Utama Berdasarkan *Tracer Study* (Studi Kasus: Fasilkom Unsika)”

berhasil menerapkan algoritma C4.5 untuk menganalisis capaian Indikator Kinerja Utama Perguruan Tinggi Negeri (IKU-PTN) berdasarkan data alumni FASILKOM UNSIKA tahun 2020–2021. Penelitian ini menggunakan teknik *feature selection* untuk memilih atribut yang berpengaruh besar seperti “Status Bekerja” dan “Gaji”, dan menghasilkan akurasi 98,77%, *recall* 96%, serta F1-score 89%.

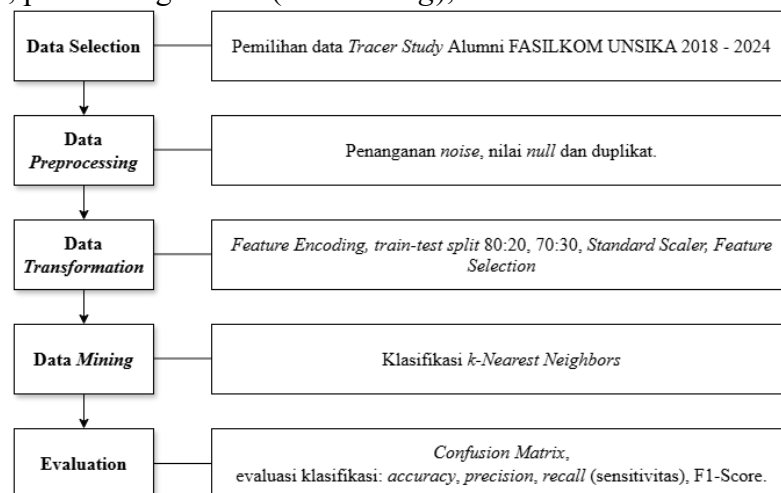
Sementara itu, studi dari (Cholil et al., 2024) berjudul “Prediksi Lama Masa Tunggu Alumni USM dalam Mendapatkan Pekerjaan dengan Algoritma KNN” menggunakan algoritma *k-Nearest Neighbors* (kNN) untuk memprediksi lama waktu tunggu alumni dalam memperoleh pekerjaan. Penelitian tersebut menunjukkan bahwa kNN mampu mencapai akurasi hingga 98,84% saat diuji dengan variasi nilai *k* dari 1 hingga 100.

Berdasarkan kompleksitas faktor-faktor yang memengaruhi kesiapan kerja lulusan, penelitian ini mengadopsi pendekatan *machine learning* untuk mengklasifikasikan kesiapan kerja alumni berdasarkan data tracer study FASILKOM UNSIKA. Metode klasifikasi yang digunakan adalah algoritma *k-Nearest Neighbors* (kNN), yang termasuk ke dalam kelompok supervised learning dan memerlukan label data sebagai acuan pelatihan. Algoritma ini dinilai cocok untuk diterapkan pada dataset dengan ukuran relatif kecil seperti *tracer study* ini, karena tidak memerlukan banyak parameter dan tidak membebani sumber daya komputasi secara berlebihan. Selain itu, kNN juga fleksibel untuk berbagai jenis data, meskipun kinerjanya sangat bergantung pada pemilihan fitur dan metode pembobotan yang tepat (Putra et al., 2024).

Dengan mempertimbangkan berbagai permasalahan yang telah diuraikan serta referensi dari studi-studi terdahulu, penelitian ini akan mengimplementasikan algoritma kNN untuk mengklasifikasikan kesiapan kerja lulusan mahasiswa FASILKOM UNSIKA, sekaligus menerapkan teknik pemilihan fitur guna menangani permasalahan ketidakseimbangan kelas pada *dataset tracer study* tersebut.

## METODOLOGI PENELITIAN

Penelitian ini menerapkan tahapan proses *Knowledge Discovery in Database* (KDD) yang terdiri dari lima tahap utama, yaitu seleksi data, prapemrosesan data, transformasi data, penambangan data (*data mining*), dan evaluasi model.



**Gambar 2.** Rancangan Penelitian KDD

### 1. *Data Selection*

Data yang digunakan dalam penelitian ini berasal dari *dataset tracer study* milik Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang (FASILKOM

UNSIKA), yang mencakup data alumni lulusan tahun 2018 hingga 2024. Pemilihan rentang tahun tersebut dilakukan untuk mengurangi ketidakseimbangan kelas yang berlebihan dalam proses klasifikasi. Selain itu, dilakukan pula eliminasi terhadap atribut-atribut yang tidak relevan dan bersifat sensitif seperti nomor telepon dan Nomor Induk Kependudukan (NIK), guna menjaga privasi responden.

2. *Data Pre-processing*

Tahapan ini mencakup proses pembersihan data dari nilai kosong (*missing values*), nilai ekstrem (*outlier*), dan data duplikat. Pembersihan ini bertujuan untuk meningkatkan konsistensi dan kualitas data, sehingga model yang dikembangkan dapat beroperasi dengan lebih andal dan akurat.

3. *Data Transformation*

Transformasi data dilakukan agar data mentah dapat disesuaikan dengan kebutuhan algoritma pembelajaran mesin. Salah satu langkah utama dalam tahap ini adalah encoding nilai-nilai kategorikal menjadi bentuk numerik. Selain itu, dilakukan juga proses *scaling* pada fitur numerik agar memiliki rentang nilai yang sebanding, serta pembagian *dataset* menjadi data latih dan data uji. Pembagian ini dilakukan di awal untuk menghindari kebocoran data (*data leakage*) yang dapat memengaruhi hasil evaluasi model.

4. *Data Mining*

Pada tahap ini, diterapkan algoritma k-Nearest Neighbors (kNN) terhadap data yang telah melalui proses transformasi. Algoritma kNN dipilih karena sifatnya yang sederhana, efektif, dan cocok diterapkan pada *dataset* dengan ukuran relatif kecil seperti data *tracer study* ini. Proses klasifikasi dilakukan dengan mempertimbangkan nilai k tertentu serta skema pembobotan yang tepat guna meningkatkan akurasi klasifikasi.

5. *Evaluation*

Tahapan evaluasi bertujuan untuk mengukur performa model yang telah dibangun. Evaluasi dilakukan dengan menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*. Langkah ini penting untuk memastikan keandalan model dalam mengenali pola baru serta menilai sejauh mana model mampu mengklasifikasikan kesiapan kerja lulusan secara tepat.

**HASIL DAN PEMBAHASAN**

Penelitian ini menerapkan tahapan proses *Knowledge Discovery in Database* (KDD) yang terdiri dari lima tahap utama, yaitu seleksi data, prapemrosesan data, transformasi data, penambangan data (*data mining*), dan evaluasi model.

1. *Data Selection*

	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Nomor Pokok	Tahun Lulus	Kode Prodi	Nar	No	Alar	I	I	I2 Menurut anda seb	I2 Menurut anda seb	I2 Menurut anda seb	I2 Menurut anda seb	I2 Menurut anda seb	I2 Menurut anda seb
2		2020	55201	Teknik Informatika S1					[2] Besar	[4] Kurang	[2] Besar	[2] Besar	[2] Besar	[4] Kurang
3		2020	55201	Teknik Informatika S1					[3] Cukup Besar	[2] Besar	[2] Besar	[2] Besar	[2] Besar	[2] Besar
4		2020	55201	Teknik Informatika S1					[2] Besar	[2] Besar	[2] Besar	[4] Kurang	[1] Sangat Besar	[2] Besar
5		2020	55201	Teknik Informatika S1					[1] Sangat Besar	[2] Besar	[1] Sangat Besar	[2] Besar	[1] Sangat Besar	[2] Besar
6		2019	55201	Teknik Informatika S1					[3] Cukup Besar	[3] Cukup Besar	[4] Kurang	[3] Cukup Besar	[3] Cukup Besar	[4] Kurang
7		2020	55201	Teknik Informatika S1					[1] Sangat Besar	[3] Cukup Besar	[3] Cukup Besar	[3] Cukup Besar	[2] Besar	[4] Kurang
8		2020	55201	Teknik Informatika S1					[1] Sangat Besar	[2] Besar	[3] Cukup Besar	[3] Cukup Besar	[1] Sangat Besar	[3] Cukup Besar
9		2020	55201	Teknik Informatika S1					[1] Sangat Besar	[2] Besar	[4] Kurang	[4] Kurang	[2] Besar	[4] Kurang
10														
11		2020	55201	Teknik Informatika S1					[3] Cukup Besar	[3] Cukup Besar	[3] Cukup Besar	[3] Cukup Besar	[3] Cukup Besar	[3] Cukup Besar
12		2020	55201	Teknik Informatika S1					[3] Cukup Besar	[2] Besar	[2] Besar	[2] Besar	[3] Cukup Besar	[2] Besar
13		2020	55201	Teknik Informatika S1					[2] Besar	[2] Besar	[4] Kurang	[3] Cukup Besar	[4] Kurang	[4] Kurang
14		2020	55201	Teknik Informatika S1					[1] Sangat Besar	[2] Besar	[3] Cukup Besar	[4] Kurang	[1] Sangat Besar	[5] Tidak Sama Sekali
15		2020	55201	Teknik Informatika S1					[4] Kurang	[4] Kurang	[4] Kurang	[4] Kurang	[3] Cukup Besar	[4] Kurang
16		2020	55201	Teknik Informatika S1					[2] Besar	[2] Besar	[1] Sangat Besar	[2] Besar	[1] Sangat Besar	[1] Sangat Besar
17		2020	55201	Teknik Informatika S1					[1] Sangat Besar	[1] Sangat Besar	[2] Besar	[3] Cukup Besar	[1] Sangat Besar	[2] Besar
18		2020	55201	Teknik Informatika S1					[1] Sangat Besar	[3] Cukup Besar	[3] Cukup Besar	[3] Cukup Besar	[3] Cukup Besar	[2] Besar
19		2020	55201	Teknik Informatika S1					[3] Cukup Besar	[4] Kurang	[4] Kurang	[5] Tidak Sama Sekali	[4] Kurang	[5] Tidak Sama Sekali
20		2020	55201	Teknik Informatika S1					[3] Cukup Besar	[4] Kurang	[4] Kurang	[4] Kurang	[3] Cukup Besar	[4] Kurang
21		2019	55201	Teknik Informatika S1					[1] Sangat Besar	[1] Sangat Besar	[1] Sangat Besar	[1] Sangat Besar	[1] Sangat Besar	[1] Sangat Besar

Gambar 3. Data Mentah *Tracer Study* Fasilkom UNSIKA

Pada Gambar 3 ditampilkan cuplikan data *tracer study* alumni Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang (FASILKOM UNSIKA). *Dataset* ini memuat informasi dari total 1.461 responden, dengan sekitar 120 entri yang tidak ditampilkan secara lengkap karena keterbatasan ruang dan panjangnya atribut yang dimiliki. Proses awal yang dilakukan dalam penelitian ini adalah melakukan seleksi data dengan memfokuskan pada lulusan tahun 2018 hingga 2024.

Tahapan selanjutnya adalah pemilihan atribut secara manual dan menghapus atribut duplikat, kosong, bersifat pribadi (seperti nama, nomor telepon, dan NIK), serta atribut yang tidak relevan dengan tujuan penelitian, seperti data tempat kerja dan gaji. Selanjutnya, dilakukan penyesuaian nama atribut agar lebih ringkas dan mudah dibaca.

Atribut Pilihan			
Sumber dana pendidikan	Pengetahuan di luar bidang atau disiplin ilmu	Manajemen waktu	Bekerja dengan orang yang berbeda budaya maupun latar belakang
Perkuliah	Pengetahuan umum	Bekerja secara mandiri	Kepemimpinan
Demonstrasi	Bahasa Inggris	Bekerja dalam tim	Kemampuan dalam memegang tanggung jawab
Partisipasi dalam proyek riset	Keterampilan internet	Kemampuan dalam memecahkan masalah	Inisiatif
Magang	Penggunaan Teknologi Informasi	Negosiasi	Manajemen proyek
Praktikum	Berpikir kritis	Kemampuan analisis	Kemampuan untuk mempresentasikan ide/produk/laporan
Kerja Lapangan	Ketrampilan riset	Toleransi	Kemampuan dalam menulis laporan, memo dan dokumen
Diskusi	Kemampuan belajar	Kemampuan adaptasi	Kemampuan untuk terus belajar sepanjang hayat
Keahlian berdasarkan bidang ilmu	Kemampuan berkomunikasi	Loyalitas	Etika
Bekerja di bawah tekanan	Integritas	Pengembangan Diri	Status Kerja

**Gambar 4.** Data Mentah *Tracer Study* Fasilkom UNSIKA

Gambar 4 menunjukkan hasil seleksi atribut, di mana diperoleh 39 atribut potensial yang memengaruhi kesiapan kerja, satu atribut “Nomor Pokok Mahasiswa” sebagai *identifier*, dan *Status Kerja* sebagai label untuk klasifikasi.

Jenis Nilai Atribut	Nilai Atribut
Beasiswa	[1] Biaya Sendiri / Keluarga, [2] Beasiswa
Lainnya	[1] Sangat Besar, [2] Besar, [3] Cukup Besar, [4] Kurang, [5] Tidak Sama Sekali
	[1] Sangat rendah, [2] Rendah, [3] Cukup, [4] Tinggi, [5] Sangat tinggi
Status Kerja	[1] Bekerja ( <i>full time/part time</i> ), [2] Belum memungkinkan bekerja, [3] Wiraswasta (termasuk <i>freelance</i> /kerja lepas), [3] Wiraswasta, [4] Melanjutkan Pendidikan, [5] Tidak Kerja tetapi sedang mencari kerja

**Gambar 5.** Keterangan Jenis Nilai Atribut

Gambar 5 menampilkan nilai-nilai pada masing-masing atribut dalam dataset. Mengingat tujuan penelitian ini adalah melakukan klasifikasi kesiapan kerja lulusan, maka atribut *Status Kerja* dipilih sebagai label. Nilai pada atribut tersebut kemudian dikonversi ke dalam dua kategori klasifikasi, yaitu “siap” dan “belum siap”, dengan ketentuan klasifikasi terlihat pada Gambar 6.

Status Kerja	Kesiapan Kerja
[1] Bekerja (full time/part time), [3] Wiraswasta (termasuk freelance/kerja lepas), [3] Wiraswasta, [4] Melanjutkan Pendidikan	Siap
[2] Belum memungkinkan bekerja, [5] Tidak Kerja tetapi sedang mencari kerja	Belum Siap

**Gambar 6.** Keterangan Jenis Nilai Atribut

2. *Data Pre-processing*

Setelah tahap seleksi data, proses selanjutnya adalah melakukan prapemrosesan data, yang mencakup penanganan nilai kosong serta duplikat. Penghapusan data duplikat dilakukan dengan menggunakan atribut Nomor Pokok Mahasiswa sebagai identifikasi unik. Dari proses ini ditemukan sebanyak 66 data duplikat yang kemudian dihapus. Selanjutnya, atribut Nomor Pokok Mahasiswa juga dieliminasi karena tidak memiliki kontribusi terhadap analisis kesiapan kerja. Setelah tahap pembersihan selesai, distribusi jumlah lulusan yang termasuk dalam kategori “siap” dan “belum siap” ditampilkan pada Gambar 7.

Kesiapan Kerja	Jumlah	Persentase
Siap	829	~62%
Belum Siap	507	~38%

**Gambar 7.** Jumlah Alumni Siap Bekerja

3. *Data Transformation*

Setelah proses pembersihan data selesai, tahap berikutnya adalah transformasi data. Langkah awal pada tahap ini adalah melakukan *feature encoding*, yaitu mengubah nilai-nilai kategorikal pada setiap atribut menjadi nilai numerik. Transformasi ini diperlukan agar data sesuai dengan format *input* yang dibutuhkan oleh algoritma pembelajaran mesin. Setelah seluruh atribut berhasil dikonversi ke dalam bentuk numerik, data kemudian dibagi ke dalam dua jenis, yaitu data latih dan data uji, yang masing-masing digunakan untuk pelatihan model klasifikasi dan evaluasi kinerjanya. Data dibagi dalam dua skenario: skenario pertama dengan rasio 70% data latih dan 30% data uji, serta skenario kedua dengan rasio 80% data latih dan 20% data uji.

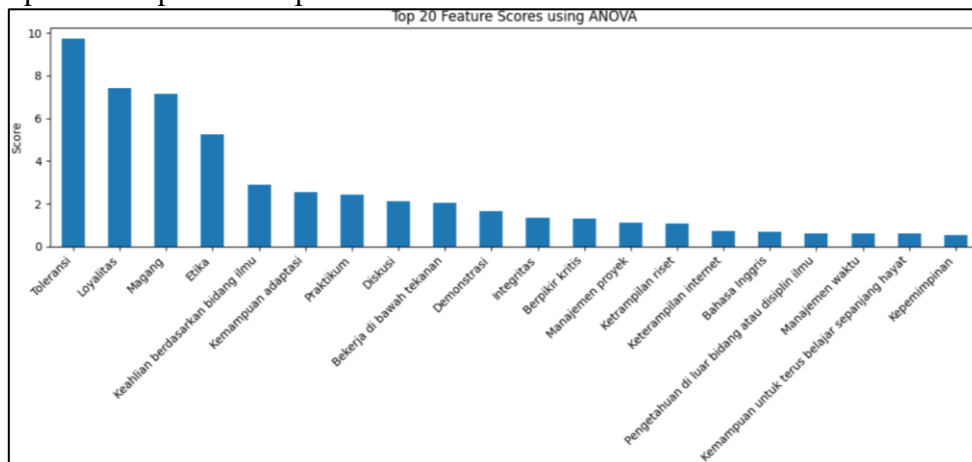
Sumber dana pendidikan	Perkuliahan	Demonstrasi	Partisipasi dalam proyek riset	Magang	Praktikum	Kerja Lapangan	Diskusi	Keahlian berdasarkan bidang ilmu	Pengetahuan di luar bidang atau disiplin ilmu
-0.355677	-1.006778	-0.341358	-0.267308	-0.422202	-0.750961	-0.277141	-0.778853	-0.467742	0.923085
-0.355677	1.435083	0.819260	1.789754	1.494103	1.376761	1.649452	1.522024	-0.467742	2.373648
-0.355677	-1.006778	-1.501976	-2.324371	-2.338508	-2.878682	-2.203733	-3.079730	-3.354466	-3.428603
-0.355677	1.435083	0.819260	-1.295839	1.494103	0.312900	-0.277141	0.371585	2.418982	0.923085
-0.355677	0.214153	0.819260	-1.295839	-0.422202	-1.814822	-1.240437	-0.778853	-0.467742	-0.527477

**Gambar 8.** Data Sesudah *Standard Scaler*

Terlihat pada gambar 8 tahap selanjutnya adalah melakukan skala data numerik. Teknik yang digunakan adalah *Standard Scaler*, yaitu dengan mengurangi setiap nilai fitur numerik dengan rata-rata (*mean*) dan membaginya dengan standar deviasi berdasarkan data latih.

Penelitian ini juga bertujuan mengidentifikasi fitur yang paling berpengaruh terhadap kesiapan kerja. Dari 39 atribut kandidat, dilakukan seleksi fitur menggunakan metode ANOVA karena mampu mengevaluasi fitur secara independen berdasarkan urutan signifikan terhadap target label dan memiliki hasil yang konsisten (tidak berubah-

ubah). Delapan skenario pengujian digunakan, terdiri dari dua rasio data (80:20 dan 70:30) serta empat variasi fitur: 10, 15, 20, dan tanpa seleksi fitur. Untuk 20 fitur terbaik yang diperoleh dapat dilihat pada Gambar 9.



**Gambar 9.** Seleksi 20 Fitur Terbaik ANOVA

#### 4. Data Mining

Tahap selanjutnya adalah pembangunan model untuk keperluan pengujian. Penelitian ini menggunakan algoritma *k-Nearest Neighbors* (kNN) untuk melakukan klasifikasi biner kesiapan kerja lulusan. Karena pembagian data latih dan data uji telah dilakukan pada tahap transformasi data, maka proses *data mining* dimulai dengan pembuatan model dan penyetelan *hyperparameter*.

```
k_values = range(1, 25)
for k in k_values:
    knn2 = KNeighborsClassifier(n_neighbors=k, weights='distance', metric='euclidean')
    knn2.fit(X2_train, y2_train)
    y2_pred = knn2.predict(X2_test)
```

**Gambar 10.** *Hyperparameter* kNN

Seperti ditunjukkan pada Gambar 10 model kNN dibuat dan dievaluasi secara berulang untuk 12 nilai *k* (ganjil dari 3 hingga 25). Parameter lainnya diatur menggunakan skema pembobotan *distance* dan metrik jarak *Euclidean*. Model diuji dalam delapan skenario yang menggabungkan variasi jumlah fitur dan rasio pembagian data.

#### 5. Evaluation

Langkah berikutnya adalah melakukan evaluasi model dengan menyusun *confusion matrix* dari setiap skenario yang menunjukkan nilai F1-Score terbaik pada data uji. Hasil evaluasi dirangkum dalam bentuk visual agar memudahkan analisis. Gambar 11 berikut menyajikan kumpulan *confusion matrix* dari delapan skenario terbaik, yang memuat nilai *true positive*, *true negative*, *false positive*, dan *false negative* untuk masing-masing model.

No	Pembagian Data (Latih : Uji)	Atribut	K	TP	TN	FP	FN	F1-Score	Accuracy
1	80:20	10	25	122	27	75	44	67.22%	55.60%
2	80:20	15	23	132	30	72	34	71.35%	60.45%
3	80:20	20	21	134	27	75	32	71.47%	60.07%
4	80:20	39	21	144	19	83	22	73.28%	60.82%
5	70:30	10	21	186	46	106	63	68.76%	57.86%
6	70:30	15	25	207	32	120	42	71.88%	59.60%
7	70:30	20	25	212	23	129	37	71.86%	58.60%
8	70:30	39	19	198	42	110	51	71.10%	59.85%

**Gambar 11.** Hasil *Confusion Matrix* 8 Skenario F1-Score Tertinggi

## KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, diperoleh beberapa kesimpulan sebagai berikut:

1. Proses klasifikasi kesiapan kerja dilakukan menggunakan data yang telah melalui tahapan data selection, preprocessing, dan transformation, kemudian diuji melalui delapan skenario pengujian. Skenario tersebut terdiri dari dua rasio pembagian data latih dan uji, yaitu 80:20 dan 70:30 dengan teknik stratified sampling, serta empat variasi jumlah fitur menggunakan metode seleksi ANOVA: tanpa seleksi, 10 fitur, 15 fitur, dan 20 fitur. Algoritma klasifikasi yang digunakan adalah k-Nearest Neighbors (kNN) dengan nilai k ganjil dari 3 hingga 25.
2. Pada skenario tanpa seleksi fitur dengan rasio pembagian data 80:20 dan nilai  $k = 21$ , model menghasilkan akurasi sebesar 60,82%, precision 63,43%, recall 86,74%, dan F1-score sebesar 73,28%. Sementara itu, nilai F1-score terbaik diperoleh pada skenario dengan pembagian data 70:30, seleksi fitur ANOVA sebanyak 15 atribut, dan nilai  $k = 25$ , yaitu dengan akurasi 59,60%, precision 63,30%, recall 83,13%, dan F1-score sebesar 71,88%.
3. Berdasarkan hasil seleksi fitur menggunakan metode ANOVA pada skenario terbaik, diperoleh 15 atribut dengan pengaruh paling signifikan terhadap kesiapan kerja lulusan. Atribut-atribut tersebut, berdasarkan urutan pengaruh tertinggi, adalah: toleransi, loyalitas, pengalaman magang, etika, keahlian berdasarkan bidang ilmu, kemampuan adaptasi, praktikum, diskusi, kemampuan bekerja di bawah tekanan, demonstrasi, integritas, kemampuan berpikir kritis, manajemen proyek, keterampilan riset, dan keterampilan internet.

## DAFTAR PUSTAKA

- Aprilliah, Riska, Edi Supratman, Universitas Bina Darma Palembang Jl Jenderal Ahmad Yani No, Kecamatan Seberang Ulu, Kota Palembang, and Sumatera Selatan. 2020. *SISTEM INFORMASI PUSAT KARIR DAN TRACER STUDY PADA UNIVERSITAS BINA DARMA BERBASIS ANDROID*. Vol. 4.
- Dafa Ayaturrahman, Jaisya, and Isti Rahayu. 2023. "Dampak Soft Skill Terhadap Kesiapan Kerja Mahasiswa Di Era Industri 4.0." 5:169–75. doi: 10.20885/ncaf.vol5.art19.
- Géron, Aurélien. 2019. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc.
- Gunawan, Hanes, and Catherine. 2021. "C4.5, K-Nearest Neighbor, Naïve Bayes and Random Forest Algorithms Comparison to Predict Students' On Time

- Graduation.” *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIMD)* 4(2):62–71. doi: [10.24014/ijaidm.v4i2.10833](https://doi.org/10.24014/ijaidm.v4i2.10833).
- Rahmayani, Fenny, Betha Nurina Sari, Iqbal Maulana, Rini Mayasari, and Universitas HS Singaperbangsa Karawang Jl Ronggo Waluyo. 2023. *PENERAPAN ALGORITMA C4.5 DENGAN FEATURE FORWARD SELECTION UNTUK ANALISIS CAPAIAN INDIKATOR KINERJA UTAMA BERDASARKAN TRACER STUDY (STUDI KASUS: FASILKOM UNSIKA)*. Vol. 7.
- Rohman Cholil, Saifur, and Vensy Vydia. 2024. *IJCIT (Indonesian Journal on Computer and Information Technology) Prediksi Lama Masa Tunggu Alumni USM Dalam Mendapatkan Pekerjaan Dengan Algoritma KNN*. Vol. 9.
- Palacios, Carlos A., José A. Reyes-Suárez, Lorena A. Bearzotti, Víctor Leiva, and Carolina Marchant. 2021. “Knowledge Discovery for Higher Education Student Retention Based on Data Mining: Machine Learning Algorithms and Case Study in Chile.” *Entropy* 23(4). doi: [10.3390/e23040485](https://doi.org/10.3390/e23040485).
- Putra, Febrianda, Hafsah Fulaila Tahiyat, Raja Muhammad Ihsan, Rahmaddeni Rahmaddeni, and Lusiana Efrizoni. 2024. “Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing Untuk Penentuan Keterangan Berat Badan Manusia.” *MALCOM: Indonesian Journal of Machine Learning and Computer Science* 4(1):273–81. doi: [10.57152/malcom.v4i1.1085](https://doi.org/10.57152/malcom.v4i1.1085).
- Sun, L., Zhang, S., Liu, W., & Hu, Y. (2021). Feature selection: A data perspective. *ACM Computing Surveys*, 54(10s), 1–38. <https://doi.org/10.1145/3447548>.