Terbit online pada laman web jurnal: **https://jurnal.plb.ac.id/index.php/tematik/index**

TEMATIK

Jurnal **Teknologi Informasi Komunikasi (e-Journal)**

Vol. 12 No. 1 (2025) 109 - 117 ISSN Media Elektronik: 2443-3640

Perbandingan Klasifikasi Tipe Kesuksesan Generasi Z Menggunakan Algoritma Naïve Bayes dan Decision Tree

Comparison of Z Generation Success Type Classification Using Naïve Bayes and Decision Tree Algorithms

Novara Aulist Zakia[1*], M. Bucci Ryando[2], Halim Agung[3]
[1,2]Teknik Informatika, Fakultas Teknologi Informasi dan Komunikasi, Institut Teknologi dan Bisnis Bina Sarana Global
[3]Bisnis Digital, Fakultas Bisnis dan Manajemen, Institut Teknologi dan Bisnis Bina Sarana Global
[1]1121130066@global.ac.id, [2]bucci@global.ac.id, [3]halim@global.ac.id

Abstract

This study aims to classify the types of success of Generation Z using the CRISP-DM method approach and using the Naïve Bayes and Decision Tree algorithms. Generation Z who grew up in a digital environment has a unique view of the meaning of success, which is no longer limited to income or position, but also includes life balance and self-development. This study identifies several important factors such as educational background, technological skills, work experience, personal branding, and use of social media as determining variables in the classification of types of success. The classification model produces four main categories of success, namely financial, career, self-development, and life balance. The results showed that life balance was the most dominant category of success among respondents. The use of the Naïve Bayes and Decision Tree algorithms showed that Decision Tree with balancing techniques (random oversampling) provided the highest classification accuracy, which was 94%, compared to Naïve Bayes which only reached 37%. This study makes an important contribution to the development of human resource strategies, education, and policies that are relevant to the characteristics and aspirations of Generation Z in the digital era.

Keywords: Z Generation, classification of success, CRISP-DM, naïve bayes, decision tree

Abstrak

Penelitian ini bertujuan untuk mengklasifikasikan tipe kesuksesan Generasi Z dengan pendekatan metode CRISP-DM serta menggunakan algoritma Naïve Bayes dan Decision Tree. Generasi Z, yang tumbuh dalam lingkungan digital, memiliki pandangan yang unik terhadap makna kesuksesan, yang tidak lagi terbatas pada penghasilan atau jabatan, tetapi juga mencakup keseimbangan hidup dan pengembangan diri. Penelitian ini mengidentifikasi beberapa faktor penting seperti latar belakang pendidikan, keterampilan teknologi, pengalaman kerja, *personal branding*, dan penggunaan media sosial sebagai variabel penentu dalam klasifikasi tipe kesuksesan. Model klasifikasi menghasilkan empat kategori utama kesuksesan: finansial, karier, pengembangan diri, dan keseimbangan hidup. Hasil menunjukkan bahwa keseimbangan hidup merupakan kategori kesuksesan yang paling dominan di antara responden. Penggunaan algoritma Naïve Bayes dan Decision Tree menunjukkan bahwa Decision Tree dengan teknik *balancing* (*random oversampling*) memberikan akurasi klasifikasi tertinggi, yakni sebesar 94%, dibandingkan Naïve Bayes yang hanya mencapai 37%. Penelitian ini memberikan kontribusi penting bagi pengembangan strategi sumber daya manusia, pendidikan, dan kebijakan yang relevan terhadap karakteristik dan aspirasi Generasi Z di era digital.

Kata kunci: Generasi Z, tipe kesuksesan, CRISP-DM, naïve bayes, decision tree

1. Pendahuluan

Generasi Z atau yang biasa dikenal dengan sebutan generasi digital merupakan sekelompok orang yang terbiasa dengan perkembangan teknologi dan lahir pada rentang tahun 1997 hingga 2010. Generasi ini terbukti mampu beradaptasi di dunia digital dan memiliki karakteristik yang khas dalam menghadapi tantangan ekonomi dan sosial. Berdasarkan informasi terkini, Generasi Z saat ini tengah memasuki fase produktif dan tengah bersiap untuk memberikan dampak yang lebih besar di pasar kerja [1]. Namun, kesuksesan generasi ini seringkali diukur secara sempit, terutama berdasarkan tingkat penghasilan [2]. Generasi Z kini mulai mendominasi dunia kerja. Generasi ini memiliki karakteristik unik dibandingkan generasi sebelumnya, seperti preferensi terhadap fleksibilitas kerja, keseimbangan kehidupan kerja, serta penggunaan teknologi dalam keseharian mereka [3].

Diterima Redaksi: 15-05-2025 | Selesai Revisi: 09-06-2025 | Diterbitkan Online: 30-06-2025

Kesuksesan karir generasi Z tidak secara langsung dipengaruhi oleh latar belakang sosial ekonomi (SES) dan disimpulkan bahwa kesuksesan generasi Z lebih dipengaruhi oleh dukungan sosial, pendidikan, dan kebijakan institusional dibandingkan faktor ekonomi [4]. Di Indonesia, penelitian tentang kesuksesan generasi Z masih terfokus pada aspek pendidikan dan keterampilan teknis. Misalnya, penelitian menemukan bahwa tingkat pendidikan memiliki korelasi positif dengan penghasilan generasi Z [5]. Penelitian lainnya menunjukkan bahwa pelatihan kerja, tingkat pendidikan, dan pengalaman kerja berpengaruh positif terhadap kinerja karyawan Generasi Z. Pendidikan memberikan dasar pengetahuan dan keterampilan, sementara pengalaman kerja memperkuat kemampuan dalam praktik [6].

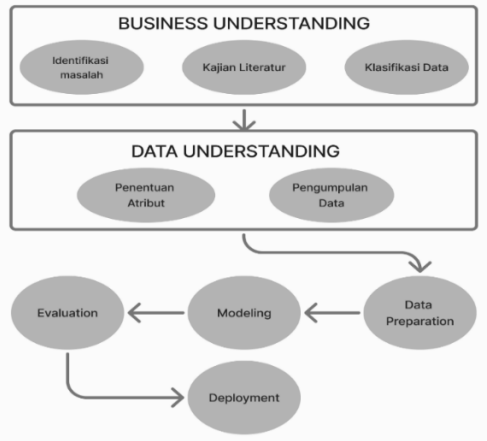
Beberapa penelitian internasional telah mencoba mengklasifikasikan kesuksesan berdasarkan tingkat penghasilan. Misalnya, penelitian oleh menggunakan algoritma Naïve Bayes dan Decision Tree untuk mengidentifikasi faktor-faktor yang memengaruhi penghasilan di kalangan milenial [7]. Hasil dari penelitian ini menunjukkan bahwa pendidikan, pengalaman kerja, dan keterampilan teknis memiliki dampak signifikan terhadap potensi penghasilan mereka. Namun, penelitian ini tidak memperhitungkan karakteristik unik generasi Z, seperti ketergantungan pada teknologi dan preferensi terhadap pekerjaan fleksibel. Di sisi lain, penelitian oleh [8] menemukan bahwa literasi digital memiliki pengaruh signifikan terhadap penghasilan generasi Z di Korea Selatan. Namun, temuan ini belum diuji dalam konteks negara berkembang seperti Indonesia, di mana akses terhadap teknologi dan infrastruktur digital masih terbatas.

Penelitian ini akan berkontribusi dalam pengembangan model klasifikasi berbasis data yang dapat membantu perusahaan atau lembaga pendidikan dalam merancang strategi peningkatan keterampilan tenaga kerja muda. Selain itu, hasil penelitian ini juga dapat digunakan oleh individu generasi Z untuk memahami faktor apa saja yang dapat meningkatkan peluang mereka dalam mencapai kesuksesan finansial di era digital [9].

Penelitian ini bertujuan untuk mengembangkan model klasifikasi yang dapat mengidentifikasi tipe kesuksesan generasi Z berdasarkan data yang dikumpulkan dari berbagai variabel terkait pekerjaan dan penghasilan. Dengan pendekatan data mining, penelitian ini akan menghasilkan model yang dapat digunakan untuk memprediksi faktor-faktor utama yang berkontribusi terhadap kesuksesan finansial generasi ini.

2. Metode Penelitian

Metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dirancang untuk memberikan penjelasan terperinci tentang berbagai tahap pengumpulan data [10]. Fase-fase ini dibagi menjadi enam fase: *business understanding, data understanding, data preparation, metodeling, evaluation, dan deployment* [11][12].



Gambar 1. Diagram CRISP-DM

Diagram pada gambar 1 menggambarkan langkah-langkah proses kerja data mining menggunakan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Diagram ini mengilustrasikan alur logis yang terstruktur dalam menghasilkan solusi, pengetahuan, dan pemahaman mendalam dari data.

Pada tahap awal *business understanding* adalah memahami tujuan penelitian dari sudut pandang bisnis, yaitu mengklasifikasikan faktor -faktor kunci seperti, pengetahuan teknologi, latar belakang pendidikan, dan keterampilan berjejaring yang akan diolah kedalam berbagai tipe kesuksesan [13]. Bagian ini terdiri dari tiga bagian utama, yaitu mengidentifikasi masalah, kajian literatur, dan klarifikasi data.

Pada tahap *data understanding* dimulai dengan pengumpulan data awal, kemudian mengidentifikasi dan mengeksplorasi data. Dilanjutkan dengan menetapkan atribut untuk dipertimbangkan dalam analisis agar data mencakup informasi yang relevan.

Tahap ketiga yaitu *data preparation*, yang dimana melibatkan kegiatan mengumpulkan data mentah sehingga dapat digunakan untuk pemodelan dengan cara menyortir, membersihkan, mengorganisasikan, dan mengintegrasikan data [14]. Pada tahap ini pula dilakukan seleksi terhadap kriteria responden yang digunakan dalam penelitian. Adapun kriteria responden yang ditetapkan yaitu individu yang termasuk dalam kelompok generasi Z yang lahir antara tahun 1997 hingga 2010. Selain itu, responden juga harus sudah memiliki pengalaman kerja atau pernah menjalani aktivitas ekonomi, baik sebagai pekerja tetap, freelancer, wirausaha, maupun pegawai kontrak. Responden diharuskan memiliki latar belakang pendidikan minimal SMA, serta aktif menggunakan teknologi digital seperti smartphone, laptop, atau media sosial dalam kegiatan sehari-hari. Kriteria ini ditetapkan

DOI: https://doi.org/10.38204/tematik.v12i1.2334
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

agar data yang diperoleh relevan dan representatif terhadap populasi Generasi Z yang menjadi fokus utama dalam klasifikasi tipe kesuksesan. Selama tahap pemodelan, berbagai teknik dan algoritma diterapkan tergantung pada jenis data yang dipilih. Tahap ini tidak terdiri dari bagian-bagian yang lebih kecil, melainkan melibatkan berbagai kegiatan seperti membersihkan data, mengubah data, dan memilih atribut. Tujuannya untuk menyediakan data untuk persiapan pemodelan dengan memastikan data terbebas dari kesalahan.

Pada tahap *modelling* data yang telah disiapkan akan digunakan untuk membuat model klasifikasi dengan menggunakan algoritma Naïve Bayes dan Decision Tree yang bertujuan untuk mengembangkan suatu model yang dapat melakukan prediksi atau klasifikasi berdasarkan data yang disediakan, untuk mencapai target [15].

Tahap *evaluation* bertujuan untuk menilai performa model apakah telah mencapai tujuan proyek yang ditetapkan [16]. Dalam penelitian ini digunakan algoritma Naïve Bayes dan Decision Tree untuk memodelkan jenis-jenis keberhasilan Gen Z selama fase pemodelan. Pada tahap evaluasi dilakukan untuk mengukur kinerja model menggunakan phyton, yang dimana data yang nantinya digunakan memiliki atribut lebih dari tiga.

2.1. Algoritma Naïve Bayes

Algoritma Naïve Bayes dipakai dalam penelitian ini bekerja dengan menghitung probabilitas setiap variabel dan menggunakan informasi tersebut untuk menentukan jenis keberhasilan yang paling mungkin terjadi. Naïve Bayes cocok digunakan karena dapat menangani data yang mengandung banyak variabel yang berinteraksi dan memberikan hasil prediksi yang cepat dan akurat [17].

Untuk klasifikasi tipe kesuksesan pada Generasi Z, kita ingin menghitung probabilitas bahwa suatu data X termasuk dalam suatu kelas tertentu C (tipe kesuksesan).

X adalah data masukan, yang terdiri dari beberapa fitur atau atribut, seperti latar belakang pendidikan, pengalaman kerja, keterampilan teknologi, dll.

C adalah kelas atau tipe kesuksesan yang ingin diprediksi, seperti: Finansial, Karir, dll.

Naïve Bayes menghitung probabilitas dari kelas C yang paling mungkin, yang dituliskan sebagai berikut.

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} \quad (1)$$

Keterangan:

P(C|X): Probabilitas kelas C diberikan data X. Ini adalah probabilitas yang ingin kita hitung, seberapa besar kemungkinan C merupakan kelas dari data X.

P(X|C): Probabilitas mendapatkan data X jika kita mengetahui kelas C. Ini adalah informasi yang dihitung berdasarkan distribusi data dalam kelas tertentu.

P(C): Probabilitas awal kelas C. Ini merupakan distribusi awal kelas sebelum melihat data X.P(X): Probabilitas data X di semua kelas. Ini berfungsi sebagai faktor normalisasi agar probabilitas total sama dengan 1.

2.2. Algoritma Decision Tree

Algoritma Decision Tree adalah metode klasifikasi yang membagi data ke dalam subset berdasarkan nilai-nilai atribut dan menghasilkan model berupa struktur pohon keputusan. Setiap simpul dalam pohon merepresentasikan pengujian terhadap atribut, dan cabangnya menunjukkan hasil pengujian, sementara simpul daun menggambarkan kelas akhir atau keputusan [18]. Keunggulan dari algoritma ini terletak pada interpretasi yang mudah dipahami dan penerapannya yang luas, baik untuk data kategorikal maupun numerik. Decision Tree sangat berguna dalam penelitian sosial-ekonomi untuk mengevaluasi faktor-faktor yang memengaruhi hasil tertentu, seperti penghasilan atau keberhasilan karier. Dalam studi terkait generasi milenial dan Z, algoritma ini telah digunakan untuk mengidentifikasi pengaruh pendidikan, keterampilan, dan pengalaman kerja terhadap penghasilan individu secara akurat dan efisien.Jelaskan metode penelitian dan teknik penelitian yang digunakan.

Pada tahap *deployment*, fokusnya adalah pada penerapan model klasifikasi yang telah dibangun ke dalam konteks nyata. Dalam penelitian ini, model Naïve Bayes dan Decision Tree yang telah dievaluasi akan diimplementasikan.

3. Hasil dan Pembahasan

3.1. Klasifikasi Data

Pada Tabel 1, ada 12 atribut yang akan digunakan pada penelitian ini yang dimana terdiri dari Usia, Jenis Kelamin, Pendidikan Terakhir, Keterampilan Digital, Pengalaman Kerja, Jenis Pekerjaan, Model Kerja, Media Sosial, *Personal Branding*, Pelatihan Keterampilan, Penghasilan Utama, Tipe Kesuksesan. Dari tipe atribut yang ada disimpulkan beberapa pertanyaan yang akan menjadi faktor penting dalam klasifikasi kesuksesan Gen Z. Kemudian, data diubah dan disederhanakan dengan mengonversi menjadi kategori dan tingkatan yang ditentukan. Deskripsi dan aturan transformasi data yang dijalankan tercantum pada Tabel 2.

Pada Tabel 2, pengumpulkan informasi/data yang relevan dari sumber seperti kuesioner atau basis data untuk mendukung analisis. Informasi-informasi yang didapatkan melalui kuesioner akan mencakup beragam

DOI: https://doi.org/10.38204/tematik.v12i1.2334
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

data. Transformasi data yang tercantum dalam Tabel 3, mengindikasikan bahwa ada 13 atribut yang dipakai untuk menganalisis berbagai variable. Setiap atribut diubah menjadi nilai numerik diskrit untuk merepresentasikan kategori-kategori tertentu dalam variabel tersebut. Ini memungkinkan visualisasi pola atau keterkaitan antara variabel menjadi lebih sederhana, serta memungkinkan pengolahan data yang lebih efisien dalam analisis yang lebih mendalam.

Tabel 1. Deskripsi Atribut dan Tipe Data

No	Nama Atribut	Deskripsi	Tipe Data
1	Usia	Rentang usia responden Generasi Z (misal: 15-27 tahun)	Numerik
2	Jenis Kelamin	Kategori jenis kelamin responden (Laki-laki/Perempuan)	Nominal
3	Pendidikan Terakhir	Tingkat pendidikan formal tertinggi yang telah dicapai (SMA, Diploma, S1, dan S2.)	Ordinal
4	Keterampilan Digital	Penguasaan keterampilan digital (misal: Sangat tidak mahir, Tidak mahir, Cukup mahir, Mahir, Sangat mahir)	Ordinal
5	Pengalaman Kerja	Jumlah tahun pengalaman kerja dalam bidang yang relevan	Ordinal
6	Jenis Pekerjaan	Kategori pekerjaan utama (Pegawai Tetap, Freelancer, Wirausaha, dll.)	Ordinal
7	Model Kerja	Jenis model kerja (Konvensional, Remote, Hybrid)	Ordinal
8	Penggunaan Media Sosial untuk Karir	Apakah media sosial digunakan untuk meningkatkan peluang kerja? (Ya/Tidak)	Nominal
9	*Personal Branding*	Tingkat kesadaran individu dalam membangun *personal branding* (Rendah, Sedang, Tinggi)	Ordinal
10	Partisipasi dalam Pelatihan/Kursus	Apakah pernah mengikuti pelatihan keterampilan tambahan? (Ya/Tidak)	Nominal
11	Sumber Pendapatan Utama	Apakah penghasilan utama berasal dari pekerjaan tetap, freelance, atau bisnis?	Ordinal
12	Tipe Kesuksesan	Klasifikasi kesuksesan berdasarkan tingkat penghasilan dan karir (Finansial, Karir, Pengembangan Diri dan Keseimbangan Hidup)	Ordinal

```
Informasi dataset:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 13 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   usia                    300 non-null    int64
 1   laki-laki               300 non-null    int64
 2   perempuan               300 non-null    int64
 3   pendidikan terakhir     300 non-null    int64
 4   keterampilan digital    300 non-null    int64
 5   pengalaman kerja        300 non-null    int64
 6   jenis pekerjaan         300 non-null    int64
 7   model kerja             300 non-null    int64
 8   media sosial            300 non-null    int64
 9   personal branding       300 non-null    int64
 10  pelatihan keterampilan  300 non-null    int64
 11  penghasilan utama       300 non-null    int64
 12  tipe kesuksesan         300 non-null    int64
dtypes: int64(13)
memory usage: 30.6 KB
None

Jumlah baris: 300
Jumlah kolom: 13
```

Gambar 2. Informasi Dataset

Informasi Dataset yang tercantum dalam Gambar 2, dataset digunakan dalam penelitian ini terdiri dari 300 baris data dan 13 kolom atribut. Setiap baris merepresentasikan satu individu dari Generasi Z, sementara setiap kolom menyimpan informasi mengenai berbagai faktor yang berpotensi memengaruhi kesuksesan seseorang. Semua kolom bertipe integer (int64) dan tidak mengandung nilai kosong (non-null), sehingga data sudah lengkap dan siap untuk dianalisis lebih lanjut.

```
duplikat = df.duplicated().sum()
print(f"\nJumlah data duplikat: {duplikat}")

if duplikat > 0:
    df.drop_duplicates(inplace=True)
    print(f"✅ Duplikat dihapus. Jumlah data sekarang: {df.shape[0]} baris.")
else:
    print("✅ Tidak ada data duplikat.")
```

```
Jumlah data duplikat: 13
✅ Duplikat dihapus. Jumlah data sekarang: 287 baris.
```

Gambar 3. Distribusi Data Setelah Penghapusan Duplikat

Pada Gambar 3, terdapat 13 baris data duplikat pada dataset. Data duplikat ini dapat mempengaruhi kualitas model karena berisiko menimbulkan bias pada proses pelatihan, terutama ketika data yang sama muncul secara berulang. Oleh karena itu, semua baris duplikat telah dihilangkan untuk memastikan keakuratan dan representasi yang seimbang dalam analisis. Setelah proses pembersihan ini dilakukan, total data yang siap digunakan untuk pemodelan menjadi 287 baris. Dengan menghilangkan data yang tidak unik, proses analisis selanjutnya dapat dilakukan dengan lebih optimal dan valid.

```
X = df.drop(columns='tipe kesuksesan')
y = df['tipe kesuksesan']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=y)

print(f"Jumlah data latih  : {X_train.shape[0]} baris")
print(f"Jumlah data uji    : {X_test.shape[0]} baris")
```

```
Jumlah data latih  : 229 baris
Jumlah data uji    : 58 baris
```

DOI: https://doi.org/10.38204/tematik.v12i1.2334
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Gambar 4. Pemisahan Data Latih Dan Data Uji

Pada Gambar 4, dataset yang telah dibersihkan kemudian dibagi menjadi dua bagian utama, yaitu data latih dan data uji. Sebanyak 229 baris data digunakan sebagai data latih, yang berfungsi untuk membangun dan melatih model agar dapat mengenali pola dari fitur-fitur yang tersedia. Sementara itu, 58 baris data lainnya digunakan sebagai data uji, yang berperan penting dalam mengevaluasi performa model terhadap data yang belum pernah dilihat sebelumnya. Pembagian ini bertujuan untuk memastikan bahwa model yang dihasilkan tidak hanya baik dalam menghafal data, tetapi juga mampu melakukan generalisasi terhadap data baru dengan akurat.

Tabel 2. Hasil Perolehan Data

Usia	Jenis Kelamin	pendidikan terakhir	keterampilan digital	pengalaman kerja	jenis pekerjaan	model kerja	media sosial	personal branding	pelatihan keterampilan	penghasilan utama	tipe kesuksesan
21	Laki-laki	SMA	5	Kurang dari 1 tahun	Pegawai kontrak	Konvensional	Ya	3	Ya	Pekerjaan tetap	Keseimbangan hidup
21	Laki-laki	SMA	5	Kurang dari 1 tahun	Freelancer	Hybrid	Ya	2	Tidak	Freelancer	Karir (Jabatan)
22	Perempuan	SMA	5	Kurang dari 1 tahun	Freelancer	Hybrid	Ya	3	Ya	Bisnis	Keseimbangan hidup
23	Perempuan	SMA	4	Kurang dari 1 tahun	Freelancer	Remote	Ya	2	Ya	Bisnis	Pengembangan diri
21	Perempuan	SMA	5	Kurang dari 1 tahun	Freelancer	Remote	Ya	2	Ya	Freelancer	Pengembangan diri
22	Perempuan	SMA	5	Kurang dari 1 tahun	Freelancer	Remote	Ya	2	Tidak	Bisnis	Pengembangan diri
24	Perempuan	SMA	4	1 - 3 tahun	Pegawai kontrak	Konvensional	Ya	2	Ya	Pekerjaan tetap	Keseimbangan hidup
22	Perempuan	SMA	2	Kurang dari 1 tahun	Freelancer	Remote	Ya	2	Ya	Freelancer	Pengembangan diri
23	Laki-laki	SMA	5	4 - 6 tahun	Pegawai kontrak	Konvensional	Ya	2	Tidak	Pekerjaan tetap	Finansial

Tabel 3. Data Setelah Tranformasi Data

Usia	Laki-laki	perempuan	pendidikan terakhir	keterampilan digital	pengalaman kerja	jenis pekerjaan	model kerja	media sosial	personal branding	pelatihan keterampilan	penghasilan utama	tipe kesuksesan
21	1	0	1	5	1	3	1	1	3	1	1	4
21	1	0	1	5	1	1	3	1	2	0	2	2
22	0	1	1	5	1	1	3	1	3	1	3	4
23	0	1	1	4	1	1	2	1	2	1	3	3
21	0	1	1	5	1	1	2	1	2	1	2	3
22	0	1	1	5	1	1	2	1	2	0	3	3
24	0	1	1	4	2	3	1	1	2	1	1	4
22	0	1	1	2	1	1	2	1	2	1	2	3
23	1	0	1	5	3	3	1	1	2	0	1	1

Berdasarkan Gambar 5, hasil evaluasi model Naïve Bayes tanpa teknik *balancing*, diperoleh akurasi sebesar 34,48%, yang menunjukkan bahwa model hanya mampu mengklasifikasikan data dengan benar sebanyak 34% dari total data yang diuji. Hasil *classification report* memperlihatkan bahwa performa model bervariasi pada setiap kelas. Label 4 memiliki performa terbaik dengan *precision* sebesar 0.44, *recall* sebesar 0.48, dan *f1-score* sebesar 0.46. Sementara itu, label 2 memiliki performa terendah dengan *precision* sebesar 0.08 dan *f1-score* sebesar 0.10, yang menunjukkan bahwa model kesulitan dalam mengenali kelas ini. Nilai *macro average f1-score* sebesar 0.30 dan *weighted average f1-score* sebesar 0.35 juga menunjukkan bahwa secara umum, model masih kurang optimal dalam melakukan klasifikasi. Hal ini diperkuat oleh hasil *confusion matrix* yang menunjukkan banyaknya prediksi yang tersebar ke kelas lain, terutama dominasi prediksi pada label 4 yang ditandai dengan warna biru tua pada visualisasi. Sebagian besar kesalahan klasifikasi terjadi karena model cenderung memetakan data ke label 4, yang mengindikasikan kemungkinan adanya ketidakseimbangan jumlah data antar kelas (*data imbalance*). Oleh karena itu, untuk meningkatkan kinerja model, disarankan untuk melakukan *balancing* data serta mempertimbangkan teknik peningkatan fitur atau penggunaan algoritma lain yang lebih sesuai.

Gambar 6 menunjukkan langkah-langkah yang dilakukan untuk menggandakan data atau teknik *random oversampling* sebanyak 10 kali dari data awal. Setelah penggandaan, jumlah total data bertambah menjadi 2.870 baris, dari semula hanya 287 (asumsi berdasarkan pembagian hasil kelas). Distribusi kelas sebelum dilakukan *balancing* ditampilkan dengan memanggil *.value_counts()* pada kolom tipe kesuksesan, yang menunjukkan bahwa kelas dengan label 4 "Keseimbangan Hidup" memiliki jumlah data

DOI: https://doi.org/10.38204/tematik.v12i1.2334
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

paling besar, yakni 1.220, sementara kelas 2 hanya berjumlah 420.

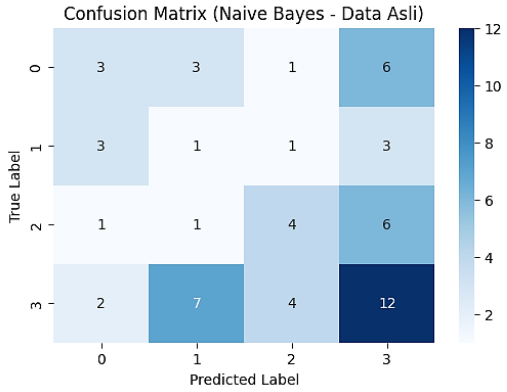
Tujuan dari *random oversampling* ini adalah untuk memperbesar jumlah data minoritas agar model pembelajaran mesin tidak bias terhadap kelas mayoritas. Dengan demikian, proses klasifikasi akan menjadi lebih adil dan memiliki performa yang lebih baik dalam mengenali tiap tipe kesuksesan yang berbeda.

```
🎯 Akurasi Naive Bayes (Tanpa Balancing): 34.48%

📋 Classification Report:
              precision    recall  f1-score   support

           1       0.33      0.23      0.27        13
           2       0.08      0.12      0.10         8
           3       0.40      0.33      0.36        12
           4       0.44      0.48      0.46        25

    accuracy                           0.34        58
   macro avg       0.32      0.29      0.30        58
weighted avg       0.36      0.34      0.35        58
```



Gambar 5. Hasil Akurasi Naïve Bayes Tanpa *Balancing*

```
# Gandakan semua data sebanyak 10x
df_digandakan = pd.concat([df] * 10, ignore_index=True)

# Cek jumlah baris setelah digandakan
print(f"Jumlah data setelah digandakan: {df_digandakan.shape[0]}")
print("Distribusi kelas sebelum balancing:")
print(df_digandakan['tipe kesuksesan'].value_counts())
```

```
Jumlah data setelah digandakan: 2870
Distribusi kelas sebelum balancing:
tipe kesuksesan
4    1220
1     630
3     600
2     420
Name: count, dtype: int64
```

Gambar 6. Proses *Random Oversampling* Untuk Menyeimbangkan Kelas Tipe Kesuksesan

```
Distribusi label setelah balancing:
tipe kesuksesan
1    288
3    288
2    288
4    288
Name: count, dtype: int64
Total baris setelah balancing: 1152
```

Gambar 7. Hasil Distribusi Data Setelah *Balancing* Menggunakan *Random Oversampling*

Gambar 7 menunjukkan bahwa setiap kelas—yang mewakili berbagai tipe kesuksesan Generasi Z seperti finansial, karir (jabatan), pengembangan diri, dan keseimbangan hidup yang mana masing-masing memiliki 288 data. Ini menghasilkan total keseluruhan 1.152 baris data yang siap digunakan untuk pelatihan model klasifikasi. Dengan menyamakan jumlah data di setiap kelas melalui *random oversampling*, kini model dapat belajar dari seluruh tipe kesuksesan secara merata. Ini penting untuk memastikan bahwa prediksi yang dilakukan lebih adil dan representatif terhadap kondisi nyata Generasi Z yang memiliki berbagai orientasi kesuksesan.

Pada Gambar 8, Setelah proses pelatihan menggunakan data latih sebanyak 921 baris dan pengujian pada 231 baris data uji, model Naïve Bayes yang dibangun dengan data yang telah diseimbangkan menunjukkan tingkat akurasi sebesar 36,80%. Metrik evaluasi lainnya menunjukkan bahwa nilai *precision* untuk keempat kelas berkisar antara 0.36 hingga 0.39, sedangkan nilai *recall* berkisar antara 0.24 hingga 0.46, dan nilai *f1-score* tertinggi ada pada label 2 dengan nilai 0.41. Hal ini menunjukkan bahwa meskipun distribusi data sudah seimbang, performa model masih tergolong rendah, dengan kecenderungan model cukup kesulitan membedakan antar kelas secara akurat.

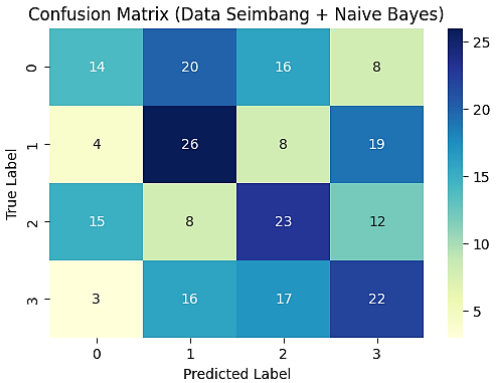
```
Jumlah data latih: 921
Jumlah data uji  : 231

🎯 Akurasi Akhir Model (Data Baru + Seimbang): 36.80%

📋 Classification Report:
              precision    recall  f1-score   support

           1       0.39      0.24      0.30        58
           2       0.37      0.46      0.41        57
           3       0.36      0.40      0.38        58
           4       0.36      0.38      0.37        58

    accuracy                           0.37       231
   macro avg       0.37      0.37      0.36       231
weighted avg       0.37      0.37      0.36       231
```

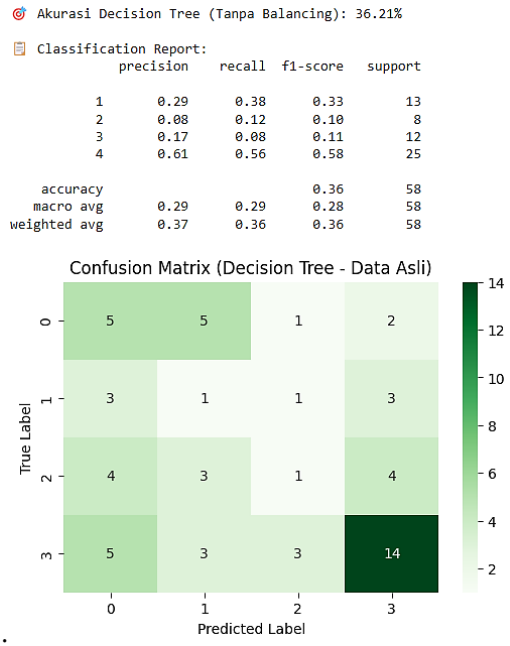


Gambar 8. Hasil Akurasi Naïve Bayes Dengan *Balancing*

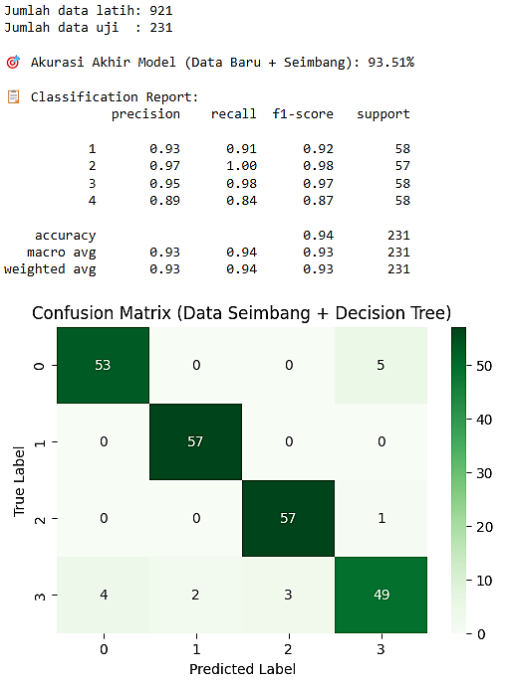
Pada Gambar 9 menampilkan hasil akurasi model Decision Tree yang diterapkan pada data tanpa *balancing*. Akurasi model yang diperoleh hanya sebesar 36,21%, yang menunjukkan performa klasifikasi yang masih lemah. Dari laporan klasifikasi, dapat dilihat bahwa *precision, recall, dan f1-score* untuk sebagian

DOI: https://doi.org/10.38204/tematik.v12i1.2334
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

besar kelas sangat rendah, khususnya untuk tipe kesuksesan kelas 2 dan 3 yang bahkan hanya memiliki *f1-score* sebesar 0.10 dan 0.11.



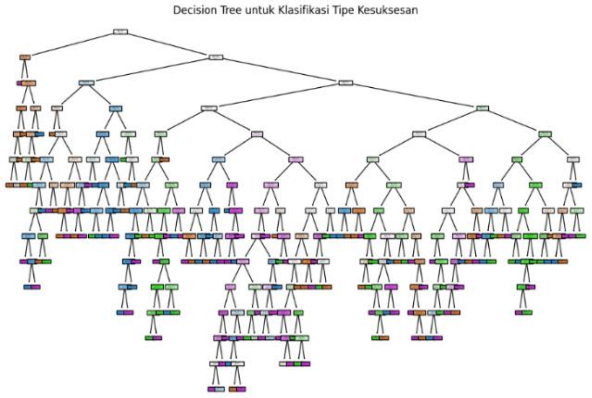
Gambar 9. Hasil Dari Akurasi Decision Tree Tanpa *Balancing*



Gambar 10. Hasil Akurasi Decision Tree Dengan *Balancing*

Pada Gambar 10 menampilkan hasil evaluasi model Decision Tree setelah dilakukan data *balancing*, yang menunjukkan peningkatan performa yang sangat signifikan. Dengan jumlah data latih sebanyak 921 dan data uji sebanyak 231, akurasi akhir model tercatat sebesar 94,51%, yang mencerminkan tingkat prediksi yang sangat tinggi dan konsisten di seluruh kelas.

Berdasarkan *classification report*, setiap label tipe kesuksesan dimulai dari kelas 1 hingga 4 memiliki nilai *precision, recall, dan f1-score* di atas 0.89, dengan nilai f1 tertinggi berada pada kelas 3 sebesar 0.97. Nilai *macro average dan weighted average* juga sama-sama tinggi di angka 0.93-0.94, menunjukkan bahwa model memiliki performa yang seimbang di semua kelas, tanpa memihak pada label tertentu.



Gambar 11. Pohon Keputusan

Berdasarkan visualisasi dan struktur Decision Tree pada Gambar 11, dapat disimpulkan bahwa model klasifikasi ini secara sistematis membentuk pola logika dari fitur-fitur utama dalam dataset untuk memprediksi tipe kesuksesan generasi Z. Model dimulai dari _root node_ yang menggunakan fitur paling berpengaruh, seperti penghasilan utama, model kerja, atau pengalaman kerja, sebagai titik awal dalam memisahkan data. Dari node utama ini, pohon bercabang secara bertingkat berdasarkan keputusan logis dari setiap fitur, membentuk jalur yang menggambarkan aturan klasifikasi, misalnya "*freelancer* + pengalaman kerja ≤ 3 tahun + aktif media sosial" yang cenderung berujung pada klasifikasi ke tipe "Pengembangan Diri".

Pohon ini membuktikan bahwa algoritma _decision tree_ mampu mengenali keterkaitan antar variabel dalam menentukan preferensi atau persepsi terhadap kesuksesan. Model ini dapat digunakan untuk melakukan prediksi terhadap data baru dengan akurasi dan transparansi yang baik. Selain itu, kedalaman dan kerumitan struktur pohon mencerminkan banyaknya kondisi dan kombinasi atribut yang berkontribusi terhadap klasifikasi, menjadikan model ini tidak hanya sebagai alat prediksi, tapi juga sebagai alat eksplorasi pemahaman pola pikir generasi Z terkait kesuksesan.

Tabel 4. Perbandingan Hasil

Metode	Hasil tanpa *Balancing*	Hasil dengan *Balancing*
Naïve Bayes	34,48%	36,80%
Decision Tree	36,21%	93,51%

Berdasarkan Tabel 4 yang menampilkan hasil perbandingan performa antara algoritma Naïve Bayes

DOI: https://doi.org/10.38204/tematik.v12i1.2334
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

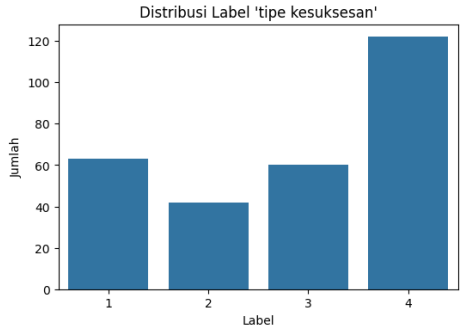
dan Decision Tree dalam klasifikasi tipe kesuksesan Generasi Z, terdapat perbedaan yang signifikan antara hasil tanpa teknik *balancing* dan dengan teknik *balancing*.

Pada algoritma Naïve Bayes, akurasi klasifikasi tanpa dilakukan *balancing* tercatat sebesar 34,48%, dan meningkat menjadi 36,80% setelah dilakukan balancing. Kenaikan ini relatif kecil, menandakan bahwa Naïve Bayes kurang sensitif terhadap perubahan distribusi data atau bahwa model ini memiliki keterbatasan dalam menangani ketidakseimbangan kelas yang kompleks dalam konteks data ini.

Sebaliknya, pada algoritma Decision Tree, terjadi peningkatan akurasi yang sangat drastis. Akurasi awal tanpa *balancing* berada pada angka 36,21%, namun setelah dilakukan *balancing*, akurasi melonjak secara signifikan menjadi 93,51%. Hal ini menunjukkan bahwa Decision Tree sangat diuntungkan oleh proses *balancing* data, karena algoritma ini sangat bergantung pada struktur distribusi data dalam pembentukan pohon keputusan. Dengan distribusi kelas yang seimbang, Decision Tree dapat menghasilkan pemisahan yang lebih efektif dan representatif untuk setiap label kelas.

Secara keseluruhan, hasil ini menunjukkan bahwa dalam kasus klasifikasi tipe kesuksesan Generasi Z dengan dataset tidak seimbang, teknik *balancing* seperti *random oversampling* sangat penting untuk meningkatkan performa model, terutama pada algoritma seperti Decision Tree. Sementara itu, Naïve Bayes mungkin memerlukan pendekatan tambahan seperti pemilihan fitur atau tuning parameter lebih lanjut agar kinerjanya meningkat secara signifikan.

Dengan hasil ini, dapat disimpulkan bahwa penerapan data *balancing* sebelum proses pelatihan model Decision Tree sangat efektif untuk meningkatkan akurasi klasifikasi tipe kesuksesan Generasi Z, khususnya berdasarkan kategori yang sebelumnya kurang terwakili.



Gambar 12. Diagram Hasil Tipe Kesuksesan

Distribusi label pada variabel “tipe kesuksesan” pada Gambar 12, menunjukkan bahwa mayoritas generasi Z dalam dataset ini lebih memaknai kesuksesan sebagai keseimbangan hidup dibandingkan indikator tradisional seperti jabatan atau kekayaan finansial. Label 4, yang merepresentasikan Keseimbangan Hidup, menjadi kategori paling dominan dengan jumlah responden lebih dari 120 orang, mengindikasikan bahwa aspek seperti waktu pribadi, kesehatan mental, dan harmoni antara kehidupan kerja dan personal sangat dihargai oleh generasi ini. Sebaliknya, Karir (Jabatan) sebagai label 2 merupakan kategori paling sedikit dipilih, menandakan menurunnya relevansi status formal dalam mendefinisikan kesuksesan. Sementara itu, label 1 Finansial dan label 3 Pengembangan Diri menunjukkan distribusi yang relatif seimbang, memperlihatkan bahwa kemandirian ekonomi dan aktualisasi diri tetap menjadi bagian penting dari makna sukses. Temuan ini mengonfirmasi hipotesis bahwa generasi Z memiliki orientasi kesuksesan yang lebih menyeluruh dan personal, yang memerlukan penyesuaian dalam kebijakan pendidikan, ketenagakerjaan, dan pengembangan individu untuk mengakomodasi nilai-nilai tersebut.

4. Kesimpulan

Penelitian ini berhasil mengembangkan model klasifikasi untuk mengidentifikasi tipe kesuksesan Generasi Z berdasarkan tingkat penghasilan dan berbagai faktor penentu lainnya, menggunakan algoritma Naïve Bayes dan Decision Tree dalam kerangka metode CRISP-DM. Hasil analisis menunjukkan bahwa atribut seperti pendidikan terakhir, keterampilan digital, pengalaman kerja, jenis pekerjaan, *personal branding*, dan penggunaan media sosial memiliki pengaruh signifikan dalam menentukan tipe kesuksesan. Dari dua algoritma yang diuji, Decision Tree dengan pendekatan *balancing* melalui teknik *random oversampling* terbukti memberikan hasil yang paling optimal dengan akurasi mencapai 94%. Sebaliknya, algoritma Naïve Bayes hanya mencapai akurasi 37% meskipun telah diterapkan *balancing* data. Temuan ini menunjukkan bahwa Decision Tree lebih unggul dalam menangani data kompleks dan multivariat dalam konteks klasifikasi sosial-ekonomi Generasi Z. Dengan demikian, penelitian ini menyimpulkan bahwa pendekatan data mining berbasis Decision Tree sangat potensial digunakan untuk memahami dan memetakan orientasi kesuksesan generasi muda secara lebih akurat dan aplikatif.

Daftar Rujukan

[1] C. T. Widiastuti, “Analisis Gaya Hidup Sebagai Variabel Mediasi Pada Perilaku Konsumtif Belanja Online,” vol. 21, no. 4, 2023.

[2] M. I. Revilliano Trisandri, B. Aji Putra, and A. Aldo Nugroho, “PENGARUH KOMPENSASI FINANSIAL TERHADAP NIAT UNTUK KELUAR MELALUI KEPUASAN KERJA GENERASI Z SEKTOR JASA DI JAKARTA,” *Jurnal Review Pendidikan dan Pengajara*, vol. 7 No 4, 2024, Dec. 2024, Accessed: Mar. 20, 2025. [Online]. Available: http://journal.universitaspahlawan.ac.id/index.php/jrpp

DOI: https://doi.org/10.38204/tematik.v12i1.2334
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

[3] M.-A. Lazar, A. Zbuchea, and F. Pînzaru, "The Emerging Generation Z Workforce in the Digital World: A Literature Review on Cooperation and Transformation," *Proceedings of the International Conference on Business Excellence*, vol. 17, no. 1, pp. 1991–2001, Jul. 2023, doi: 10.2478/picbe-2023-0175.

[4] D. Mavhungu, E. N. Barkhuizen, and M. Magau, "Predicting the career success of Generation Z employees: A socioeconomic perspective," *South African Journal of Economic and Management Sciences*, vol. 27, no. 1, Nov. 2024, doi: 10.4102/sajems.v27i1.5788.

[5] M. G. Herlina and D. H. Syahchari, "Soft Skill Analysis: Rasch Model on Generation Z in Greater Jakarta," *Jurnal Pendidikan Ekonomi Dan Bisnis (JPEB)*, vol. 10, no. 2, pp. 100–110, May 2022, doi: 10.21009/JPEB.010.2.1.

[6] T. A. Suciati and D. Deswarta, "Pengaruh Pelatihan Kerja, Tingkat Pendidikan, dan Pengalaman Kerja terhadap Kinerja Karyawan Generasi Z di Selat Panjang," *Al Qalam: Jurnal Ilmiah Keagamaan dan Kemasyarakatan*, vol. 18, no. 1, p. 58, Jan. 2024, doi: 10.35931/aq.v18i1.3008.

[7] M. Zaid and RajendranT, "Higher Classification Accuracy of Income Class Using Decision Tree Algorithm over Naive Bayes Algorithm," 2022. doi: 10.3233/APC220079.

[8] S. Sang-Cheol Moon, "Syncretistic Scientism and Dissonant Plausibility Structure: An Analysis of Worldview Orientation Among Generation Zers in South Korea," *Humanities and Social Sciences*, vol. 9, no. 1, p. 14, 2021, doi: 10.11648/j.hss.20210901.13.

[9] Arya Satya Pratama, Suci Maela Sari, Maila Faiza Hj, Moh Badwi, and Mochammad Isa Anshori, "Pengaruh Artificial Intelligence, Big Data Dan Otomatisasi Terhadap Kinerja SDM Di Era Digital," *Jurnal Publikasi Ilmu Manajemen*, vol. 2, no. 4, pp. 108–123, Oct. 2023, doi: 10.55606/jupiman.v2i4.2739.

[10] D. Kurniawan and D. M. Yasir, "Optimization Sentiment Analysis Using CRISP-DM And Naïve Bayes Methods Implemented On Social Media," Oct. 2022.

[11] Salma Navisa, Luqman Hakim, and Aulia Nabilah, "Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM," *Jurnal Sistem Cerdas*, pp. 114–125, 2021.

[12] M. B. Ryando and S. Sutarman, "Clustering of Eligibility and Characteristics of Smart Indonesia Card Recipients for College using Agglomerative Hierarchical Clustering," *JURNAL SISFOTEK GLOBAL*, vol. 14, no. 2, p. 137, Sep. 2024, doi: 10.38101/sisfotek.v14i2.15707.

[13] A. Handijono, "Memanfaatkan Data Mining danKnowledge Management System Untuk Mengoptimalkan Strategi Pemasaran pada Bank," 2022.

[14] A. Pratiwi, A. T. Sasongko, and D. K. Pramudito, "Analisis Prediksi Gilingan Plastik Terlaris Menggunakan Algoritma K-Nearest Neighbor Di CV Menembus Batas," *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 5, no. 3, pp. 437–445, 2023.

[15] R. R. Adhitya, Wina Witanti, and Rezki Yuniarti, "Perbandingan Metode CART Dan Naïve Bayes Untuk Klasifikasi Customer CHURN," *INFOTECH journal*, vol. 9, no. 2, pp. 307–318, Jul. 2023, doi: 10.31949/infotech.v9i2.5641.

[16] R. R. Rahmah, I. R. Setiawan, and F. Frazna Az-Zahra, "Prediksi Keterlambatan Mahasiswa Dalam Membayar Biaya Kuliah Menggunakan Algoritma C4.5 (Studi Kasus : Universitas Muhammadiyah Sukabumi)," 2024.

[17] E. G. G. Sukmo, G. Firmansyah, and B. Tjahjono, "Perbandingan Prediksi Keberlanjutan Polis Asuransi Dengan Menggunakan Algoritma Naive Bayes, KNN, dan SVM," *Jurnal Bisnis dan Manajemen*, vol. 4, no. 2, pp. 24–43, 2024.

[18] A.- Husaini, I. Hariyanti, and A. R. Raharja, "Perbandingan Algoritma Decision Tree dan Naive Bayes dalam Klasifikasi Data Pengaruh Media Sosial dan Jam Tidur Terhadap Prestasi Akademik Siswa," *Technologia : Jurnal Ilmiah*, vol. 15, no. 2, p. 332, Apr. 2024, doi: 10.31602/tji.v15i2.14381.

DOI: https://doi.org/10.38204/tematik.v12i1.2334
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)