

PENERAPAN MODEL KLASIFIKASI PADA KELULUSAN MAHASISWA DI UNIVERSITAS PERJUANGAN TASIKMALAYA MENGGUNAKAN ALGORITMA C4.5

Salsabila Dewi Nugraha ^{1*}, Missi Hikmatyar ², Ruuhwan ³

^{1,2,3} Jurusan Teknik Informatika, Fakultas Teknik, Universitas Perjuangan Tasikmalaya

^{1,2,3} JL.Peta No. 177, Kahuripan, Kec. Tawang, Kota Tasikmalaya, 461115

Received: 1 Agustus 2024

Accepted: 6 Agustus 2024

Published: 7 Agustus 2024

Keywords:

Classification

Decision Tree

Prediction of student graduation

Correspondent Email:

salsabiladewinugraha@gmail.com

Abstrak. Penelitian ini mengatasi tantangan penumpukan mahasiswa dan bertujuan untuk memprediksi kelulusan tepat waktu di Universitas Perjuangan Tasikmalaya. Menurut UNESCO, peningkatan kualitas pendidikan merupakan kunci untuk meningkatkan kualitas suatu bangsa secara keseluruhan. Kelulusan tepat waktu adalah faktor penting dalam menilai akreditasi pendidikan tinggi dan efisiensi pendidikan. Namun, memprediksi kelulusan mahasiswa bisa menjadi tantangan tersendiri. Untuk mengatasi masalah ini, penelitian ini menggunakan teknik data mining dengan Algoritma Decision Tree untuk memprediksi kelulusan mahasiswa. Penelitian sebelumnya telah menunjukkan efektivitas metode klasifikasi pada berbagai jenis data. Dalam penelitian ini, Algoritma Decision Tree memberikan hasil yang mengesankan, dengan tingkat akurasi sebesar 92,233% ketika menggunakan seluruh dataset sebagai set pelatihan, dan akurasi sebesar 90,24% dengan validasi silang 80%. Hasil ini menyoroti tingginya akurasi Algoritma Decision Tree dalam mengklasifikasikan hasil kelulusan mahasiswa. Oleh karena itu, penelitian ini akan menggunakan Algoritma Decision Tree untuk menganalisis faktor-faktor yang mempengaruhi kelulusan di Universitas Perjuangan Tasikmalaya, memberikan wawasan dalam meningkatkan hasil pendidikan dan mendukung kelulusan mahasiswa tepat waktu.

Abstract. This study addresses the challenge of student accumulation and seeks to predict timely graduation at the University of Perjuangan Tasikmalaya. According to UNESCO, enhancing the quality of education is key to improving a nation's overall quality. Timely graduation is a critical factor in assessing higher education accreditation and educational efficiency. However, predicting student graduation can be challenging. To tackle this issue, this study utilizes data mining techniques with the Decision Tree Algorithm to predict student graduation. Previous research has demonstrated the effectiveness of classification methods across various data types. In this study, the Decision Tree Algorithm yielded impressive results, with an accuracy rate of 92.233% when using the entire dataset as the training set, and 90.24% accuracy with 80% cross-validation. These results highlight the high accuracy of the Decision Tree Algorithm in classifying student graduation outcomes. Consequently, this study will employ the Decision Tree Algorithm to analyze the factors influencing graduation at the University of Perjuangan Tasikmalaya, providing insights into improving educational outcomes and supporting timely student graduation.

1. PENDAHULUAN

Dalam usaha meningkatkan kualitas sebuah bangsa, UNESCO menyatakan bahwa peningkatan mutu pendidikan merupakan satu-satunya cara untuk mencapai hal tersebut.[1] Salah satu unsur penilaian dalam akreditasi perguruan tinggi adalah tingkat kelulusan mahasiswa tepat waktu, yang merupakan parameter penting dalam penilaian akreditasi.[2] Oleh karena itu, penting untuk menemukan teknik yang dapat memprediksi kelulusan mahasiswa agar dapat membantu dalam penilaian akreditasi perguruan tinggi.

Terdapat permasalahan dalam hal ketidakseimbangan antara jumlah mahasiswa yang masuk dan jumlah mahasiswa yang lulus tepat waktu dalam menyelesaikan studi mereka. Mahasiswa yang masuk dalam jumlah besar, tetapi jumlah mahasiswa yang lulus tepat waktu (dalam 4 tahun) jauh lebih kecil dibandingkan mahasiswa yang masuk setiap tahunnya. Hal ini menyebabkan akumulasi mahasiswa dengan jumlah yang tinggi setiap periode. Untuk mengatasi permasalahan ini, diperlukan teknik prediksi kelulusan mahasiswa. Salah satu teknik yang sering digunakan adalah data mining, khususnya metode klasifikasi.

Dalam penelitian ini, metode klasifikasi yang digunakan adalah Algoritma C.45, yang menggunakan probabilitas dan statistik untuk mencari jawaban dan membantu dalam pengambilan keputusan mengenai kelulusan mahasiswa dengan mempertimbangkan berbagai faktor yang mempengaruhinya. Model klasifikasi ini dapat digunakan untuk mengklasifikasikan data numerik dan kategorikal sesuai dengan kriteria yang diinginkan.[3]

Penelitian sebelumnya telah membuktikan bahwa metode klasifikasi dapat digunakan untuk berbagai tipe data dengan karakteristik tertentu.[4] Studi sebelumnya juga telah mengimplementasikan algoritma klasifikasi untuk prediksi kelulusan mahasiswa dengan tingkat akurasi yang cukup baik. [5] [6]

Penelitian lainnya menggunakan algoritma C.45 untuk menganalisis faktor-faktor yang mempengaruhi kelulusan mahasiswa dan telah mencapai tingkat akurasi yang tinggi, mencapai 92,233% dengan menggunakan seluruh data sebagai training set dan 90,24% dengan menggunakan cross-validation 80% . [7] [8]

Berdasarkan penelitian sebelumnya, pada penelitian ini akan digunakan model klasifikasi dengan menggunakan algoritma C.45 untuk mencari faktor-faktor yang mempengaruhi kelulusan mahasiswa di Universitas Perjuangan Tasikmalaya. Metode ini diharapkan dapat membantu meningkatkan ketepatan prediksi kelulusan mahasiswa dan memberikan wawasan yang berharga dalam mengatasi masalah penumpukan mahasiswa dalam jumlah tinggi di setiap periode

2. TINJAUAN PUSTAKA

2.1 Kelulusan

Kelulusan mahasiswa bukan sekadar formalitas administratif, tetapi merupakan pencapaian penting yang mencerminkan terpenuhinya standar kompetensi yang telah ditetapkan. Standar ini mencakup kemampuan dalam sikap, pengetahuan, dan keterampilan yang harus dimiliki oleh lulusan. Hal ini menunjukkan bahwa kelulusan bukan hanya soal menyelesaikan rangkaian pembelajaran, tetapi juga tentang membentuk individu yang siap berkontribusi secara profesional dan sosial. Oleh karena itu, kelulusan adalah pengakuan resmi atas keberhasilan mahasiswa dalam mencapai kualifikasi yang diharapkan dari suatu program studi. [9]

2.2 Klasifikasi

Klasifikasi adalah salah satu teknik data mining yang sangat penting dan serbaguna, dengan aplikasi yang luas dalam berbagai bidang. [10] Proses ini tidak hanya memungkinkan pengelompokan data berdasarkan ciri-ciri tertentu, tetapi juga menawarkan fleksibilitas dalam metode.

2.3 Decision Tree

Decision Tree adalah metode klasifikasi yang unggul dalam menangani missing value dan memberikan interpretasi hasil yang mudah dipahami. Dengan struktur seperti flowchart, Decision Tree menawarkan pendekatan yang sistematis dalam melakukan klasifikasi, di mana setiap keputusan diambil berdasarkan evaluasi atribut tertentu. Metode ini menunjukkan performa

yang baik, terutama pada dataset yang besar, dengan tingkat akurasi yang lebih tinggi dibandingkan algoritma pohon keputusan lainnya. [12] Selain itu, penggunaan ukuran statistik seperti information gain dan entropy membantu dalam mengevaluasi efektivitas atribut, menjadikan Decision Tree sebagai alat yang sangat efektif dalam pembentukan model klasifikasi yang andal.

2.2.1 Entropy

Entropy secara istilah adalah keberbedaan atau keberagaman, semakin heterogeny suatu himpunan data, semakin besar pula nilai entropy-nya. Cara menghitung nilai entropy sebagai berikut :

$Entropy(S) = - \sum_{i=1}^n p_i \log_2 p_i$
 Dimana n adalah jumlah nilai yang terdapat pada atribut target (jumlah kelas). Sedangkan pi menyatakan rasio antara jumlah sample di kelas i dengan jumlah semua sample pada himpunan data

2.2.2 Information Gain

Secara istilah information gain adalah perolehan informasi, di definisikan sebagai ukuran efektivitas suatu atribut dalam mengklasifikasikan data, cara mengukur information gain adalah :

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

Keterangan:

- S : himpunan
- A : atribut
- n : jumlah partisi atribut A
- | Si | : jumlah kasus pada partisi ke-i
- | S | : jumlah kasus dalam S

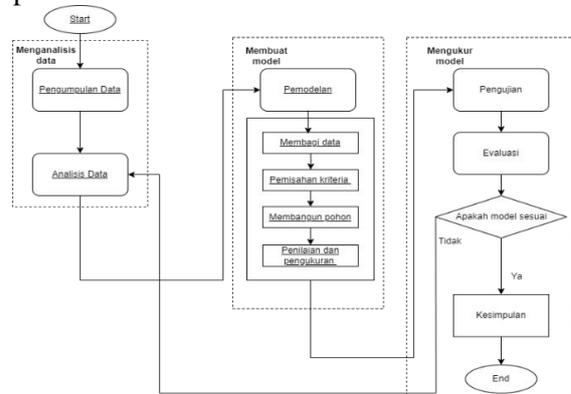
2.4 Phyton

Secara keseluruhan, Python adalah bahasa pemrograman yang sangat fleksibel dan mudah digunakan, terutama dalam pengembangan aplikasi yang memerlukan integrasi dengan berbagai bahasa pemrograman dan penyimpanan data. [11] Python dirancang dengan fitur-fitur yang memudahkan programmer dalam penulisan, pembacaan, dan pemeliharaan kode. Dengan dukungan

library yang luas dan modularitas yang tinggi, Python memungkinkan pengembangan yang cepat dan efisien. Selain itu, kemampuannya dalam memanfaatkan berbagai algoritma klasifikasi menjadikan Python alat yang kuat dalam analisis data, yang pada akhirnya dapat memberikan hasil yang andal untuk pengambilan keputusan.

3. METODE PENELITIAN

Metodologi yang digunakan dalam penelitian ini adalah:



Gambar 1. Metodologi penelitian

3.1. Menganalisis data

Langkah pertama dalam menganalisis adalah pengumpulan data secara langsung dari bagian akademik dan UPTIK. Dataset mahasiswa yang akan digunakan adalah data wisudawan Universitas Perjuangan Tasikmalaya. selanjutnya data di proses dengan membersihkan data dan mengintegrasikan data yang akan digunakan dalam menunjang pengolahan datanya adalah Microsoft excel, dan menggunakan bahasa pemrograman Phyton.

3.2 Pembuatan model

Dalam fase ini merupakan implementasi algoritma C.45 . Dalam tahap ini pertama kali di lakukan adalah dilakukan pembagian data training dan data testing. Data training yang nantinya akan digunakan sebagai sample dalam klasifikasi data yang telah di analisis, data testing digunakan untuk pengujian data.

Setelah data di bagi, pilih kriteria pemisahan yang sesuai untuk pohon keputusan. Pilihan kriteria yang digunakan dalam penelitian ini adalah information gain.

Proses pembangunan pohon ini untuk membuat struktur pohon di mana setiap simpul daun mewakili parameter yang diprediksi. Mulailah dengan simpul akar dan pisahkan data secara rekursif berdasarkan kriteria pemisahan yang dipilih. Di setiap langkah, pilih fitur yang memberikan pemisahan terbaik dan buat simpul anak yang sesuai. Lanjutkan proses ini sampai kondisi berhenti terpenuhi. Pada tahap pemodelan ini akan di lakukan menggunakan bahasa pemrograman *python*. [11]

3.3. Pengukuran

Hasil implementasi algoritma C.45 pada tahapan modelling dapat dilihat pada tahapan pengukuran. Setelah dirasa sesuai maka akan dilakukan pengujian menggunakan metrik evaluasi umum untuk masalah klasifikasi meliputi *accuracy*, *presision*, *recall*, dan *F1-Score*.

Hasil dari pengujian kemudian di evaluasi untuk mengetahui apakah hasil dari pemodelan sesuai dengan kebutuhan atau tidak. Model yang telah terbentuk kemudian diuji juga akurasi, nilai akurasi yang baik dapat menjadi tolak ukur model ini dapat digunakan untuk pengambilan sebuah keputusan di Universitas Perjuangan Tasikmalaya. Nilai akurasi di dapat dari hasil pembuatan pohon keputusan dari data yang telah di tetapkan, nilai akurasi dinyatakan dalam bentuk nilai yang telah di konversi ke dalam persentasi.

4. HASIL DAN PEMBAHASAN

4.1 Analisis Data

Pada penelitian ini peneliti menggunakan data skunder data yang diperoleh dari database mahasiswa yang dimiliki oleh Universitas Perjuangan Tasikmalaya. Data yang dikumpulkan adalah data mahasiswa yang telah melakukan wisuda, Jumlah data yang di proses adalah 1932 sample data.

Data yang di dapatkan adalah nama, nim, jenis kelamin, nama fakultas, nama prodi, tahun masuk, tanggal yudisium, tahun wisuda, dan lama studi dalam bentuk hari, bulan dan tahun. Untuk menambah variasi data maka peneliti mengumpulkan data wisudawan yang pernah melakukan cuti dan aktif organisasi. Namun pada penelitian ini data yang akan di gunakan adalah nama prodi, jenis kelamin, IPK, tahun masuk, cuti, dan aktif organisasi

Setelah data terkumpul dan atribut penentu telah di tetapkan selanjutnya data di beri label untuk mengetahui apakah siswa tersebut lulus tepat waktu atau terlambat, pelabelan ini berdasarkan masa studi, apabila masa studinya kurang lebih sama dengan 4 tahun di nyatakan lulus tepat waktu, apabila lama studinya lebih dari 4 tahun di nyatakan terlambat.

	A	B	C	D	E	F	G	H
1	nama_prodi	Gender	ipk	tahun_masuk	lama_studi	Organisasi	Cuti	Kelulusan
2	farmasi	P	3,22	2015	1	Non	Tidak	Tepat
3	agribisnis	P	3,79	2015	2	Non	Tidak	Tepat
4	akuntansi	L	3,24	2015	2	Non	Tidak	Tepat
5	akuntansi	L	3,09	2015	2	Non	Tidak	Tepat
6	akuntansi	L	3,13	2015	2	Non	Tidak	Tepat
7	akuntansi	L	3,2	2015	2	Non	Tidak	Tepat
8	farmasi	P	3,32	2015	2	Non	Tidak	Tepat
9	farmasi	L	3,35	2015	2	Non	Tidak	Tepat
10	farmasi	P	3,00	2015	2	Non	Tidak	Tepat
11	manajemen	P	3,14	2015	2	Non	Tidak	Tepat
12	manajemen	L	3,48	2015	2	Non	Tidak	Tepat
13	manajemen	P	2,9	2015	2	Non	Tidak	Tepat
14	manajemen	L	3,77	2015	2	Non	Tidak	Tepat
15	manajemen	P	3,23	2015	2	Non	Tidak	Tepat
16	manajemen	P	3,00	2015	2	Non	Tidak	Tepat
17	manajemen	L	3,19	2015	2	Non	Tidak	Tepat
18	manajemen	P	3,81	2015	2	Non	Tidak	Tepat
19	manajemen	L	3,25	2015	2	Non	Tidak	Tepat
20	manajemen	P	3,3	2015	2	Non	Tidak	Tepat
21	manajemen	P	2,82	2015	2	Non	Tidak	Tepat
22	pendidikan guru sekolah dasar	L	3,25	2015	2	Non	Tidak	Tepat
23	pendidikan guru sekolah dasar	L	3,42	2015	2	Non	Tidak	Tepat
24	pendidikan guru sekolah dasar	L	3,36	2015	2	Non	Tidak	Tepat

Gambar 2 Pemberian label

Setelah seluruh atribut diberi label maka simpan file dengan menggunakan CSV. Kemudian file tersebut di upload di *Google Drive*. Selanjutnya dataset yang telah dibuat akan di panggil di goole collab menggunakan *query import csv* untuk diolah. Langkah pertama *import library* yang akan digunakan pada penelitian ini.

```
[38] from sklearn.tree import DecisionTreeClassifier, export_graphviz
04 from sklearn import tree
import pandas as pd
import pydotplus
from IPython.display import Image
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
```

Gambar 3 Import file

Selanjutnya mendefinisikan atribut yang akan di gunakan sebagai nama kolom saat membuat *dataframe*.

```
Atribut=
['nama_prodi', 'Gender', 'IPK', 'lama_studi', 't
ahun_masuk', 'kelulusan', 'cuti', 'organisasi']
```

Disini peneliti menggunakan atribut nama prodi, gender, IPK, lama studi, tahun masuk, cuti, aktif organisasi dan kelulusan sesuai dengan penamaan data yang ada di dalam data yang telah di import ke *google drive*. Untuk memuat data set yang berada di *google drive* peneliti menggunakan fungsi *read_csv* dari *libary pandas*. Setelah data set sudah bisa di tampilkan selanjutnya di analisis untuk menguji apakah ada data yang memiliki nilai *missing value*.

```

0d [74] data.isnull().sum()

nama_prodi      0
Gender          0
IPK             0
lama_studi     0
tahun_masuk    0
Kelulusan      0
Cuti           0
Organisasi     0
dtype: int64
    
```

Gambar 4 Menghitung missing value

Dari gambar di atas kita bisa melihat nilai missing value nya sama dengan nol, maka itu berarti dataset lengkap dan tidak ada nilai yang hilang dalam setiap kolomnya, maka langkah selanjutnya adalah melakukan one-hot encoding pada setiap kolom yang akan digunakan untuk melakukan pemodelan.

```

Siswa=pd.get_dummies(data['nama_prodi','Gender','IPK','tahun_masuk','cuti','organisasi']
    
```

One-hot encoding adalah proses mengubah kolom-kolom object menjadi vektor biner. dengan nilai 0 atau 1 untuk setiap kategori dimana 0 melambangkan ketiadaan atribut dalam kolom yang telah dipisah sesuai kategori yang ada dalam atribut tersebut sedangkan 1 melambangkan sebaliknya

4.2 Hasil model

Pada pemodelan kali ini peneliti akan membuat model dengan data dan pengukuran yang berbeda. Model yang akan dibangun menggunakan data :

1. Data nama prodi, gender, IPK tahun masuk, lama studi, cuti, dan aktif organisasi.
2. Data gender, IPK, tahun masuk, lama studi, cuti, dan aktif organisasi.
3. Data gender, IPK, lama studi, cuti, dan aktif organisasi.

Selanjutnya data di pisahkan menjadi data train dan data test adalah langkah penting dalam pengembangan model decision tree.

Pada penelitian ini akan menggunakan 2 proporsi pengukuran yaitu 20:80 dan 30 : 70 pada data siswa dan kelulusan.

```

0d # Memisahkan data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(siswa, data['Kelulusan'], test_size=0.2, random_state=42)
    
```

Gambar 5 proporsi 20:80

```

0d # Memisahkan data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(siswa, data['Kelulusan'], test_size=0.3, random_state=42)
    
```

Gambar 6 proporsi 30:70

Langkah selanjutnya dilakukan pemisahan kriteria pada penelitian ini menggunakan fungsi entropi yang merupakan langkah penting dalam pembuatan decision tree dan merupakan pendekatan algoritma C.45. Berikut mencari nilai entropi menggunakan rumus python di google collab.

```

0d # Membuat instance DecisionTreeClassifier
model = DecisionTreeClassifier(criterion='entropy')
    
```

Gambar 7 Memisahkan karakter

Langkah selanjutnya setelah memisahkan kriteria dan memilih fitur pemisah dalam decision tree adalah membangun pohon menggunakan model.fit() pada fungsi Python. Pada langkah ini, data train (x_train) dan label train (y_train) digunakan untuk melatih model decision tree.

```

0d model.fit(X_train, y_train)

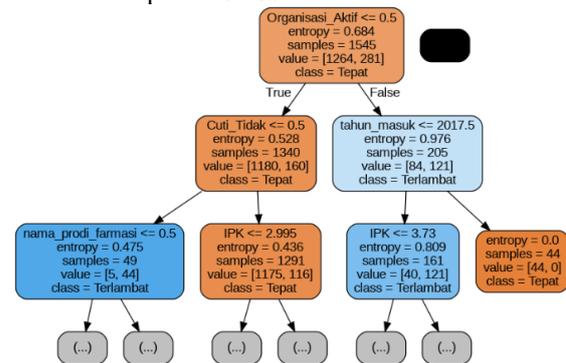
DecisionTreeClassifier
DecisionTreeClassifier(criterion='entropy')
    
```

Gambar 8 Melatih model

Hasil dari pohon keputusan ini di visualisasikan menggunakan modul 'tree' dari Scikit-learn.

1. Model 1

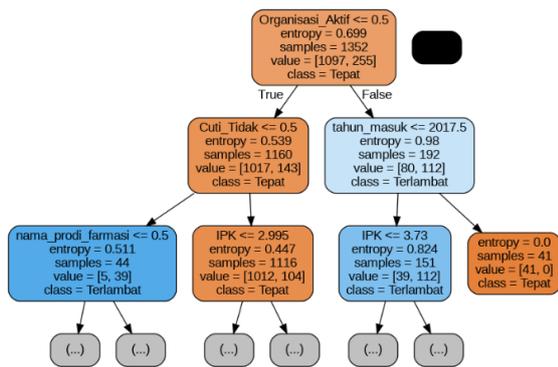
a. Proporsi 20 : 80



Gambar 9 Model 1 proporsi 20:80

Berdasarkan hasil pohon diatas dapat diinterpretasikan sebagai berikut: Jika seseorang tidak aktif dalam organisasi kemungkinan akan lulus tepat waktu, Jika seseorang aktif dalam organisasi tahun lulusnya kurang dari 2017, maka kemungkinan ia akan terlambat. Jika seseorang aktif dalam organisasi tahun lulusnya 2017 atau lebih baru, dan tidak melakukan cuti, maka kemungkinan akan lulus tepat waktu. Jika seseorang aktif dalam organisasi tahun lulusnya 2017 atau lebih baru, tetapi melakukan cuti, maka kemungkinan akan lulus terlambat.

b. Proporsi 30 : 70

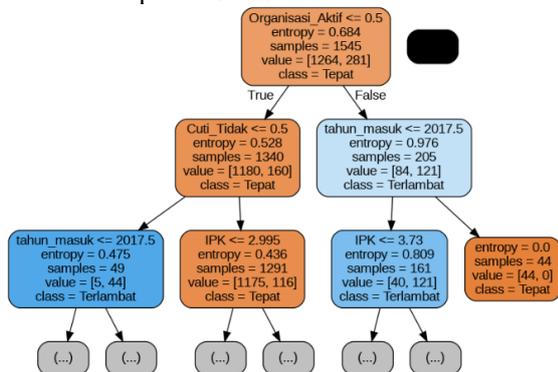


Gambar 10 Model 2 proporsi 30:70

Berdasarkan hasil pohon diatas dapat diinterpretasikan sebagai berikut : Jika seseorang aktif dalam organisasi maka kemungkinan lulus dengan tepat. Jika seseorang aktif dalam organisasi tahun lulusnya kurang dari 2017, maka kemungkinan akan terlambat. Jika seseorang aktif dalam organisasi tahun lulusnya 2017 atau lebih baru, dan tidak melakukan cuti, maka kemungkinan akan lulus tepat. Jika seseorang aktif dalam organisasi tahun lulusnya 2017 atau lebih baru, tetapi melakukan cuti, maka kemungkinan akan lulus tepat.

2. Model 2

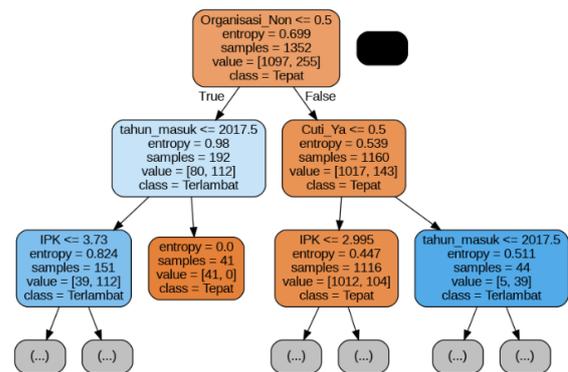
a. Proporsi 20 : 80



Gambar 11 Model 2 proporsi 20:80

Berdasarkan hasil pohon diatas dapat diinterpretasikan sebagai berikut: Jika seseorang tidak aktif dalam organisasi maka kemungkinan lulus tepat. Jika seseorang aktif dalam organisasi tetapi tahun masuknya kurang dari 2017, maka pengamatan tersebut diklasifikasikan sebagai tepat waktu. Jika seseorang aktif dalam organisasi dan tahun masuknya 2017 atau lebih, maka kemungkinan ia lulus terlambat. Jika seseorang aktif dalam organisasi memiliki tahun masuk yang kurang dari 2017, tetapi melakukan cuti, maka kemungkinan akan terlambat. Jika seseorang aktif dalam organisasi memiliki tahun masuk yang kurang dari 2017, dan tidak melakukan cuti, maka kemungkinan akan lulus tepat.

b. Proporsi 30 : 70

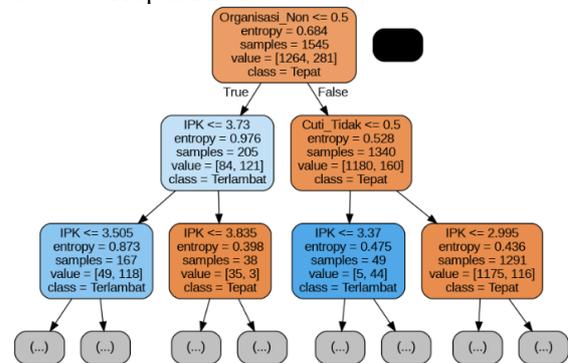


Gambar 12 Model 2 proporsi 30:70

Berdasarkan hasil pohon diatas dapat diinterpretasikan sebagai berikut : Jika seseorang tidak aktif dalam organisasi maka kemungkinan lulus tepat. Jika seseorang aktif dalam organisasi tetapi tahun masuknya kurang dari 2017, maka pengamatan tersebut diklasifikasikan sebagai tepat waktu. Jika seseorang aktif dalam organisasi dan tahun masuknya 2017 atau lebih, maka kemungkinan ia lulus terlambat. Jika seseorang aktif dalam organisasi memiliki tahun masuk yang kurang dari 2017, tetapi melakukan cuti, maka kemungkinan akan terlambat. Jika seseorang aktif dalam organisasi memiliki tahun masuk yang kurang dari 2017, dan tidak melakukan cuti, maka kemungkinan akan lulus tepat waktu.

3. Model 3

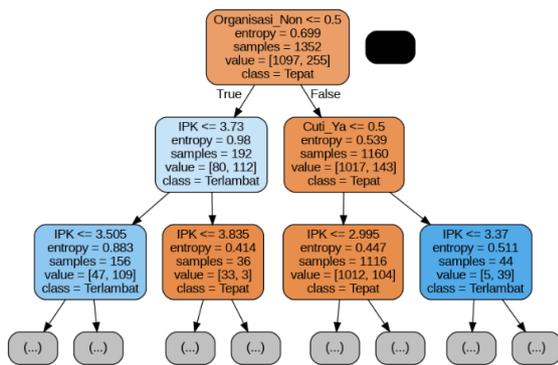
a. Proporsi 20 : 80



Gambar 13 Model 1 proporsi 20:80

Berdasarkan hasil pohon diatas dapat diinterpretasikan sebagai berikut: Jika seseorang aktif dalam organisasi maka kemungkinan akan lulus tepat waktu. Jika seseorang aktif dalam organisasi tetapi memiliki IPK kurang dari 3.73, maka kemungkinan ia akan lulus terlambat. Jika seseorang aktif dalam organisasi memiliki IPK 3.73 atau lebih tinggi, dan tidak melakukan cuti maka kemungkinan akan lulus tepat waktu. Jika seseorang aktif dalam organisasi memiliki IPK 3.73 atau lebih tinggi, tetapi melakukan cuti maka kemungkinan akan terlambat.

b. Proporsi 30 : 70



Gambar 13 Model 2 proporsi 30:70

Dari gambar di atas sampul akhirnya adalah organisasi_non berarti jika seseorang tidak aktif dalam organisasi maka akan lulus dengan tepat. Jika seseorang aktif dalam organisasi, tetapi memiliki IPK kurang dari 3.73, ia kemungkinan akan terlambat. Jika seseorang aktif dalam organisasi, memiliki IPK 3.73 atau lebih tinggi, tetapi melakukan cuti, maka kemungkinan ia akan terlambat. Dan jika seseorang aktif dalam organisasi memiliki IPK 3.73 atau lebih tinggi, dan tidak melakukan cuti, maka kemungkinan ia akan lulus tepat

Setelah pohon keputusan dibuat kemudian melakukan penilaian dan pengukuran pada semua model untuk mengevaluasi dan memahami kinerja model melalui pengukuran akurasi, presisi, recall, dan F1-score pada pohon keputusan (decision tree).

1. Model 1

a. Proporsi 20:80

```

04 print("Akurasi: ", akurasi)
print()
print("Variabel Tepat:")
print("Presisi: ", precision_tepat)
print("Recall: ", recall_tepat)
print("F1-Score: ", f1_score_tepat)
print()

print("Variabel Terlambat:")
print("Presisi: ", precision_terlambat)
print("Recall: ", recall_terlambat)
print("F1-Score: ", f1_score_terlambat)
print()

Akurasi: 0.896640826873385

Variabel Tepat:
Presisi: [0.92941176]
Recall: [0.95180723]
F1-Score: [0.94047619]

Variabel Terlambat:
Presisi: [0.65957447]
Recall: [0.56363636]
F1-Score: [0.60784314]
    
```

Gambar 14 Hasil perhitungan model 1 proporsi 20:80

Model 1 dengan perbandingan data 20:80 memiliki akurasi 0.896640826873385. Model ini memiliki presisi yang lebih tinggi untuk variabel tepat (0.92941176) dibandingkan dengan variabel terlambat (0.65957447). Namun, presisi dan recall untuk variabel terlambat cenderung lebih rendah, menunjukkan bahwa model mungkin memiliki

kesulitan dalam mengklasifikasikan variabel terlambat dengan baik.

b. Proporsi 30 :70

```

04 print("Akurasi: ", akurasi)
print()
print("Variabel Tepat:")
print("Presisi: ", precision_tepat)
print("Recall: ", recall_tepat)
print("F1-Score: ", f1_score_tepat)
print()

print("Variabel Terlambat:")
print("Presisi: ", precision_terlambat)
print("Recall: ", recall_terlambat)
print("F1-Score: ", f1_score_terlambat)
print()

Akurasi: 0.9051724137931034

Variabel Tepat:
Presisi: [0.93700787]
Recall: [0.95390782]
F1-Score: [0.94538232]

Variabel Terlambat:
Presisi: [0.68055556]
Recall: [0.60493827]
F1-Score: [0.64852288]
    
```

Gambar 15 Hasil perhitungan model 1 proporsi 30:70

Model 1 dengan perbandingan data 30:70 memiliki akurasi 0.9051724137931034. Model ini memiliki presisi yang tinggi untuk variabel tepat (0.93700787) dan variabel terlambat (0.68055556). Recall untuk kedua variabel juga cukup baik.

2. Model 2

a. Proporsi 20:80

```

04 print("Akurasi: ", akurasi)
print()
print("Variabel Tepat:")
print("Presisi: ", precision_tepat)
print("Recall: ", recall_tepat)
print("F1-Score: ", f1_score_tepat)
print()

print("Variabel Terlambat:")
print("Presisi: ", precision_terlambat)
print("Recall: ", recall_terlambat)
print("F1-Score: ", f1_score_terlambat)
print()

Akurasi: 0.9069767441860465

Variabel Tepat:
Presisi: [0.94047619]
Recall: [0.95180723]
F1-Score: [0.94610778]

Variabel Terlambat:
Presisi: [0.68627451]
Recall: [0.63636364]
F1-Score: [0.66037736]
    
```

Gambar 16 Hasil perhitungan model 2 proporsi 20:80

Model 2 dengan perbandingan data 20:80 memiliki akurasi 0.9069767441860465. Model ini memiliki presisi yang tinggi untuk variabel tepat (0.94047619) dan variabel terlambat (0.68627451). Recall untuk kedua variabel juga cukup baik.

b. Proporsi 30 :70

```
[231] print("Akurasi: ", akurasi)
print()
print("Variabel Tepat:")
print("Presisi: ", precision_tepat)
print("Recall: ", recall_tepat)
print("F1-Score: ", f1_score_tepat)
print()

print("Variabel Terlambat:")
print("Presisi: ", precision_terlambat)
print("Recall: ", recall_terlambat)
print("F1-Score: ", f1_score_terlambat)
print()

Akurasi: 0.906896551724138

Variabel Tepat:
Presisi: [0.94589178]
Recall: [0.94589178]
F1-Score: [0.94589178]

Variabel Terlambat:
Presisi: [0.66666667]
Recall: [0.66666667]
F1-Score: [0.66666667]
```

Gambar 17 Hasil perhitungan model 2 proporsi 30:70 Model 5 dengan perbandingan data 30:70 memiliki akurasi 0.906896551724138. Model ini memiliki presisi yang tinggi untuk kedua variabel (0.94589178) dan recall yang seimbang.

3. Model 3
a. Proporsi 20:80

```
[231] print("Akurasi: ", akurasi)
print()
print("Variabel Tepat:")
print("Presisi: ", precision_tepat)
print("Recall: ", recall_tepat)
print("F1-Score: ", f1_score_tepat)
print()

print("Variabel Terlambat:")
print("Presisi: ", precision_terlambat)
print("Recall: ", recall_terlambat)
print("F1-Score: ", f1_score_terlambat)
print()

Akurasi: 0.9069767441860465

Variabel Tepat:
Presisi: [0.94589178]
Recall: [0.94589178]
F1-Score: [0.94589178]

Variabel Terlambat:
Presisi: [0.66666667]
Recall: [0.66666667]
F1-Score: [0.66666667]
```

Gambar 18 Hasil perhitungan model 3 proporsi 20:80 Model 3 dengan perbandingan data 20:80 memiliki akurasi 0.9069767441860465. Model ini memiliki presisi yang baik untuk variabel tepat (0.9180791), tetapi presisi untuk variabel terlambat (0.78787879) masih dapat ditingkatkan. Recall untuk kedua variabel juga bisa ditingkatkan.

b. Proporsi 30 :70

```
print("Akurasi: ", akurasi)
print()
print("Variabel Tepat:")
print("Presisi: ", precision_tepat)
print("Recall: ", recall_tepat)
print("F1-Score: ", f1_score_tepat)
print()

print("Variabel Terlambat:")
print("Presisi: ", precision_terlambat)
print("Recall: ", recall_terlambat)
print("F1-Score: ", f1_score_terlambat)
print()

Akurasi: 0.9189655172413793

Variabel Tepat:
Presisi: [0.93129771]
Recall: [0.97795591]
F1-Score: [0.9540567]

Variabel Terlambat:
Presisi: [0.80357143]
Recall: [0.55555556]
F1-Score: [0.65693431]
```

Gambar 19 Hasil perhitungan model 3 proporsi 30:70 Model 6 dengan perbandingan data 30:70 memiliki akurasi 0.9189655172413793. Model ini memiliki presisi yang tinggi untuk variabel tepat (0.93129771) dan variabel terlambat (0.80357143). Recall untuk kedua variabel juga cukup baik.

Selanjutnya melakukan pengujian pada semua model menggunakan salah satu data apakah sesuai atau tidak dengan status yang akan dihasilkan.

1. Model 1 Data yang akan diuji adalah nilai IPK 3.5 gender nya laki-laki pernah mengambil cuti dan tidak aktif organisasi.

```
# Data yang akan diprediksi
data_prediksi = pd.DataFrame({'IPK': [3.5], 'tahun_masuk': [2016],
                               'nama_prodi_agribisnis': [0], 'nama_prodi_agroteknologi': [0],
                               'nama_prodi_akuntansi': [0], 'nama_prodi_farmasi': [0], 'nama_prodi_manajemen': [0],
                               'nama_prodi_pendidikan_bahasa_Inggris': [0], 'nama_prodi_pendidikan_guru_sekolah_dasar': [0],
                               'nama_prodi_peternakan': [0], 'nama_prodi_teknik_informatika': [1], 'nama_prodi_teknik_sipil': [0],
                               'Gender_L': [1], 'Gender_P': [0],
                               'Cuti_Tidak': [0], 'Cuti_Ya': [1], 'Organisasi_Aktif': [0],
                               'Organisasi_Non': [0],
                               })
```

Gambar 20 Pengujian model 1 Apakah status yang dihasilkan akan sesuai pada data atau tidak.

```
prediksi = model.predict(data_prediksi)
print(f"Prediksi Status: {prediksi}")

Prediksi Status: ['Terlambat']
```

Gambar 21 Hasil pengujian model Ternyata data yang di dihasilkan sesuai dimana nilai IPK 3.5 gender nya laki-laki pernah mengambil cuti dan tidak aktif organisasi kelulusannya terlambat, prediksi yang di keluarkan sesuai dengan data.

4.3 Evaluasi

Berdasarkan hasil proses pemodelan di atas dapat di lihat pada tabel berikut :

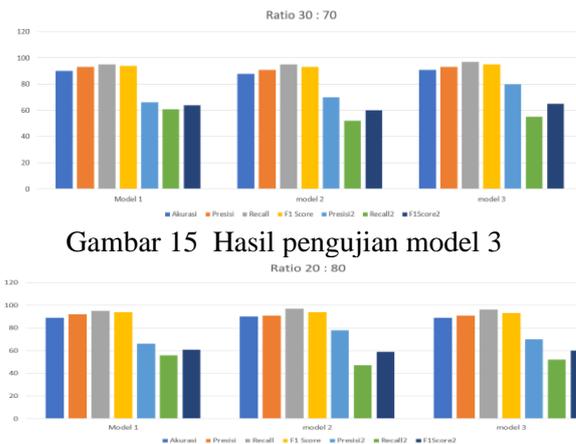
Tabel 1 Hasil Evaluasi

No	Model	Hasil pohon keputusan	Metrik evaluasi
1	Model 1 (rasio 80:20)	Sampul akar : Non Organisasi Cabang kiri :	Akurasi: 89,66% Variabel Tepat:

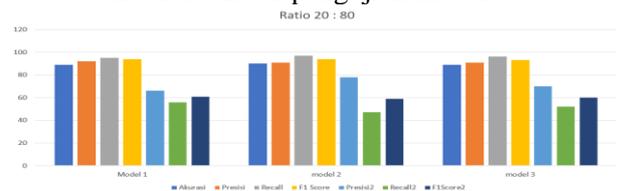
No	Model	Hasil pohon keputusan	Metrik evaluasi
		Data : tahun masuk <= 2017 Nama prodi Gender IPK Tahun masuk Cuti Aktif organisasi Kelulusan	Presisi: 92 % Recall: 95 % F1-Score: 0.94%
2	Model 1 (rasio 70:30)	Sampul akar : Aktif Organisasi Cabang kiri : a. Tidak cuti , b. IPK <=2.95, c. prodi farmasi Cabang kanan : a. tahun masuk <= 2017 b. IPK <=7.37 c. IPK <=3.505, d. prodi manajemen	Akurasi: 90.34% Variabel Tepat: Presisi: 93.86% Recall: 94.99% F1-Score: 94.42% Variabel Terlambat: Presisi : 66.67% Recall: 61.73% F1-Score: 64.10%
3	Model 2 (rasio 80:20)	Sampul akar : Non Organisasi Cabang kiri : a. tahun masuk <= 2017 b. IPK <=7.37 c. IPK <=3.505 Cabang kanan : a. Ya cuti b. IPK <=2.95	Akurasi: 90.70% Variabel Tepat Presisi: 91.81% Recall: 97.89% F1-Score: 94.75% Variabel Terlambat Presisi: 78.79% Recall: 47.27% F1-Score: 59.09%
4	Model 2 (rasio 70:30)	Sampul akar : Aktif Organisasi Cabang kiri : a. Tidak cuti , b. IPK <=2.95 Cabang kanan : a. tahun masuk <= 2017 b. IPK <=7.37 c. IPK <=3.505	Akurasi: 88.62% Variabel Tepat Presisi: 91.16% Recall : 95.67% F1-Score: 93.36% Variabel Terlambat Presisi: 70.42% Recall: 52.63% F1-Score: 60.24%
5	Model 3 (rasio 80:20) :	Sampul akar : Non Organisasi Cabang kiri : a. IPK <=7.37 b. IPK <=3.505 Cabang kanan : a. Ya cuti	Akurasi: 89.41% Variabel Tepat Presisi: 91.84% Recall: 96.04% F1-Score: 93.89% Variabel Terlambat

No	Model	Hasil pohon keputusan	Metrik evaluasi
		Kelulusan b. IPK <=2.95	Presisi: 70.45% Recall: 52.54% F1-Score: 60.19%
6	Model 3 (rasio 70:30)	Sampul akar : Aktif Organisasi Cabang kiri : a. Tidak cuti , b. IPK <=2.95 Data : Gender IPK Cuti Aktif organisasi Kelulusan Cabang kanan : a. IPK <=7.37 b. IPK <=3.505	Akurasi: 91.90% Variabel Tepat Presisi: 93.13% Recall: 97.80% F1-Score: 95.41% Variabel Terlambat Presisi: 80.36% Recall: 55.56% F1-Score: 65.69%

Dari Tabel di atas parameter yang paling berpengaruh dari semua model adalah aktif atau tidaknya organisasi, setelahnya adalah dia pernah melakukan cuti atau tidak serta nilai IPK walaupun saat menambahkan data nama prodi dan tahun masuk node selanjutnya adalah nama prodi atau tahun masuk namun setelahnya adalah nilai IPK lebih dari sama dengan 37,3.



Gambar 15 Hasil pengujian model 3



Gambar 16 Hasil pengujian model 3

Berdasarkan hasil tersebut, model yang menonjol adalah Model 3 dengan proporsi rasio data 30:70. Model ini memiliki akurasi yang baik dan mencapai presisi yang tinggi serta recall yang baik untuk kedua variabel

5. KESIMPULAN

Berdasarkan hasil melakukan pemodelan klasifikasi menggunakan algoritma *c.45* dengan menggunakan 3 model data dengan 2 proporsi rasio pengukuran dapat disimpulkan atribut yang dapat mempengaruhi kelulusan di universitas perjuangan adalah organisasi, cuti

dan IPK. Model klasifikasi algoritma c.45 yang di bangun pada penelitian ini menghasilkan nilai akurasi terbaik 91% dengan nilai *presisi* tepat 93% dan *presisi* 80% terlambat nilai *recall* tepat 98% dan *recall* terlambat 55% dan *F1-Score* tepat 95% dan *F1-Score* terlambat 66%. Berdasarkan hasil pengujian terhadap data yang di uji memunculkan hasil yang sesuai dengan lab .

Penelitian akan lebih maksimal jika menambahkan lebih banyak data kelulusan agar tingkat akurasi dapat lebih maksimal dan dapat menambahkan atribut lain yang dapat menunjang pemodelan klasifikasi yang lebih baik

6. UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam penelitian ini. Dukungan Anda sangat berharga bagi keberhasilan penelitian ini.

DAFTAR PUSTAKA

- [1] I. Kurniawati, R. E. Indrajit, and M. Fauzi, "Peran Bussines Intelligence Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru," *Ikraith-Informatika*, vol. 1, no. 2, pp. 70–79, 2017.
- [2] S. A. Aradea, Ariyan Z, and Yuliana A, "Penerapan Decision Tree Untuk Penentuan Pola Data Penerimaan Mahasiswa Baru," *Jurnal Penelitian Sitrotika*, vol. 7, no. 1, pp. 1693–9670, 2011.
- [3] W. E. Nugroho, T. Prihandoyo, and O. Somantri, "Optimalisasi Metode Naive Bayes untuk Menentukan Program Studi bagi Calon Mahasiswa Baru dengan Pendekatan Unsupervised Discretization," *Infotekmesin*, vol. 13, no. 1, pp. 161–167, 2022, doi: 10.35970/infotekmesin.v13i1.1048.
- [4] F. A. D. Aji Prasetya Wibawa, Muhammad Guntur Aji Purnama, Muhammad Fathony Akbar, "Metode-metode Klasifikasi," *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 1, p. 134, 2018.
- [5] M. S. Simanjuntak *et al.*, "Analisis Decision Tree Pada Menentukan Lulus," vol. 2, no. 1, pp. 34–42.
- [6] S. Shedriko and M. Firdaus, "Penentuan Klasifikasi Dengan Crisp-Dm Dalam Memprediksi Kelulusan Mahasiswa Pada Suatu Mata Kuliah," *Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi)*, vol. 6, no. 1, pp. 826–831, 2022, doi: 10.30998/semnasristek.v6i1.5814.
- [7] Y. T. Samuel and B. Jonathan, "Memprediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree J48 di Universitas Advent Indonesia Predicting Timely Students Graduation Using the Decision Tree J48 Method at Universitas Advent Indonesia," *Telka*, vol. Vol 9 No 1, pp. 43–52, 2019.
- [8] C. N. Dengen, K. Kusriani, and E. T. Luthfi, "Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *Sisfotenika*, vol. 10, no. 1, p. 1, 2020, doi: 10.30700/jst.v10i1.484.
- [9] A. F. A. Naibaho and A. Zahra, "Prediksi Kelulusan Siswa Sekolah Menengah Pertama Menggunakan Machine Learning," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, Jul. 2023, doi: 10.23960/jitet.v11i3.3056.
- [10] Nugroho, W. E., Prihandoyo, T., & Somantri, O. (2022). Optimalisasi Metode Naive Bayes untuk Menentukan Program Studi bagi Calon Mahasiswa Baru dengan Pendekatan Unsupervised Discretization. *Infotekmesin*, 13(1), 161–167. <https://doi.org/10.35970/infotekmesin.v13i1.1048>
- [11] Mustika *et al.*, *Data Mining dan Aplikasinya*. 2021
- [12] Simanjuntak, M. S., Simangunsong, J., Matondang, A. A., Kelamin, J., Daerah, A., Sekolah, A., Tree, D., & Waktu, T. (n.d.). *Analisis Decision Tree Pada Menentukan Lulus*. 2(1), 34–42.