



Klasifikasi Mahasiswa Berpotensi Putus Studi Menggunakan Algoritma Decision Tree pada Fakultas Teknik Unismuh Makassar

<u>INFO PENULIS</u>	<u>INFO ARTIKEL</u>
Yumi Universitas Muhammadiyah Makassar 105841102120@student.unismuh.ac.id	ISSN: 3026-3603 Vol. 2, No. 2 Oktober 2024 http://jurnal.ardenjaya.com/index.php/ajst
Lukman Universitas Muhammadiyah Makassar	
Rizki Yusliana Bakti Universitas Muhammadiyah Makassar	

© 2024 Arden Jaya Publisher All rights reserved

Saran Penulisan Referensi:

Yumi, Lukman, & Bakti, R. Y. (2024). Klasifikasi Mahasiswa Berpotensi Putus Studi Menggunakan Algoritma Decision Tree pada Fakultas Teknik Unismuh Makassar. *Arus Jurnal Sains dan Teknologi*, 2 (2), 532-542.

Abstrak

Universitas Muhammadiyah Makassar (Unismuh Makassar) menghadapi tantangan signifikan dalam menangani masalah mahasiswa yang berpotensi putus studi, terutama di Fakultas Teknik. Faktor-faktor seperti rendahnya kemampuan akademik, keterbatasan biaya, dan kendala tempat tinggal menjadi pemicu utama masalah ini, yang pada gilirannya dapat menghambat kemajuan perguruan tinggi. Penelitian ini bertujuan untuk mengklasifikasikan mahasiswa yang berpotensi putus studi menggunakan Algoritma Decision Tree C4.5 dan mengevaluasi tingkat akurasi sistem klasifikasi tersebut. Penelitian ini menggunakan metode kuantitatif dengan analisis statistik deskriptif. Dataset yang dianalisis terdiri dari 5657 mahasiswa Fakultas Teknik Unismuh Makassar, dengan atribut-atribut seperti pekerjaan dan penghasilan orang tua/wali, IPK, SKS, dan variabel lainnya yang relevan dengan status akademik mahasiswa. Hasil penelitian menunjukkan bahwa Algoritma Decision Tree C4.5 mampu mengklasifikasikan mahasiswa yang berpotensi putus studi dengan akurasi sebesar 100%. Nilai rata-rata precision dan recall masing-masing adalah 100%, sedangkan nilai rata-rata f1-score mencapai 100%. Temuan ini mengindikasikan bahwa Algoritma Decision Tree C4.5 memiliki performa yang tinggi dan merupakan metode yang efektif dalam mengidentifikasi mahasiswa dengan risiko putus studi, sehingga memungkinkan perguruan tinggi untuk mengambil langkah-langkah preventif yang lebih tepat sasaran.

Kata kunci : Putus Studi, Algoritma Decision Tree, Klasifikasi, Fakultas Teknik, Universitas Muhammadiyah Makassar

Abstract

Muhammadiyah University of Makassar (Unismuh Makassar) faces significant challenges in addressing the issue of students at risk of dropping out, particularly in the Faculty of Engineering. Factors such as low academic performance, financial constraints, and housing issues are the main triggers of this problem, which in turn can hinder the progress of the university. This study aims to classify students at risk of dropping out using the C4.5 Decision Tree Algorithm and evaluate the accuracy level of the classification system. This research employs a quantitative method with descriptive statistical analysis. The analyzed dataset consists of 5657 students from the Faculty of Engineering at Unismuh Makassar, with attributes including the occupations and income of parents/guardians, IPK, SKS, and other variables relevant to students' academic status. The results indicate that the C4.5 Decision Tree Algorithm can classify students at risk of dropping out with an accuracy of 100%. The average precision and recall values are both 100%, while the average F1-score reaches 100%. These findings suggest that the C4.5 Decision Tree Algorithm has high performance and is an effective method for identifying students at risk of dropping out, enabling universities to take more targeted preventive measures.

Keywords: Dropout, Decision Tree Algorithm, Classification, Faculty Of Engineering, Muhammadiyah University Of Makassar.

A. Pendahuluan

Universitas Muhammadiyah Makassar atau dikenal juga dengan sebutan Unismuh Makassar didirikan pada tanggal 19 Juni 1963 sebagai cabang dari Universitas Muhammadiyah Jakarta. Sejak berdirinya hingga saat ini, Unismuh Makassar telah meluluskan alumni sebanyak lebih dari 15.000 orang. Salah satu fakultas yang ada di Universitas Muhammadiyah Makassar yaitu Fakultas Teknik. Dimana dalam Fakultas Teknik Muhammadiyah Makassar terdiri dari beberapa program studi yaitu Teknik Pengairan, Teknik Elektro, Arsitektur, Informatika dan Perencanaan Wilayah dan Kota (PWK).

Apabila ada beberapa mahasiswa yang terlambat lulus atau tidak tepat pada waktunya sehingga menjadi kendala untuk kemajuan dari perguruan tinggi tersebut. Salah satu masalah yang dihadapi perguruan tinggi adalah terdapat mahasiswa yang Putus Studi. Kasus putus studi juga bisa terjadi karena beberapa faktor seperti rendahnya kemampuan akademik, faktor biaya dan tempat tinggal saat menempuh pendidikan. Putus studi termasuk masalah yang serius dan merupakan situasi yang pernah dihadapi oleh sebagian mahasiswa di perguruan tinggi (Samasil et al., 2022). Untuk mengidentifikasi mahasiswa yang berpotensi putus studi saat ini dilakukan secara manual. Proses ini melibatkan pencacatan nilai, pembuatan tabel, dan perhitungan manual, yang membuatnya rumit dan memakan waktu. Ketidakpraktisan ini menjadi salah satu alasan mengapa sistem otomatis kemudian dikembangkan.

Untuk mencegah putus studi mahasiswa, perguruan tinggi dapat mengambil langkah-langkah preventif dan efektif. Salah satu strategi yang dapat diterapkan adalah dengan melakukan prediksi mahasiswa yang berpotensi untuk putus studi, dengan demikian, pihak kampus dapat mengambil langkah-langkah pencegahan sejak awal dan mengurangi kemungkinan terjadinya kasus putus studi.

Decision Tree merupakan sebuah algoritma yang termasuk pada teknik klasifikasi pada data mining. Klasifikasi adalah proses menemukan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya, untuk dapat digunakan memprediksi data yang belum memiliki kelas data tertentu. Data mining merupakan sebuah teknik yang dipergunakan untuk menyelesaikan permasalahan pada proses data yang besar. Proses pada decision tree bertujuan untuk mengelompokkan atau mengklasifikasi pada data terhadap kelasnya masing-masing. Pada decision tree hasil dari proses yang dilakukan berupa sebuah pohon keputusan dengan rule yang didapatkan dari pohon keputusan tersebut. Pohon keputusan pada decision tree terbentuk berdasarkan dengan perhitungan nilai gain dan entropy. Atribut yang memiliki nilai gain tersebut yang didapatkan dari proses perhitungan pada pengolahan data (Arfyanti et al., 2022)

Kelebihan yang dimiliki Decision Tree adalah sifatnya yang fleksibel sehingga mampu meningkatkan kualitas keputusan yang dihasilkan, sedangkan kekurangan dari algoritma ini adalah akan terjadi overlap jika menggunakan data yang memiliki kelas dan kriteria dengan jumlah yang sangat banyak (Rahmadayan and Mustakim 2023).

Penelitian terkait permasalahan putus studi banyak dilakukan para peneliti sebelumnya seperti Kotsiantis, Pierrakeas dan Pintelas menyebutkan bahwa sangat penting bagi dosen untuk mendeteksi mahasiswa yang cenderung putus studi sebelum mereka memasuki pertengahan masa studi. Gerben W Dekker menyebutkan bahwa monitoring dan dukungan terhadap mahasiswa di tahun pertama sangat penting dilakukan (Hermanto, 2020).

B. Metodologi

1) Klasifikasi

Klasifikasi adalah proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya, untuk dapat digunakan dalam memprediksi data yang belum memiliki kelas data tertentu (Nasrullah, 2021).

2) Mahasiswa Putus Studi

Mahasiswa merupakan salah satu aspek penting dalam menentukan keberhasilan penyelenggaraan program studi. Mahasiswa yang selesai tepat waktu merupakan salah satu elemen penilaian akreditasi Universitas. Tetapi tidak semua mahasiswa bisa selesai tepat waktu sesuai jangka waktu yang telah ditentukan, tidak sedikit mahasiswa yang menempuh studi S1 lebih dari batas waktu maksimal yang telah ditentukan atau bahkan sampai terancam putus studi (Paskalis et al., 2019).

3) Decision Tree

Algoritma Decision Tree merupakan algoritma yang umum digunakan untuk pengambilan keputusan. Decision tree akan mencari solusi permasalahan dengan menjadikan kriteria sebagai node yang saling berhubungan membentuk seperti struktur pohon. Decision tree adalah model prediksi terhadap suatu keputusan menggunakan struktur hirarki atau pohon. Setiap pohon memiliki cabang, cabang mewakili suatu atribut yang harus dipenuhi untuk menuju cabang yang selanjutnya hingga berakhir di daun (tidak ada cabang lagi) (Sartika & Sensuse, 2017).

4) Cross Validation

Cross validation adalah sebuah tindakan pembuktian dari sebuah metode atau performa suatu algoritma. Cross validation merupakan pembuktian dengan membagi data menjadi data training dan data testing dengan komposisi tertentu. Pembagian paling sering digunakan dalam penelitian klasifikasi data mining ialah dengan membagi data secara acak menjadi 10 bagian. Satu bagian sebagai data testing dan 9 bagian sebagai data training. Validasi ini disebut juga dengan k-fold cross validation dengan k=10.

K-fold cross validation adalah salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak. Setiap kelas pada kelompok data harus diwakili dalam proporsi yang tepat antara data training dan data testing. Data dibagi secara acak pada masing-masing kelas dengan perbandingan yang sama. Untuk mengurangi bias yang disebabkan oleh sampel tertentu, seluruh proses training dan testing diulangi beberapa kali dengan sampel yang berbeda. Tingkat kesalahan pada iterasi yang berbeda akan dihitung rata-ratanya untuk menghasilkan error rate secara keseluruhan. Model yang memberikan kesalahan terkecil adalah model yang terbaik (Salmawati, Yuyun, 2021)

5) Confusion matrix

Confusion matrix adalah sebuah hasil evaluasi dari sebuah klasifikasi data mining yang diwujudkan dalam sebuah tabel. Confusion matrix berisi tentang perhitungan jumlah objek data testing yang diprediksikan ke dalam sebuah kelas dengan klasifikasi yang sebenarnya. Confusion matrix menggunakan tabel matriks yang mana jika dataset hanya terdiri dari dua kelas maka kelas tersebut dianggap sebagai positif negatif.

Tabel 1. Confusion matrix

<i>Correct Classification</i>	<i>Classified as</i>	
	<i>+</i>	<i>-</i>
<i>+</i>	<i>True Positives</i>	<i>False Negative</i>
<i>-</i>	<i>False Positive</i>	<i>True Negative</i>

Perhitungan performance dalam confusion matrix menggunakan persamaan sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \dots\dots\dots(1)$$

$$Precision = \frac{TP}{FP+TP} \times 100\% \dots\dots\dots(2)$$

$$Recall = \frac{TP}{FN+TP} \times 100\% \dots\dots\dots(3)$$

Keterangan:

TP = True Positive (data positif yang terklasifikasi benar positif)

TN = True Negative (data negatif yang terklasifikasi benar negatif)

FP = False Positive (data positif yang terklasifikasi negatif)

FN = False Negative (data negatif yang terklasifikasi positif)

Selain menggunakan accuracy, precision dan recall, kinerja dari algoritma juga dapat diukur menggunakan F-Measure (F1-Score) merupakan harmonic mean dari precision dan recall, yang mana memiliki persamaan sebagai berikut (Salmawati, Yuyun, 2021):

$$F1 - Score = 2 \times \frac{Precision \times recall}{Precision + Recall} \dots\dots\dots(4)$$

C. Hasil dan Pembahasan

1. Pengumpulan Data

Data yang dikumpulkan merupakan data dari Fakultas Teknik Universitas Muhammadiyah Makassar dengan jumlah data sebesar 1493 data. Data yang diperoleh merupakan data mahasiswa program S1 angkatan 2013-2014. Data dalam penelitian ini merupakan data awal sebelum melakukan praproses data yang mana terdiri dari 18 atribut. Pengumpulan data mentah dikumpulkan dari server GraphQL secara asinkron untuk beberapa mahasiswa berdasarkan nim mereka. Proses pengumpulan data ini dilakukan dalam beberapa langkah, yang memanfaatkan concurrency untuk mempercepat pengambilan data.

2. Data Mentah

Data yang diolah dalam penelitian ini berasal data yang sudah dikumpulkan sebelumnya dari mahasiswa Fakultas Teknik Universitas Muhammadiyah Makassar

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	kodeProdi	angkatan	semester	nim	nama	jenis_kelamin	tahun_akad	tanggal_lahir	lulus	masa_studi	ayah_pekerja	ayah_pencayah	status_ibu_pekerja	ibu_penghidup	status_wali_pekerja	wali_per	
2	22202	2013	20131	105811821	MUHAMM L		20182		False	5 Tahun, 1	TANI	500000	HIDUP	-	500000	HIDUP	
3	22202	2013	20131	105811822	MUHAMM L				False								
4	22202	2013	20131	105811821	FAISAL AIL				True	4 Tahun, 8	WIRASW/500001-1	CHIDUP	WIRASW/500001-1	CHIDUP			
5	22202	2013	20131	105811821	MUHAMM L		20172		True	6 Tahun, 2	WIRASW/500001-1	CHIDUP	WIRASW/500000		HIDUP		
6	22202	2013	20131	105811822	ST FATIMP				True								
7	22202	2013	20131	105811821	SAIFUL R L				False								
8	22202	2013	20131	105811821	JUPRI L		20182		False	5 Tahun, 9	WIRASW/2000001-3	HIDUP	IRT	500000	HIDUP		
9	22202	2013	20131	105811831	HERA A L				True								500000
10	22202	2013	20131	105811831	MOHAMM L				False		TANI	500000	MENINGE-	500001-1	CHIDUP		
11	22202	2013	20131	105811831	HILMAN A L				False		TANI	500001-1	CHIDUP	WIRASW/500001-1	CHIDUP		
12	22202	2013	20131	105811831	ILHAM L				False		TANI	500001-1	CHIDUP	URT	500001-1	CHIDUP	
13	22202	2013	20131	105811831	FIRDAUS L				False								
14	22202	2013	20131	105811831	SUPARDI L				False								
15	22202	2013	20131	105811831	NURHIDAP		20172		True	4 Tahun, 1	TANI	2000001-3	HIDUP	-	500000	HIDUP	
16	22202	2013	20131	105811831	GAPARU L		20182		True	5 Tahun, 1	TANI	500000	HIDUP	IBU RUM/500000	HIDUP		
17	22202	2013	20131	105811831	JARTIA P		20172		True	4 Tahun, 8	Nelayan	500001-1	CHIDUP	WIRASW/500000	HIDUP		WIRASWASTA
18	22202	2013	20131	105811831	ARIZAL A L				False								
19	22202	2013	20131	105811841	ASMAUL P		20172		True	4 Tahun, 9	PNS	3000001-5	HIDUP	IRT	500000	HIDUP	PNS
20	22202	2013	20131	105811841	NINING H P		20172		True	4 Tahun, 1	TANI	500000	HIDUP	TANI	500000	HIDUP	
21	22202	2013	20131	105811841	MUR ILHAF		20182		True	5 Tahun, 9	TANI	500001-1	CHIDUP	-	500000	HIDUP	
22	22202	2013	20131	105811841	HERMAWL				False								
23	22202	2013	20131	105811841	USMANS L		20182		True	5 Tahun, 9	TANI	500001-1	CHIDUP	IRT	500001-1	CHIDUP	TANI
24	22202	2013	20131	105811841	SYAFRI Z L		20182		True	5 Tahun, 9	WIRASW/1500001-2	HIDUP	-	500000	HIDUP		150000

Gambar 5. Data mentah

3. Data Preprocessing

Preprocessing data merupakan proses mempersiapkan data sebelum dilakukannya proses klasifikasi. Preprocessing data dalam penelitian ini terdiri dari data selection, data cleaning, dan data transformation.

1) Data Selection

Data selection atau seleksi data merupakan proses pemilihan data yang benar-benar diperlukan dan sesuai untuk proses klasifikasi, yang mana dalam proses ini akan dilakukan beberapa hal diantaranya:

- a. Menambahkan Atribut IPK Terakhir, SKS Lulus dan Semester

Pada atribut IPK, SKS dan semester data yang dimiliki terdiri dari banyak variasi atribut data sehingga perlu dikelompokkan menjadi satu atribut. Berikut tambahan atribut data yang dibuat untuk mengelompokkan data dari SKS, IPK dan Semester.

	SKS_LULUS	IPK_TERAKHIR	semesterTerakhir
0	60.0	2.90	8
1	156.0	3.08	12
2	22.0	2.91	3
3	30.0	2.80	3
4	156.0	3.18	10

Gambar 6. Hasil Penambahan Atribut

b. Menentukan Atribut Target

Dalam penelitian ini akan ditentukan atribut target yaitu status putus studi yang dikelompokkan kedalam dua *class* yaitu Potensial Putus Studi dan Tidak Potensial. Data yang digunakan dalam penelitian ini untuk menentukan atribut status putus studi ialah semua variable yang telah di seleksi yaitu ayah_pekerjaan, ayah_penghasilan, ayah_status, ibu_pekerjaan, ibu_penghasilan, ibu_status, wali_pekerjaan, wali_penghasilan, lulus, semester Terakhir, kategori_IPK, kategori_SKS .

c. Mengkategorikan Atribut IPK Terakhir dan SKS Lulus

Pada atribut IPK terakhir dan SKS lulus data yang dimiliki terdiri dari banyak variasi data sehingga perlu dikelompokkan kedalam beberapa kategori. Berikut kategori data pada atribut ipk terakhir dan sks lulus:

	IPK_TERAKHIR	kategori_IPK	SKS_LULUS	semesterTerakhir	kategori_SKS
0	2.90	Memuaskan	60.0	8	Tidak Memenuhi
1	3.08	Memuaskan	156.0	12	Memenuhi
2	2.91	Memuaskan	22.0	3	Tidak Memenuhi
3	2.80	Memuaskan	30.0	3	Tidak Memenuhi
4	3.18	Memuaskan	156.0	10	Memenuhi

Gambar 8. Hasil kategori

d. Seleksi Atribut

Atribut yang digunakan untuk proses klasifikasi hanya terdiri dari beberapa atribut yaitu yaitu angkatan, semester Awal, NIM, nama, jenis_Kelamin, ayah_pekerjaan, ayah_penghasilan, ayah_status, ibu_pekerjaan, ibu_penghasilan, ibu_status, wali_pekerjaan, wali_penghasilan, lulus, semester_Terakhir, kategori_IPK, kategori_SKS dan status_putus_studi.

e. Menghapus Atribut Nim, Nama, Jenis Kelamin

Pada penelitian ini, atribut nim, nama dan jenis kelamin tidak digunakan dalam proses klasifikasi. Atribut nim, nama, jenis kelamin hanya digunakan untuk menjelaskan identitas mahasiswa sehingga atribut tersebut akan dihapus. Atribut nim, nama dan jenis kelamin akan dihapus menggunakan kode program

	angkatan	semesterAwal	ayah_pekerjaan	ayah_penghasilan	ayah_status	ibu_pekerjaan	ibu_penghasilan	ibu_status	wali_pekerjaan
0	2013	20131	TANI	500000	HIDUP	-	500000	HIDUP	WIRASWASTA
1	2013	20131	TANI	500000	HIDUP	-	500000	HIDUP	WIRASWASTA
2	2013	20131	TANI	500000	HIDUP	-	500000	HIDUP	WIRASWASTA
3	2013	20131	TANI	500000	HIDUP	-	500000	HIDUP	WIRASWASTA
4	2013	20131	WIRASWASTA	500001-1000000	HIDUP	WIRASWASTA	500001-1000000	HIDUP	WIRASWASTA

2) Data Cleaning

Data cleaning atau pembersihan data pada penelitian ini dilakukan dengan menghapus data-data missing value. Data missing value merupakan data yang bernilai kosong (null) yang tidak dapat digunakan dalam penelitian, sehingga didalam penelitian saya, saya mengisi data yang missing value dengan nilai mean atau rata-rata. Jumlah data missing value pada tiap atribut dapat dilihat pada tabel berikut:

Atribut	Jumlah Data Missing Value
angkatan	0
semesterAwal	0
ayah_pekerjaan	410
ayah_penghasilan	421
ayah_status	416
ibu_pekerjaan	411
ibu_penghasilan	428
ibu_status	416
wali_pekerjaan	1266
wali_penghasilan	1298
lulus	0
semesterTerakhir	0
kategori_IPK	4
kategori_SKS	0

Jumlah *missing value* diatas merupakan jumlah data kosong pada setiap kolom atribut, dalam satu baris data terdapat lebih dari satu atribut yang memiliki data *missing value*. Dari 14 atribut yang digunakan, 9 diantaranya memiliki data *missing value*. Berikut contoh data *missing value*:

ayah_status	ibu_pekerjaan	ibu_penghasilan	ibu_status	wali_pekerjaan	wali_penghasilan	lulus	semesterTerakhir
HIDUP	WIRASWASTA	3000001-5000000	HIDUP	NaN	500000	False	6
HIDUP	PNS	1500001-2000000	HIDUP	NaN	500000	True	9
HIDUP	PENSUIN	1000001-1500000	HIDUP	SWASTA	2000001-3000000	False	15
MENINGGAL	PNS	3000001-5000000	HIDUP	NaN	500000	True	13
MENINGGAL	-	500000	HIDUP	NaN	500000	True	14

Atribut	Jumlah Data Missing Value
angkatan	0
semesterAwal	0
ayah_pekerjaan	0
ayah_penghasilan	0
ayah_status	0
ibu_pekerjaan	0
ibu_penghasilan	0
ibu_status	0
wali_pekerjaan	0
wali_penghasilan	0
lulus	0
semesterTerakhir	0
kategori_IPK	0
kategori_SKS	0

Dari gambar diatas dapat dilihat bahwa tidak ada lagi data *missing value* atau data yang kosong. Jumlah data pada proses pembersihan data dapat dilihat pada tabel berikut:

Tabel 7. Jumlah Data Hasil Penghapusan

	Jumlah Data
Data awal	1493
Data <i>missing</i>	5070
Data bersih	1493

Jumlah data bersih yang diperoleh setelah mengatasi data *missing value* berjumlah 1493 data. Setelah semua data bersih, selanjutnya melakukan transformasi data agar sesuai untuk proses data mining.

3) Data Transformation

Data dalam penelitian ini terdiri dari data kategorikal dan numerial sehingga perlu diubah kedalam bentuk data yang sama. Data akan diubah menggunakan LabelEncoder. Data yang diubah yaitu ayah_pekerjaan, ayah_penghasilan, ayah_status, ibu_pekerjaan, ibu_penghasilan, ibu_status, wali_pekerjaan, wali_penghasilan, lulus, kategori_IPK, kategori_SKS dan status_putus_studi. Berikut hasil pelabelannya:

	angkatan	semesterAwal	ayah_pekerjaan	ayah_penghasilan	ayah_status	ibu_pekerjaan	ibu_penghasilan	ibu_status	wali_pekerjaan	wali_penghasilan	Lulus	semesterTerakhir	kategori_IPK	kategori_SKS	status_potensi_putus_studi
0	2013	20131	2	0	0	5	0	0	0	0	0	8	2	1	0
1	2013	20131	2	0	0	5	0	0	0	0	1	12	2	0	1
2	2013	20131	2	0	0	5	0	0	0	0	0	3	2	1	0
3	2013	20131	2	0	0	5	0	0	0	0	0	3	2	1	0
4	2013	20131	0	0	0	0	0	0	0	0	1	10	2	0	1

4. Pembagian Data

Dalam penelitian ini data akan dibagi dengan proporsi 80% data *training* dan 20% data *testing*. Berikut jumlah data setelah dilakukan pembagian:

Tabel 20. Pembagian Data

Klasifikasi	Jumlah Data	Data Training (80%)	Data Testing (20%)
Potensial Putus Studi	771	612	159
Tidak Potensial	722	582	140
Total	1493	1194	299

Dari tabel diatas diketahui bahwa jumlah data *training* yang dihasilkan adalah 1194 data dan data *testing* berjumlah 299 data.

5. Klasifikasi Menggunakan Algoritma Decision Tree C4.5

Setelah data dibagi menjadi data *training* dan data *testing*, tahap selanjutnya ialah proses klasifikasi data. Pada tahap klasifikasi data, akan dilakukan pembangunan model menggunakan *algoritma decision tree C4.5*. pada proses ini data *training* akan mempelajari pola dari algoritma sehingga didapat model yang dihasilkan. Setelah itu dapat dilihat dari performa dari algoritma yang telah dilatih berdasarkan akurasi yang dihasilkan dari data *testing* atau data uji. Berikut hasil akurasi dari *algoritma decision tree C4.5*:

Tabel 21. Hasil Akurasi

Algoritma	Akurasi(%)
<i>Decision Tree C4.5</i>	100%

6. Pengujian Menggunakan Cross Validation

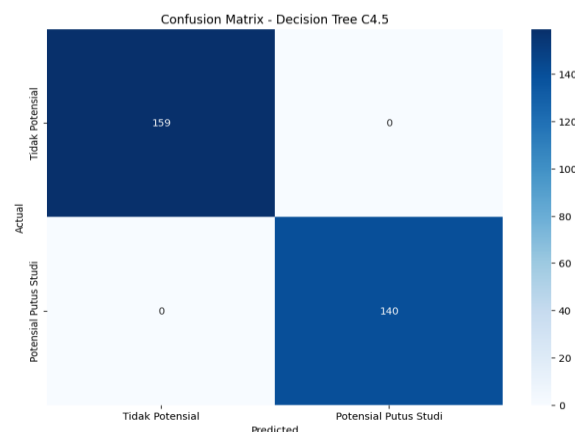
Setelah didapat hasil akurasi dari proses klasifikasi, selanjutnya ialah melakukan pengujian menggunakan *cross validation*. Jenis pengujian *cross validation* yang digunakan ialah *k-fold cross validation* dimana $k=10$ yang berarti data akan dibagi menjadi 10 bagian dengan 9 bagian sebagai data *training* dan 1 bagian sebagai data *testing* (*10-fold cross validation*). *K-fold cross validation* adalah salah satu metode dengan melakukan perulangan dengan mengacak atribut masukkan sehingga sistem teruji untuk beberapa atribut input yang acak. Berikut ini hasil akurasi dari *algoritma decision tree C4.5* menggunakan pengujian *10-fold cross validation*.

<i>k-Fold</i>	Akurasi(%)
1	99,33%
2	100%
3	100%
4	99,33%
5	100%
6	99,33%
7	99,33%
8	100%
9	99,33%
10	100%
Rata-Rata	99,66%

Berdasarkan hasil dari pengujian diatas dapat dilihat bahwa akurasi *algoritma decision tree C4.5* mengalami penurunan dari hasil akurasi sebelum dilakukan pengujian walau tidak terlalu signifikan. Rata-rata akurasi yang didapat ialah 99,66%.

7. Evaluasi Hasil Klasifikasi

Evaluasi hasil klasifikasi pada *algoritma decision tree* juga menggunakan metode *confussion matrix*, hasil evaluasi dapat dilihat pada gambar berikut:



Lebih jelasnya dapat dilihat pada tabel berikut:

Data Sebenarnya	Klasifikasi	
	Tidak Potensial	Potensial Putus Studi
Tidak Potensial	159 (TP)	0 (FN)
Potensial Putus Studi	0 (FP)	140 (TN)

Dari hasil klasifikasi pada tabel *confussion matrix* di atas dapat dilihat bahwa,

- True Positive* (TP) menjelaskan dimana data terklasifikasi Tidak Potensial, memang benar Tidak Potensial. Dalam hal ini jumlah data yang didapat sebanyak 159 data.
- False Positive* (FP) menjelaskan bahwa data yang terklasifikasi Tidak Potensial ternyata Potensial Putus Studi. Jumlah data yang didapat sebesar 0 data.
- False Negative* (FN) menjelaskan bahwa data yang terklasifikasi Potensial Putus Studi, memang benar Potensial Putus Studi. Jumlah data yang didapat adalah 0 data.
- True Negative* (TN) menjelaskan bahwa data yang terklasifikasi Potensial Putus Studi, memang benar Potensial Putus Studi. Jumlah data yang didapat sebanyak 140 data.

Dari hasil *confussion matrix* yang telah diperoleh, dihasilkan nilai akurasi, *precision*, *recall*, dan *f1-score* dengan hasil sebagai berikut:

Classification Report:				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	159
1	1.00	1.00	1.00	140
accuracy			1.00	299
macro avg	1.00	1.00	1.00	299
weighted avg	1.00	1.00	1.00	299

Gambar 14. Hasil Pengukuran Kinerja *Algoritma Decision Tree C4.5*

Berdasarkan gambar di atas hasil akurasi yang didapat sebesar 100% dengan nilai *precision* untuk data Potensial Putus Studi dan Tidak Potensial masing-masing sebesar 100% dan 100%. Nilai *recall* dari data Potensial Putus Studi dan Tidak Potensial yang dihasil yaitu 100% dan 100%. selanjutnya untuk nilai *f1-score* dari data Potensial Putus Studi dan Tidak Potensial yang didapat yaitu 100% dan 100%. Untuk nilai *precision*, *recall* dan *f1-score* secara keseluruhan dapat dilihat pada nilai rata-rata *macro (macro average)* dari *precision*, *recall* dan *f1-score*. Nilai rata-rata (*macro average*) dari *precision*, *recall* dan *f1-score* masing-masing sebesar 100%, 100% dan 100%.

Nilai Gain Informasi Pada Masing-Masing Atribut

Angkatan	0.0003
SemesterAwal	0.0003
Ayah pekerjaan	0.00759
Ayah penghasilan	0.0601
Ayah status	0.0004
Ibu pekerjaan	0.0405
Ibu penghasilan	0.0205
Ibu status	0.0085
Wali pekerjaan	0.0377
Wali penghasilan	0.0097
Lulus	0.5370
SemesterTerakhir	0.6819
Kategori IPK	0.3625
Kategori SKS	0.8256

D. Kesimpulan

1. *Algoritma Decision Tree C4.5* telah berhasil diterapkan untuk mengklasifikasikan mahasiswa yang berpotensi putus studi. Proses klasifikasi dilakukan dengan mengolah dataset yang terdiri dari atribut-atribut seperti pekerjaan dan penghasilan orang tua/wali, IPK, SKS, serta variabel lainnya yang relevan dengan status akademik mahasiswa. Penggunaan *Decision Tree C4.5* memungkinkan pembentukan pohon keputusan yang efektif dalam mengidentifikasi mahasiswa dengan risiko putus studi berdasarkan pola yang ada dalam data, serta menghasilkan aturan-aturan klasifikasi yang dapat dengan mudah diinterpretasikan.
2. Hasil penelitian menunjukkan bahwa *Algoritma Decision Tree C4.5* menghasilkan nilai akurasi 100%. Dengan nilai rata-rata *precision* dan *recall* yaitu 100% dan 100%. Nilai rata-rata *f1-score* sebesar 100%. Menunjukkan bahwa *algoritma* ini memiliki performa yang tinggi, sehingga menjadi salah satu metode yang efektif dalam Mengklasifikasi Mahasiswa Berpotensi Putus Studi Pada Fakultas Teknik Universitas Muhammadiyah Makassar.

Saran

Saran untuk penelitian selanjutnya yaitu:

1. Untuk proses klasifikasi dapat menggunakan algoritma yang lain seperti *Knearest Neighbour*, *Naive Bayes*, *Support Vector Machines* (SVM) atau menggunakan *Decision tree* yang lain contoh salah satunya adalah CART (*Classification and Regression Trees*). Bisa juga dengan melakukan 2 perbandingan algoritma klasifikasi atau bahkan lebih.
2. Untuk pengukuran kinerja klasifikasi dapat ditambah dengan metode lain agar hasil lebih optimal.

E. Referensi

- Arfyanti, I., Fahmi, M., & Adytia, P. (2022). Penerapan Algoritma Decision Tree Untuk Penentuan Pola Penerima Beasiswa KIP Kuliah. *Building of Informatics, Technology and Science (BITS)*, 4(3), 1196-1201.
- Gaol, N. Y. L. (2020). Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4. 5. *Jurnal Informasi dan Teknologi*, 23-29.
- Hakam, M., Sudarno, S., & Hoyyi, A. (2015). Analisis jalur terhadap faktor-faktor yang mempengaruhi indeks prestasi kumulatif (IPK) mahasiswa statistika UNDIP. *Jurnal Gaussian*, 4(1), 61-70.
- Hermanto, H. (2020). Prediksi Kelulusan dan Putus Studi Mahasiswa dengan Pendekatan Bertingkat pada Perguruan Tinggi. *Jurnal SIMADA (Sistem Informasi dan Manajemen Basis Data)*, 3(2), 140-148.
- Mudawama, I. R. (2022). *Analisis Perbedaan Prestasi Akademik PAI Siswa Ditinjau Dari Input Jalur Penerimaan Peserta Didik Baru (Studi Kasus di SMP Negeri 5 Nganjuk)* (Doctoral dissertation, IAIN Kediri).
- Nasrullah, A. H. (2021). Implementasi algoritma Decision Tree untuk klasifikasi produk laris. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, 7(2), 45-51.
- Orpa, E. P. K., Ripanti, E. F., & Tursina, T. (2019). Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma Decision Tree C4. 5. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, 7(4), 272-278.
- Quadri, M. M., & Kalyankar, N. V. (2010). Drop out feature of student data for academic performance using decision tree techniques. *Global Journal of Computer Science and Technology*, 10(2), 2-5.
- Rahmadeyan, A., & Mustakim, M. (2023). Seleksi Fitur pada Supervised Learning: Klasifikasi Prestasi Belajar Mahasiswa Saat dan Pasca Pandemi COVID-19. *Jurnal Nasional Teknologi dan Sistem Informasi*, 9(1), 21-32.
- Rahmadeyan, A., & Mustakim, M. (2023). Seleksi Fitur pada Supervised Learning: Klasifikasi Prestasi Belajar Mahasiswa Saat dan Pasca Pandemi COVID-19. *Jurnal Nasional Teknologi dan Sistem Informasi*, 9(1), 21-32.
- Rahmadeyan, A., & Mustakim, M. (2023). Seleksi Fitur pada Supervised Learning: Klasifikasi Prestasi Belajar Mahasiswa Saat dan Pasca Pandemi COVID-19. *Jurnal Nasional Teknologi dan Sistem Informasi*, 9(1), 21-32.
- Samasil, S., Yuyun, Y., & Hazriani, H. (2022). Klasifikasi Mahasiswa Berpotensi Drop Out Menggunakan Algoritma Naive Bayes Dan Decision Tree. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, 8(2), 108-114.
- Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine. *eProceedings of Engineering*, 8(5).
- Sartika, D., & Sensuse, D. I. (2017). Perbandingan algoritma klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada studi kasus pengambilan keputusan pemilihan pola pakaian. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 3(2), 151-161.
- Sinaga, D., Solaiman, E. J., & Kaunang, F. J. (2021). Penerapan Algoritma Decision Tree C4. 5 Untuk Klasifikasi Mahasiswa Berpotensi Drop out Di Universitas Advent Indonesia. *TelKa*, 11(2), 167-173.
- Umamah, K. N., Anggraini, M. P., Edyta, N., & Faradiba, A. T. (2018). Prestasi Akademik Ditinjau Dari Keterlibatan Remaja Dalam Kegiatan Ekstrakurikuler. *Jurnal Muara Ilmu Sosial, Humaniora, Dan Seni*, 2(1), 108-114.

Susilo, U., & Arifin, M. (2020). Analisis Hubungan Indeks Prestasi Semester Dan Indeks Prestasi Kumulatif Dengan Prestasi Mahasiswa Fakultas Ekonomi Universitas Kadiri. *RISK: Jurnal Riset Bisnis dan Ekonomi*, 1(1), 12-22.