

Prediksi Keberhasilan Studi Mahasiswa Menggunakan Metode *Iterative Dichotomiser 3*

DOI: <http://dx.doi.org/10.35889/jutisi.v14i3.3146>

Creative Commons License 4.0 (CC BY – NC)

Ester Gea¹, Yessica Nataliani^{2*}

Sistem Informasi, Universitas Kristen Satya Wacana, Salatiga, Indonesia

*e-mail Corresponding Author: yessica.nataliani@uksw.edu

Abstract

Student academic success is an important indicator in evaluating the quality of higher education in Indonesia. This study aims to predict student academic success based on several variables, such as second-year GPA, number of credits passed, and student's province of origin, using the Iterative Dichotomiser 3 (ID3) algorithm. The data used were 650 graduates of the Information Systems Department of the Faculty of Information Technology, Satya Wacana Christian University, from 2013 to 2017. The ID3 classification results indicate that second-year GPA is the most significant attribute in determining academic success, while province of origin also influences the likelihood of students graduating or failing. Model evaluation using a confusion matrix yielded an accuracy of 85.38%, a precision of 88.8%, a recall of 89.8%, and F-measure of 89.3. These findings demonstrate that the ID3 method can be used in predicting student academic success.

Keywords: Academic success; Student; Prediction; Iterative Dichotomiser 3

Abstrak

Keberhasilan studi mahasiswa merupakan indikator penting dalam mengevaluasi kualitas pendidikan perguruan tinggi di Indonesia. Penelitian ini bertujuan untuk memprediksi keberhasilan studi mahasiswa berdasarkan sejumlah variabel seperti IPK tahun kedua, jumlah SKS lulus, dan provinsi asal mahasiswa dengan algoritma *Iterative Dichotomiser 3* (ID3). Data yang digunakan berasal dari lulusan Fakultas Teknologi Informasi Program Studi Sistem Informasi Universitas Kristen Satya Wacana (UKSW) tahun 2013–2017 sebanyak 650 mahasiswa. Hasil klasifikasi ID3 menunjukkan bahwa IPK tahun kedua merupakan atribut paling signifikan dalam menentukan keberhasilan studi, sedangkan provinsi asal juga berpengaruh terhadap kecenderungan kelulusan atau kegagalan mahasiswa. Evaluasi model dilakukan menggunakan *confusion matrix*, hasilnya menunjukkan akurasi sebesar 85.38%, *precision* sebesar 88.8%, *recall* sebesar 89.8% dan *F-measure* sebesar 89.3%. Temuan ini menunjukkan bahwa metode ID3 dapat digunakan dalam memprediksi keberhasilan studi mahasiswa.

Kata kunci: Keberhasilan studi; Mahasiswa; Prediksi; Iterative Dichotomiser 3

1. Pendahuluan

Akreditasi program studi merupakan salah satu indikator penting dalam menjamin mutu pendidikan tinggi di Indonesia. Salah satu indikator utama yang dinilai oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) dalam proses akreditasi merupakan presentase kelulusan mahasiswa tepat waktu yaitu dalam masa studi maksimal empat tahun untuk jenjang sarjana (S1). Penelitian yang dilakukan oleh [1] pada program studi sarjana di Universitas PQR untuk menilai potensi kepatuhan terhadap indikator kelulusan tepat waktu dalam kerangka pemantauan akreditasi menunjukkan presentase yang cukup tinggi untuk tidak lolos pemantauan program studi oleh Badan Akreditasi. Kondisi ini menegaskan bahwa memahami faktor-faktor yang memengaruhi keberhasilan studi merupakan kebutuhan mendesak dalam penguatan mutu pendidikan tinggi.

IPK merupakan faktor utama yang berperan dalam menentukan pencapaian akademik mahasiswa [2]. Namun tidak menutup kemungkinan jika faktor lain dapat mempengaruhi

keberhasilan studi mahasiswa. Pada konteks Fakultas Teknologi Informasi UKSW, data lulusan menunjukkan adanya variasi signifikan dalam capaian studi mahasiswa, baik dalam hal ketepatan waktu kelulusan maupun tingkat kegagalan. Perbedaan ini dipengaruhi oleh sejumlah faktor seperti IPK tahun kedua, jumlah SKS yang diselesaikan, serta provinsi asal mahasiswa. Namun hingga kini, belum tersedia analisis terukur yang secara spesifik memetakan faktor mana yang paling berpengaruh serta daerah asal mana yang menyumbang tingkat kegagalan tertinggi.

Dalam upaya menyelesaikan permasalahan mengenai identifikasi variabel yang paling berpengaruh terhadap keberhasilan studi mahasiswa serta pemetaan provinsi yang memiliki tingkat kegagalan studi tertinggi, penelitian ini menawarkan solusi berbasis metode *data mining*, yaitu algoritma *Iterative Dichotomiser 3* (ID3). Metode ini dipilih karena memiliki kemampuan yang kuat dalam membangun model klasifikasi yang mampu mengekstraksi pola penting dari data akademik mahasiswa secara terstruktur dan mudah diinterpretasikan. ID3 bekerja dengan konsep *information gain* untuk menentukan atribut paling relevan sebagai dasar pengambilan keputusan, sehingga efektif untuk mengidentifikasi faktor dominan yang mempengaruhi hasil studi mahasiswa. Rasionalisasi pemilihan ID3 didukung oleh berbagai penelitian sebelumnya seperti penelitian yang dilakukan oleh Waridho dan Fauzan ditahun 2021 yang menunjukkan efektivitas dan akurasi dalam berbagai konteks, termasuk prediksi keberlangsungan studi mahasiswa dengan akurasi 79,22% [3], klasifikasi potensi drop out dengan akurasi 83,1%. serta klasifikasi risiko kesehatan dan kelayakan kredit dengan tingkat akurasi mencapai 90%.

Penelitian ini dilakukan dengan tujuan untuk menemukan variabel apa saja yang berpengaruh secara signifikan mempengaruhi keberhasilan mahasiswa dalam studinya, serta mahasiswa dari provinsi mana saja yang memiliki tingkat kegagalan studi yang paling tinggi. Penelitian ini diharapkan memberikan berbagai manfaat bagi berbagai pihak, terutama bagi institusi pendidikan tinggi untuk merancang kebijakan yang tepat dalam mendukung keberhasilan studi mahasiswa dari berbagai latar belakang. Untuk pemerintah, hasil penelitian ini diharapkan memberikan masukan dalam merancang kebijakan pendidikan yang lebih komprehensif, sehingga dapat memperhatikan kebutuhan pendidikan di berbagai wilayah di Indonesia.

2. Tinjauan Pustaka

Penelitian ini menggunakan *Iterative Dichotomiser* (ID3). Algoritma ID3 ditingkatkan oleh J. Ross Quinlan di tahun 1993 dan merupakan algoritma pembelajaran pohon keputusan yang sederhana [4]. ID3 berfungsi untuk membangun sebuah pohon keputusan untuk membantu menemukan solusi untuk masalah. Keputusan yang dibuat melalui proses pencarian nilai terbaik (*the best classifier*) akan dijadikan sebagai akar [5]. Metode ID3 memiliki keunggulan untuk mudah dipahami sehingga memudahkan untuk proses pengambilan keputusan [6]. Algoritma ini mampu memprediksi Faktor-faktor yang bisa memengaruhi kemampuan mahasiswa dalam menyelesaikan studi mereka. ID3 sendiri sudah banyak digunakan untuk memprediksi berbagai topik penelitian tertentu. Sebagai contoh penelitian terkait implementasi algoritma ID3 untuk prediksi keberlangsungan studi mahasiswa yang dilakukan oleh Bharin Rizqi Waridhon dan Abdul Charis Fauzan pada tahun 2021 tentang Penerapan ID3 untuk Memprediksi kemampuan mahasiswa dalam melanjutkan studi. yang dimana parameter yang digunakan berupa IPK, IPS, Jalur Masuk, dan SKS lulus. Studi ini menggunakan metode ID3 dan menghasilkan hasil evaluasi Confusion Matrix dengan akurasi 79.22%, presisi 70.83%, sensitivitas 94.44%, dan spesifisitas 65.85% [3]. Pada penelitian yang dilakukan oleh D. Irmayanti, Y. Muhyidin, dan D. A. Nurjaman dengan judul *Prediksi Mahasiswa Berpotensi Drop Out Menggunakan metode ID3*, algoritma ID3 dimanfaatkan untuk menganalisis pola kelulusan mahasiswa. Parameter yang digunakan dalam penelitian tersebut meliputi Indeks Prestasi Semester 1 hingga 4 serta jumlah SKS yang diambil. Hasil penelitian menunjukkan tingkat akurasi sebesar 83,1%, presisi 86,2%, dan recall 73,2% [4].

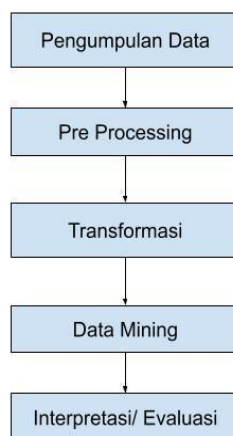
Dalam sektor kesehatan, algoritma ID3 dimanfaatkan untuk mengklasifikasikan faktor risiko diabetes melitus. Penelitian ini menggunakan beberapa atribut klasifikasi, yaitu jenis kelamin, usia, kadar gula darah sewaktu (GDS), lipoprotein berdensitas tinggi (HDL), lipoprotein berdensitas rendah (LDL), serta trigliserida. Hasil perhitungan menunjukkan bahwa metode ID3 memiliki akurasi 90% pada data pengujian [6]. Dalam bidang perbankan, ID3 digunakan untuk menemukan pola dan elemen yang mempengaruhi penilaian kelayakan kredit kendaraan bermotor dengan tingkat akurasi yang memuaskan [7].

Penelitian ini menggunakan IPK tahun kedua, Jumlah SKS tahun kedua, dan menambahkan variabel provinsi asal sebagai faktor sosial-geografis yang belum banyak dieksplorasi dalam penelitian prediksi keberhasilan studi. Hasil penelitian menunjukkan bahwa

provinsi asal memiliki kontribusi nyata terhadap kecenderungan keberhasilan atau kegagalan studi, sehingga memberi perspektif baru bahwa latar belakang geografis dapat berperan penting dalam keberhasilan studi.

3. Metodologi

Tahapan penelitian ini terdiri dari pengumpulan data, *pre-processing*, transformasi data, data mining, dan interpretasi/evaluasi data, yang dapat dilihat dari Gambar 1. Penelitian ini menggunakan aplikasi *Weka* sebagai alat bantu dalam pengolahan data.



Gambar 1. Diagram Tahapan Penelitian

3.1. Pengumpulan Data

Data dalam penelitian ini menggunakan data sekunder, yaitu data mahasiswa Fakultas Teknologi Informasi Universitas Kristen Satya Wacana yang diperoleh dari Lembaga Penjaminan Mutu (LPM) dalam format Excel. Setelah mendapatkan data tersebut, akan dipilih beberapa atribut tertentu yang nantinya digunakan dalam proses penelitian. Atribut yang digunakan berupa Provinsi Asal SMA merepresentasikan daerah asal mahasiswa. Atribut SKS Tahun Kedua menggambarkan jumlah beban akademik yang berhasil diselesaikan mahasiswa hingga akhir tahun kedua. Atribut IPK Tahun Kedua menjadi indikator akademik yang sangat menentukan karena memberikan gambaran mengenai performa belajar mahasiswa pada tahap awal masa studi. Atribut terakhir adalah *Keberhasilan Studi*, yang digunakan sebagai variabel target (kelas) dalam proses klasifikasi.

3.2. Pre-Processing

Pre-processing adalah proses mengubah data mentah menjadi format atau data yang sesuai untuk tahap analisis selanjutnya [8]. Penanganan data yang hilang, penghapusan duplikasi, normalisasi data, dan pengubah variabel kategorikal menjadi bentuk numerik yang dapat diproses oleh algoritma data mining adalah beberapa bagian dari proses *preprocessing* [9]. Proses ini dilakukan untuk meningkatkan mutu data yang digunakan sehingga menjadi lebih relevan dan akurat.

3.3. Transformasi

Proses selanjutnya yaitu transformasi, proses ini meliputi normalisasi data, pembuatan atribut baru, pengkodean data kategorik menjadi numerik, atau agregasi data dari level detail ke level yang lebih umum. Dengan melakukan transformasi, kualitas data akan meningkat sehingga mampu menghasilkan pola atau informasi yang lebih akurat pada tahap analisis selanjutnya. Selain itu, tahap transformasi juga membantu dalam mengurangi kompleksitas data dan meminimalisir adanya bias atau ketidaksesuaian data saat masuk ke proses data mining.

3.4. Data Mining

Penelitian ini menggunakan metode ID3, yang merupakan metode yang menggunakan pohon keputusan dengan menggunakan beberapa variabel. Nilai Gain tertinggi dari atribut yang

ada digunakan untuk memilih atribut sebagai akar. Gain dapat dihitung dengan menggunakan Persamaan (1) sebagai berikut [10].

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

dimana S adalah himpunan kasus, A adalah atribut, n menunjukkan jumlah partisi pada atribut A , $|S_i|$ adalah jumlah kasus pada partisi ke- i , $|S|$ adalah jumlah kasus dalam S , dan $Entropy(S_i)$ merupakan nilai *entropy* untuk sampel-sampel yang memiliki nilai i .

Untuk mendapatkan nilai *Gain* dilakukan perhitungan untuk mencari nilai *Entropy*. *Entropy* adalah teori yang dapat mengukur tingkat homogenitas dan impurity dalam kumpulan data [11]. Secara matematis *entropy* dapat dirumuskan pada Persamaan (2).

$$Entropy(S) = \sum_{i=1}^n -P_i \log_2 P_i \quad (2)$$

dimana S adalah himpunan kasus, n adalah jumlah partisi S , P_i adalah proporsi dari S_i terhadap S .

3.5. Interpretasi/ Evaluasi

Setelah pohon keputusan dibuat, model dievaluasi dengan menggunakan nilai F1 dan confusion matrix. Confusion matrix adalah cara untuk menghitung akurasi dalam konsep data mining [12]. Tujuan dari perhitungan *confusion matrix* ini bertujuan untuk mengukur akurasi algoritma yang ada. *F-measure* lebih direkomendasikan untuk mengevaluasi data yang bersifat tidak seimbang dibandingkan *accuracy*. [13]. *F-measure* memberikan gambaran yang lebih menyeluruh terhadap performa model pada kedua kelas, karena hanya akan tinggi jika baik precision maupun recall sama-sama tinggi. Persamaan (3) merupakan rumus dari *F1-measure* [14]. Persamaan (4) sampai (6) merupakan rumus dari akurasi, *precision*, dan *recall* [15].

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

dimana TP adalah *True Positif*, FN adalah *False Negatif*, FP adalah *False Positif*, dan TN adalah *True Negatif*.

4. Hasil dan Pembahasan

4.1 Pengumpulan Data

Data yang digunakan untuk penelitian ini bersumber dari Lembaga Penjaminan Mutu (LPM) UKSW. Data yang diperoleh berupa data lulusan mahasiswa Fakultas Teknologi Informasi dari angkatan 2013 sampai 2017 dalam bentuk *Excel*. Data dari masa tersebut digunakan untuk mengetahui apakah mahasiswa tersebut lulus tepat waktu ataupun tidak. Total data yang digunakan sebanyak 650 data mahasiswa dengan atribut provinsi asal berdasarkan tempat lahir, jumlah SKS tahun kedua, IPK tahun kedua, dan status keberhasilan studi. Atribut status keberhasilan studi dijadikan sebagai label atau *class attribute* dalam proses klasifikasi. Label awal terdiri dari tiga kategori yaitu lulus tepat waktu, lulus terlambat, dan gagal.

4.2 Pre-Processing

Setelah data dikumpulkan, tahap berikutnya adalah *pre-processing*. Tahapan ini melibatkan pembersihan data dari nilai kosong dengan melengkapi data tersebut melalui pencocokan dan pengambilan data dari sistem fakultas, sehingga tidak dilakukan penghapusan data. Kesalahan input dengan memastikan data numerik dalam rentang yang wajar seperti IPK

antara 0 sampai 4.00 dan jumlah SKS tidak melebihi dari batas jumlah SKS di tahun kedua. Data juga diklasifikasikan agar sesuai untuk proses klasifikasi, sehingga label "lulus tepat waktu" dan "lulus terlambat" dikategorikan menjadi satu kategori yaitu berhasil, sedangkan label "tidak lulus" atau "drop out" dikategorikan sebagai gagal. Hasil dari *pre-processing* bisa dilihat pada Gambar 2.

PROVINSI ASAL SMA	SKS THN 2	IPK THN 2	KEBERHASILAN STUDI
Lampung	75	3.24	Berhasil
Jawa Tengah	13	2.42	Gagal
Jawa Tengah	86	3.44	Berhasil
Jawa Tengah	88	3.46	Berhasil
Jawa Barat	87	3.71	Berhasil
Jawa Tengah	82	3.29	Berhasil
Jawa Tengah	74	3.15	Berhasil
Jawa Tengah	29	2.63	Berhasil
Maluku	85	3.39	Berhasil
Maluku	85	3.49	Berhasil
Maluku	0	0.00	Gagal
Maluku	87	3.44	Berhasil
Jawa Tengah	71	2.78	Berhasil
Jawa Tengah	80	3.33	Gagal
Jawa Tengah	74	3.63	Berhasil
Kalimantan Barat	48	2.33	Gagal
Lampung	77	3.19	Berhasil
Sulawesi Tengah	87	3.53	Berhasil
Maluku	61	1.73	Gagal
Sulawesi Utara	73	2.97	Gagal

Gambar 2. Data Hasil *Pre-Processing*

4.3. Transformasi

Pada tahap ini data disederhanakan dan dikelompokkan. Tahap transformasi bertujuan untuk menyiapkan data dalam format yang dapat digunakan oleh algoritma klasifikasi. Atribut numerik seperti IPK dan SKS digunakan tanpa perlu transformasi karena sudah dalam bentuk numerik kontinu, sehingga dapat dikenali oleh algoritma ID3 dalam *Weka*. Data dalam bentuk *Excel* dibuat dalam format CSV kemudian diimpor dalam *Weka*. Proses ini bisa dilihat pada gambar 3.

PROVINSI ASAL	SKS THN 2	IPK THN 2	KEBERHASILAN STUDI
Lampung	75	3.24	Berhasil
Jawa Tengah	13	2.42	Gagal
Jawa Tengah	86	3.44	Berhasil
Jawa Tengah	88	3.46	Berhasil
Jawa Barat	87	3.71	Berhasil
Jawa Tengah	82	3.29	Berhasil
Jawa Tengah	74	3.15	Berhasil
Jawa Tengah	29	2.63	Berhasil
Maluku	85	3.39	Berhasil
Maluku	85	3.49	Berhasil
Maluku	0	0.00	Gagal
Maluku	87	3.44	Berhasil
Jawa Tengah	71	2.78	Berhasil
Jawa Tengah	80	3.33	Gagal
Jawa Tengah	74	3.63	Berhasil
Kalimantan Barat	48	2.33	Gagal
Lampung	77	3.19	Berhasil
Sulawesi Tengah	87	3.53	Berhasil
Maluku	61	1.73	Gagal

Gambar 3. Hasil Transformasi

4.4. Data Mining

Data diproses menggunakan aplikasi *Weka* setelah proses pengumpulan, *preprocessing*, dan transformasi selesai. Gambar 4 menunjukkan data set yang telah berhasil di-upload di aplikasi *Weka*.

Relation: Data Penelitian Fix

No.	1: PROVINSI ASAL	2: SKS THN 2	3: IPK THN 2	4: KEBERHASILAN STUDI
	Nominal	Numeric	Numeric	
1	Lampung	75.0	3.24	Berhasil
2	Jawa Tengah	13.0	2.42	Gagal
3	Jawa Tengah	86.0	3.44	Berhasil
4	Jawa Tengah	88.0	3.46	Berhasil
5	Jawa Barat	87.0	3.71	Berhasil
6	Jawa Tengah	82.0	3.29	Berhasil
7	Jawa Tengah	74.0	3.15	Berhasil
8	Jawa Tengah	29.0	2.63	Berhasil
9	Maluku	85.0	3.39	Berhasil
10	Maluku	85.0	3.49	Berhasil
11	Maluku	0.0	0.0	Gagal
12	Maluku	87.0	3.44	Berhasil
13	Jawa Tengah	71.0	2.78	Berhasil
14	Jawa Tengah	80.0	3.33	Gagal
15	Jawa Tengah	74.0	3.63	Berhasil
16	Kalimantan Barat	48.0	2.33	Gagal
17	Lampung	77.0	3.19	Berhasil
18	Sulawesi Tengah	87.0	3.53	Berhasil
19	Maluku	61.0	1.73	Gagal
20	Sulawesi Utara	73.0	2.97	Gagal

Gambar 4. Data Set

Data yang sudah masuk akan dipecah menjadi dua bagian, yaitu data untuk latihan sebanyak 80% dan data untuk pengujian sebanyak 20%. Pembagian ini menggunakan metode percentage split sebesar 80%. Dari total 650 data, akan ada 520 data yang digunakan untuk training serta 130 data untuk testing. Selain itu, nilai ConfidenceFactor yang digunakan adalah 0.55. Untuk mengetahui pengaruh nilai *confidence factor* terhadap akurasi model, dilakukan pengujian dengan mengubah nilai *ConfidenceFactor*. Hasil perbandingan dapat dilihat dalam tabel 1, untuk parameter lainnya dalam kondisi *default* disajikan pada gambar 5.

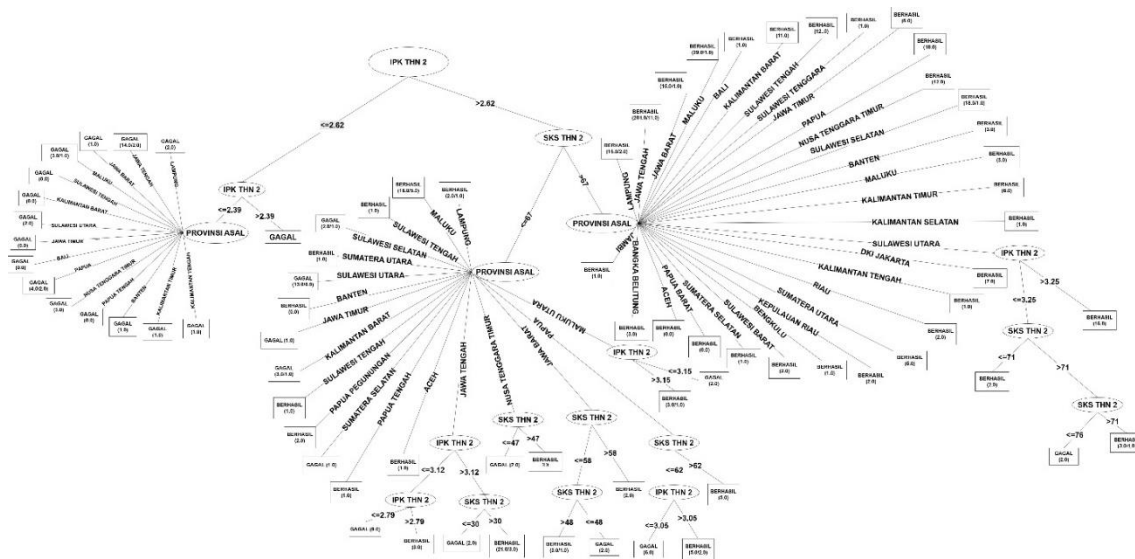
Tabel 1. Hasil perbandingan nilai *ConfidenceFactor*

<i>ConfidenceFactor</i>	Akurasi
0.25	92.307%
0.30	92.307%
0.55	85.380%
0.60	85.846%

batchSize	100	numFolds	3
binarySplits	False	reducedErrorPruning	False
collapseTree	True	saveInstanceData	False
confidenceFactor	0.55	seed	1
debug	False	subtreeRaising	True
doNotCheckCapabilities	False	unpruned	False
doNotMakeSplitPointActualValue	False	useLaplace	False
minNumObj	2	useMDLcorrection	True
numDecimalPlaces	2		

Gambar 5. Pararameter

Hasil model prediksi keberhasilan studi dengan menggunakan algoritma ID3 disajikan pada Gambar 6.



Gambar 6. Decision Tree

Dari Gambar 4 bisa dilihat bahwa atribut yang paling berpengaruh pada prediksi keberhasilan studi mahasiswa yaitu IPK tahun ke-2, di mana IPK THN 2 menjadi simpul akar (*root node*) dan menjadi dasar utama dalam memisahkan kelas “Berhasil” dan “Gagal”. Data set test yang digunakan adalah 650 data yang dimana total untuk kelas berhasil sebanyak 455 data dan untuk total kelas gagal sebanyak 195 data. Dari pohon keputusan diatas, kita dapat melakukan perhitungan *Gain* dan *Entropy* dari setiap atribut yang ada dengan rumus persamaan (1) dan persamaan (2). Tahap awal yang dilakukan yaitu dengan menghitung nilai *Entropy* total secara keseluruhan yang nantinya akan menjadi dasar untuk perhitungan nilai *Gain* pada setiap atribut.

$$\begin{aligned}
 Entropy(S) &= -\left(\frac{455}{650} \times \log_2\left(\frac{455}{650}\right)\right) - \left(\frac{195}{650} \times \log_2\left(\frac{195}{650}\right)\right) \\
 &= -(0.70 \times \log_2(0.70)) - (0.30 \times \log_2(0.30)) \\
 &= 0.360 + 0.521 \\
 &= 0.881
 \end{aligned}$$

Nilai entropy total sebesar 0.881 yang artinya model *decision tree* perlu memecah data lagi berdasarkan atribut tertentu. Setelah itu dihitung nilai *Entropy* IPK THN 2 yang menghasilkan dua cabang yaitu ($IPK\ THN\ 2 \leq 2.62$) dan ($IPK\ THN\ 2 > 2.62$). Batasan 2.62 ini diambil dari perhitungan nilai gain paling tinggi untuk setiap nilai IPK yang ada.

$$\begin{aligned}
 Entropy(S) &= -\left(\frac{7}{132} \times \log_2\left(\frac{7}{132}\right)\right) - \left(\frac{125}{132} \times \log_2\left(\frac{125}{132}\right)\right) \\
 &= -(0.053 \times \log_2(0.053)) - (0.946 \times \log_2(0.946)) \\
 &= 0.229 + 0.069 \\
 &= 0.299
 \end{aligned}$$

$$\begin{aligned}
 Entropy(S) &= -\left(\frac{448}{518} \times \log_2\left(\frac{448}{518}\right)\right) - \left(\frac{70}{518} \times \log_2\left(\frac{70}{518}\right)\right) \\
 &= -(0.864 \times \log_2(0.864)) - (0.135 \times \log_2(0.135)) \\
 &= 0.174 + 0.397 \\
 &= 0.571
 \end{aligned}$$

$$\begin{aligned}
 Gain(S, A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \\
 &= 0.881 - ((0.203 \times 0.299) + (0.796 \times 0.571)) \\
 &= 0.881 - 0.516 \\
 &= 0.365
 \end{aligned}$$

Dari perhitungan diatas menunjukkan bahwa atribut “IPK THN 2” memiliki nilai gain paling tinggi sebesar 0.365 sehingga mampu dijadikan sebagai pemisah data pada tahap pembentukan pohon keputusan.

Provinsi yang paling sering gagal dalam studi berdasarkan pohon keputusan tersebut berasal dari provinsi Lampung, Jawa Tengah, Sulawesi Utara. Ketiganya muncul dalam beberapa cabang pohon sebagai daerah asal mahasiswa yang gagal. Hal ini terjadi karena nilai IPK yang rendah maupun kombinasi nilai IPK dan jumlah SKS yang tidak mencukupi di tahun kedua. Sementara itu, provinsi yang diprediksi sering berhasil adalah Jawa Tengah.

4.5. Evaluasi Dan Interpretasi

Setelah mendapatkan model keputusan, model klasifikasi dievaluasi melalui penerapan Confusion Matrix dan F-measure. Adapun hasil evaluasi model klasifikasi algoritma ID3 dapat dilihat melalui Tabel 2.

Tabel 2. Hasil *Confusion Matrix*

<i>Actual</i>	<i>Predicted</i>		<i>Total Actual</i>
	Berhasil	Gagal	
Berhasil	79	9	88
Gagal	10	32	42
<i>Total Predicted</i>	89	41	130

Jika dilihat dari hasil Tabel 2, maka dapat dihitung tingkat ketepatan akurasi modelnya sebagai berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{79+32}{130} = \frac{111}{130} = 85.38 \% \\
 \text{Precision} &= \frac{TP}{TP+FP} = \frac{79}{79+10} = \frac{79}{89} = 0.888 = 88.8 \% \\
 \text{Recall} &= \frac{TP}{TP+FN} = \frac{79}{79+32} = \frac{79}{111} = 0.709 = 70.9 \% \\
 \text{F - Measure} &= 2 \times \frac{0.888 \times 0.709}{0.888 + 0.709} = 0.793 = 79.3 \%
 \end{aligned}$$

Dari hasil perhitungan di atas diperoleh akurasi sebesar 85.38%, *precision* sebesar 88.8%, *recall* sebesar 70.9 %, dan *F-measure* sebesar 79.3 %. Jika dilihat dari nilai *F-measure* dan akurasi, model tersebut termasuk baik dan dapat digunakan untuk mengklasifikasikan prediksi keberhasilan studi mahasiswa.

4.6 Pembahasan

Hasil pengujian menunjukkan bahwa IPK tahun kedua merupakan atribut paling berpengaruh dalam memprediksi keberhasilan studi, dengan akurasi model ID3 mencapai 85,38%. Besarnya nilai *precision*, *recall*, dan *F-measure* memperlihatkan bahwa model mampu membedakan kelas “Berhasil” dan “Gagal” secara konsisten. Temuan ini sejalan dengan pandangan bahwa performa akademik awal merupakan indikator kuat keberhasilan studi, sehingga mengonfirmasi hasil penelitian sebelumnya seperti [3] yang juga menempatkan IPK sebagai prediktor utama meskipun dengan akurasi lebih rendah (79,22%). Temuan ini juga sejalan dengan [4] yang memperoleh akurasi 83,1% pada studi prediksi drop out yang memperlihatkan bahwa algoritma ID3 tetap kompetitif pada domain pendidikan.

Kontribusi utama penelitian ini adalah penambahan provinsi asal sebagai variabel yang terbukti muncul pada cabang-cabang pohon keputusan. Hal ini memperluas temuan penelitian terdahulu yang sebagian besar hanya berfokus pada variabel akademik dan jarang memasukkan faktor geografis.

5. Simpulan

Metode *Iterative Dichotomiser 3 (ID3)* dapat digunakan untuk memprediksi keberhasilan studi mahasiswa dengan tingkat akurasi yang tinggi. Dengan data mahasiswa Fakultas Teknologi Informasi UKSW dari tahun 2013-2017, penelitian ini berhasil mengidentifikasi bahwa variabel yang paling berpengaruh terhadap keberhasilan studi adalah IPK tahun kedua, yang menjadi akar dalam pohon keputusan. Provinsi asal mahasiswa juga ditemukan memiliki hubungan terhadap

hasil studi, di mana mahasiswa dari provinsi seperti Lampung, Jawa Tengah, dan Sulawesi Utara lebih sering diklasifikasikan gagal, terutama bila memiliki IPK dan jumlah SKS yang rendah. Evaluasi model dengan *confusion matrix* menunjukkan akurasi sebesar 85.38%, kemampuan dalam mengklasifikasikan data dengan baik. Hasil ini diharapkan dijadikan dasar dalam menyusun kebijakan pendidikan yang lebih tepat sasaran oleh institusi pendidikan tinggi di UKSW maupun pemerintah, terutama dalam mendukung keberhasilan studi mahasiswa dari berbagai latar belakang daerah.

Daftar Referensi

- [1] S. Rachardian and E. Sedyono, "Prediksi kelulusan tepat waktu mahasiswa untuk pemantauan program studi menggunakan metode data mining," *Aiti*, vol. 21, no. 2, pp. 168–182, 2024, doi: 10.24246/aiti.v21i2.168-182.
- [2] A. Karin Octaviana, A. Wulan Agustin, H. Nur Taqiyyah, S. Naufal Almubarraq, and T. Syifa Zahira, "Hubungan Jarak Kos dengan IPK Mahasiswa Pendidikan Geografi di Universitas Pendidikan Indonesia," *J. Sos. dan sains*, vol. 3, no. 5, pp. 538–546, 2023, doi: 10.59188/jurnalsosains.v3i5.794.
- [3] B. R. Waridhon and A. C. Fauzan, "Implementasi Algoritma Iterative Dichotomiser 3 (ID3) untuk Prediksi Keberlangsungan Studi Mahasiswa," vol. 1, no. 2, pp. 64–74, 2021.
- [4] D. Irmayanti, Y. Muhyidin, Y & D.A. Nurjaman, "Prediksi Mahasiswa Berpotensi Drop Out Dengan Metode Iteratif Dichotomiser 3 (ID3). *Jurnal Teknologi Informasi*, vol. 5, no. 2, pp. 103-113, 2021.
- [5] V. M. Sitorus, "Sistem Pendukung Keputusan Branding Brand Berdasarkan Kualitas Produk Menggunakan Metode Iterative Dichotomiser 3 Pada Matahari Thamrin Plaza," vol. 1, no. 1, pp. 1–8, 2022.
- [6] F. Ferdina, N. Satyahadewi, and D. Kusnandar, "Penerapan Algoritma Iterative Dichotomiser 3 (Id3) Dalam Klasifikasi Faktor Risiko Penyakit Diabetes Melitus," *Var. J. Stat. Its Appl.*, vol. 5, no. 2, pp. 139–146, 2023, doi: 10.30598/variancevol5iss2page139-146.
- [7] Amrin, Rudianto, and E. Irfiani, "Analisis Algoritma Iterative Dichotomiser 3 (ID3) untuk Penilaian Kelayakan Kredit Kendaraan Bermotor," *IMTechno J. Ind. Manag. Technol.*, vol. 5, no. 2, pp. 36–40, 2024.
- [8] N. L. Utami, A. Nazir, E. Budianita, and F. Insani, "Jurnal Computer Science and Information Technology (CoSciTech) algoritma apriori," vol. 5, no. 1, pp. 75–83, 2024.
- [9] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and D. N. Rutledge, "New data preprocessing trends based on ensemble of multiple preprocessing techniques," *TrAC - Trends Anal. Chem.*, vol. 132, p. 116045, 2020, doi: 10.1016/j.trac.2020.116045.
- [10] J. Yulinda, R. S. Lubis, and R. Aprilia, "... Metode Classification Analysis Regression Trees dan Iterative Dichotomizer 3 Dalam Mengklasifikasikan Pasien Hipertensi Di Rumah Sakit Umum Daerah Dr ...," *Justek J. Sains dan ...*, vol. 6, no. 4, pp. 482–492, 2023.
- [11] T. Thi Bi Dan, S. Widya Sihwi, and R. Anggrainingsih, "Implementasi Iterative Dichotomiser 3 Pada Data Kelulusan Mahasiswa S1 Di Universitas Sebelas Maret," *J. Teknol. Inf. ITSmart*, vol. 4, no. 2, p. 84, 2016, doi: 10.20961/its.v4i2.1770.
- [12] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix," *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.
- [13] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845.
- [14] D. W. Astuti, M. Muslim, and T. A. Pahlevi, "Substrate Integrated Waveguide Bandpass Filter dengan Complementary Split Ring Resonator," *J. Rekayasa Elektr.*, vol. 15, no. 1, 2019, doi: 10.17529/jre.v15i1.12266.
- [15] N. Hidayah and Dodiman, "Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin," *J. Akad. Pendidik. Mat.*, vol. 10, no. 1, pp. 8–15, 2024.