Komparasi Algoritma Naive Bayes dan Decision Tree Untuk Memprediksi Lama Studi Mahasiswa

Indera Cahyo Wibowo^{1,*}, Abd Charis Fauzan², Marshella Dwi Putri Yustiana³, Fiqih Ainul Qhabib⁴

Program Studi Ilmu Komputer, Universitas Nahdlatul Ulama Blitar, Indonesia

¹inderacw1902@gmail.com; ²abdcharis@unublitar.ac.id; ³marshelladwi17@gmail.com; ⁴fiqihainul61@gmail.com

ARTIKEL INFO

Article history

Diterima: 3 Oktober 2019 Direvisi: 11 November 2019 Diterbitkan: 30 Desember 2019

Keywords

Naïve bayes Decision tree Komparasi algoritma

ABSTRAK

Penelitian ini bertujuan memprediksi lama studi mahasiswa lulus tepat waktu atau terlambat lulus. Data yang digunakan yaitu: gender, status mahasiswa, nilai, dan beasiswa dari semester awal sampai semester akhir di tahun ajaran 2018 - 2019. Penelitian ini menggunakan 2 metode, untuk metode yang pertama adalah Naïve Bayes dan yang kedua adalah Decision Tree. Gunakan semua data yang telah diperoleh dan hitung dengan kedua metode sampai mendapatkan hasil akhir dan akurasinya lalu komparasikan keduanya. Hasil pengkomparasian dari data dan menggunakan dua metode perhitungan keakurasiannya yaitu 30% untuk Naïve bayes, dan 55% untuk metode Decision Tree. Dari komparasi menggunakan dua metode dapat diambil kesimpulan bahwa metode Decision Tree memiliki persentase keakuratan yang lebih tinggi dibandingkan Naïve Bayes sehingga dapat dikatakan metode Decision Tree lebih akurat dan lebih detail daripada Naïve Bayes.

PENDAHULUAN

Salah satu faktor yang menentukan kualitas perguruan tinggi adalah kemampuan mahasiswa untuk menyelesaikan masa studi secara tepat waktu karena kegagalan mahasiswa dalam menyelesaikan program sarjana (S1) akan menyebabkan masalahmasalah baru yang berkaitan dengan masa depan mahasiswa tersebut. Pada dasarnya mahasiswa diberi waktu untuk menyelesaikan program sarjana (S1) dalam kurun waktu 8 semester agar studi selesai tepat waktu. Oleh karenanya, setiap perguruan tinggi pun dirasa perlu untuk melakukan prediksi untuk memperoleh informasi yang tepat dan akurat mengenai mahasiswa yang mampu menyelesaikan masa studinya dengan tepat waktu 8 semester sehingga perguruan tinggi dapat mencegah kegagalan kegiatan akademik mahasiswa secara dini [1]. Pada penelitian terdahulu, algoritma K-Nearest Neighbor btelah digunakan untuk klarifikasi data [2]. Pada penelitian ini akan dirancang perhitungan lama studi mahasiswa dengan menggunakan metode algoritma naive bayes dan decision tree. Naive bayes dan decision tree merupakan metode pengklasifikasi yang terkenal dengan tingkat keakuratan yang baik [3]. Banyak peneliti yang melakukan penelitian dengan menggunakan klarifikasi dua metode tersebut [4]. Kebanyakan dari peneliti mempunyai dasar pada ilmu komputer dan informatika komputer sehingga pada pembahasan ini lebih ditekankan pada hasil pemrograman, keakuratan, dan yang terpenting adalah pada probabilitasnya [5]. Berdasarkan masalah yang diuraikan maka penelitian ini membahas komparasi algoritma naive bayes dan decision tree. Tujuan diadakannya komparasi ini adalah diharapkan akan diketahui metode yang mendapatkan akurasi lebih besar. Berikutnya, metode yang terbaik diantara keduanya dipakai dalam menentukan prediksi kelulusan mahasiswa, yaitu antara tepat waktu atau terlambat.



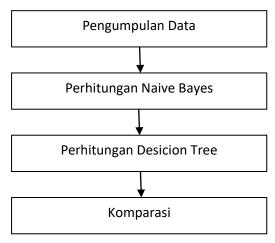
^{*} corresponding author

METODE

E-ISSN: 2715-2731

(1)

Pengumpulan data beserta tahapan untuk komparasi algoritma pada Gambar 1. Gambar 1 menunjukkan urutan dalam pengerjaan komparasi Naive Bayes dan Decision Tree untuk mendapatkan tingkat akurasi yang paling besar diantara dua metode.



Gambar 1. Urutan Pengerjaan Komparasi Naive Bayes dan Decision

Pengumpulan Data

Data studi mahasiswa didapat dari Universitas Nahdlatul Ulama Blitar dari semester awal sampai semester akhir pada tahun ajaran 2018 - 2019, pada data tersebut terdapat atribut asli yaitu: Gender, Status Mahasiswa, Nilai, dan Beasiswa seperti pada Tabel 1.

Perhitungan Naive Bayes

Naive bayes merupakan salah satu metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [6]. Dalam metode ini perhitungan ditunjukkan melalui Persamaan 1:

$$P(H|X) = \frac{P(X|H)}{P(X)}.P(H)$$

Keterangan

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (*posteriori probabilitas*)

P(H): Probabilitas hipotesis H (*prior probabilitas*)

P(X|H): Probabilitas X berdasarkan kondisi pada hipotesis H

P(X): Probabilitas X [7]

Perhitungan Decision Tree

Konsep Decision Tree atau Pohon keputusan adalah mengubah data menjadi aturan - aturan keputusan. Manfaat utama dari penggunaan decision tree adalah kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi simple, sehingga pengambilan keputusan akan lebih memudahkan solusi dari sebuah permasalahan [8]. Decision Tree adalah salah satu metode klasifikasi yang paling populer karena mudah diinterprestasikan oleh manusia [9]. Decision Tree digunakan untuk pengenalan pola dan termasuk dalam pengenalan pola secara statistic [10]. Decision Tree menggunakan 2

perhitungan yang pertama adalah perhitungan Gain pada Persamaan 2 dan perhitungan Entropy pada Persamaan 3 [5].

Perhitungan Gain

Gain (S,A) = Entropy (S) -
$$\sum_{i=1}^{n}$$
 * Entropy (Si)

Keterangan:

S : himpunan A :atribut

n : jumlah partisi atribut A |Si| : jumlah kasus pada partikel ke-i

|S|: jumlah kasusdalam S

Menghitung Nilai Entropy

Entropy (S) =
$$\sum_{i=1}^{n} -pi * \log_2 pi$$
 (3)

Keterangan:

S: himpunan kasus

A : fitur

n : jumlah partisi S

pi : Proporsi dari Si terhadap S [5]

PEMBAHASAN

Untuk menentukan masa studi mahasiswa, pada penelitian ini akan dilakukan analisis menggunakan Algoritma Decision Tree dan juga Naive bayes untuk mengetahui mana yang lebih baik diantara keduanya, sehingga perlu dilakukan proses analisa data terlebih dahulu sebelum data yang digunakan tersebut di komparasikan. Dalam menganalisa data mahasiswa alumni tersebut, maka ada atribut yang digunakan untuk menentukan masa studi mahasiswa yaitu: Gender, Status Mahasiswa, Nilai dan juga Status Beasiswa. Beberapa dari empat atribut tersebut akan dijadikan atribut prediktor atau atribut input untuk menghasilkan atribut target, di mana atribut target tersebut menjadi class output untuk menentukan masa studi yaitu dibedakan menjadi 2 class (Tepat Waktu dan Terlambat). Berikut ini adalah keterangan data mahasiswa yang akan digunakan dalam menentukan masa studi mahasiswa terlihat pada Tabel 1.

Tabel 1. Keterangan Data Mahasiswa

NO	ATRIBUT	KETERANGAN			
1	Gender	Menentukan jenis kelamin dari mahasiswa (pria/wanita)			
2	Status Mahasiswa	Menentukan status mahasiswa (mahasiswa/bekerja)			
3	Nilai	Menentukan nilai mahasiswa (A, B / C)			
4	Beasiswa	Menentukan status beasiswa (YA / TIDAK)			

Berdasarkan data tersebut kemudian dilakukan pemilihan atribut dan sebagian dari data dalam atribut yang ada akan ditransformasikan untuk memudahkan proses mining dalam menentukan masa studi. Adapun data yang digunakan untuk menentukan masa studi mahasiswa ditunjukkan pada Tabel 2, yakni 100 data yang diambil secara acak. Data pada Tabel 3 didapat dari data mahasiswa Universitas Nahdlatul Ulama Blitar pada tahun akademik 2018-2019.

Tabel 2. Sample Data Mahasiswa

NO	GENDER	STATUS MAHASISWA	NILAI	BEASISWA	LAMA STUDI	PREDIKSI	
1	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TERLAMBAT	
2	WANITA	BEKERJA	В	YA	TEPAT WAKTU	TEPAT WAKTU	
3	PRIA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
4	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
5	WANITA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
6	PRIA	BEKERJA	В	YA	TERLAMBAT	TEPAT WAKTU	
7	WANITA	MAHASISWA	В	YA	TEPAT WAKTU	TEPAT WAKTU	
8	PRIA	BEKERJA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
)	PRIA	BEKERJA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
10	WANITA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
11	WANITA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
12	WANITA	BEKERJA	В	YA	TERLAMBAT	TEPAT WAKTU	
13	WANITA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
14	PRIA	BEKERJA	В	YA	TEPAT WAKTU	TEPAT WAKTU	
15	PRIA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
16	WANITA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
17	WANITA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
18	PRIA	BEKERJA	В	YA	TEPAT WAKTU	TERLAMBAT	
19	PRIA	BEKERJA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
20	PRIA	BEKERJA	В	YA	TERLAMBAT	TEPAT WAKTU	
21	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
22	WANITA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
23	WANITA	MAHASISWA	В	YA	TEPAT WAKTU	TEPAT WAKTU	
24	PRIA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
25	PRIA	BEKERJA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
26	WANITA	MAHASISWA	В	YA	TERLAMBAT	TEPAT WAKTU	
27	WANITA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
28	WANITA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
29	WANITA	MAHASISWA	В	YA	TERLAMBAT	TEPAT WAKTU	
30	WANITA	BEKERJA	В	YA	TEPAT WAKTU	TEPAT WAKTU	
31	PRIA	BEKERJA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
32	PRIA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
33	WANITA	BEKERJA	С	TIDAK	TERLAMBAT	TEPAT WAKTU	
34	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TERLAMBAT	
35	WANITA	BEKERJA	В	YA	TEPAT WAKTU	TEPAT WAKTU	
36	PRIA	BEKERJA	В	YA	TERLAMBAT	TEPAT WAKTU	
37	WANITA	BEKERJA	В	YA	TEPAT WAKTU	TEPAT WAKTU	
38	PRIA	MAHASISWA	В	YA	TEPAT WAKTU	TEPAT WAKTU	
39	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
10	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
11	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU	
12	WANITA	BEKERJA	В	YA	TERLAMBAT	TEPAT WAKTU	
43	WANITA	MAHASISWA	В	YA	TERLAMBAT	TEPAT WAKTU	
44	PRIA	BEKERJA	В	YA	TEPAT WAKTU	TEPAT WAKTU	
45	PRIA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU	

46	WANITA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
47	WANITA	BEKERJA	С	TIDAK	TERLAMBAT	TERLAMBAT
48	PRIA	BEKERJA	С	TIDAK	TERLAMBAT	TERLAMBAT
49	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU
50	PRIA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU
51	PRIA	BEKERJA	В	YA	TERLAMBAT	TEPAT WAKTU
52	WANITA	MAHASISWA	В	YA	TEPAT WAKTU	TEPAT WAKTU
53	WANITA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU
54	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU
55	WANITA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU
56	PRIA	BEKERJA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
57	WANITA	BEKERJA	В	YA	TERLAMBAT	TEPAT WAKTU
58	WANITA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
59	WANITA	MAHASISWA	В	YA	TEPAT WAKTU	TERLAMBAT
60	PRIA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU
61	PRIA	BEKERJA	В	YA	TEPAT WAKTU	TEPAT WAKTU
62	PRIA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU
63	WANITA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
64	WANITA	MAHASISWA	С	TIDAK	TERLAMBAT	TERLAMBAT
65	PRIA	MAHASISWA	С	TIDAK	TERLAMBAT	TERLAMBAT
66	WANITA	BEKERJA	В	YA	TEPAT WAKTU	TEPAT WAKTU
67	PRIA	MAHASISWA	В	YA	TERLAMBAT	TEPAT WAKTU
68	WANITA	BEKERJA	В	YA	TEPAT WAKTU	TEPAT WAKTU
69	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU
70	PRIA	BEKERJA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
71	PRIA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
72	PRIA	MAHASISWA	В	YA	TERLAMBAT	TEPAT WAKTU
73	WANITA	MAHASISWA	В	YA	TEPAT WAKTU	TEPAT WAKTU
74	WANITA	BEKERJA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
75	WANITA	MAHASISWA	A	YA	TEPAT WAKTU	TERLAMBAT
76	PRIA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU
77	WANITA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
78	PRIA	BEKERJA	С	TIDAK	TERLAMBAT	TERLAMBAT
79	WANITA	MAHASISWA	В	YA	TEPAT WAKTU	TEPAT WAKTU
80	PRIA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU
81	PRIA	MAHASISWA	В	YA	TEPAT WAKTU	TEPAT WAKTU
82	PRIA	BEKERJA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
83	WANITA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
84	WANITA	MAHASISWA	С	TIDAK	TERLAMBAT	TERLAMBAT
85	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TERLAMBAT
86	WANITA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU
87	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU
88	PRIA	BEKERJA	В	YA	TERLAMBAT	TEPAT WAKTU
89	WANITA	MAHASISWA	В	YA	TEPAT WAKTU	TEPAT WAKTU
90	WANITA	BEKERJA	В	YA	TERLAMBAT	TEPAT WAKTU
91	PRIA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU
92	PRIA	MAHASISWA	C	TIDAK	TERLAMBAT	TEPAT WAKTU
93	WANITA	BEKERJA	В	YA	TEPAT WAKTU	TEPAT WAKTU
94	PRIA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
95	WANITA	MAHASISWA	В	YA	TEPAT WAKTU	TEPAT WAKTU
96	PRIA	MAHASISWA	С	TIDAK	TERLAMBAT	TEPAT WAKTU
97	PRIA	MAHASISWA	В	YA	TERLAMBAT	TEPAT WAKTU
98	WANITA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU
99	PRIA	BEKERJA	A	YA	TEPAT WAKTU	TEPAT WAKTU

4							
	100	WANITA	MAHASISWA	A	YA	TEPAT WAKTU	TEPAT WAKTU

Perhitungan Menggunakan Metode Naive Bayes

Pada kasus ini, data training bisa dilihat pada Tabel 3, data training sendiri terdiri dari 4 atribut yaitu, Gender, Status Mahasiswa, Nilai, dan Status Beasiswa dan juga terdapat 2 kelas yaitu Tepat Waktu dan Terlambat. Selanjutnya, dilakukan proses untuk menghitung jumlah kelas/label "P(H)" dihitung untuk probabilitas masing-masing kelas:

a. P(Tepat Waktu) = Jumlah data "Tepat Waktu" dibagi Jumlah data keseluruhan = 55:100 = 0,55 = 55%

b. P(Terlambat) = 45:100 = 0.45 = 45 %

Tabel 2. Probabilitas Kelas Tepat waktu & Terlambat

P(TEPAT/TERLAMBAT)	55%	45%

Tabel 4 menunjukkan bahwa probabilitas Kelas Tepat sebesar 55% dan Kelas Terlambat sebesar 45%. Proses berikutnya adalah menghitung jumlah kasus per kelas "P(X|H)", Berdasarkan Tabel 5 dihitung probabilitas per kelas pada tiap atributnya. Tabel 6 adalah proses mengalikan semua variabel kelas dengan hasil probabilitas yang telah didapatkan.

P(TEPAT/TERLAMBAT) TEPAT WAKTU P(GENDER) WANITA 53% 40% **PROBABILITAS** P(STATUS MAHASISWA) TEPAT TERLAMBAT 55% 51% 45% JUMLAH 100% P(NILAI) TEPAT TERLAMBAT 62% 0% 38% 33% JUMLAH 100% 100% P(BEASISWA) TEPAT TERLAMBAT 100% 33%

Tabel 5. Probabilitas Masing-Masing Atribut

Tabel 6. Hasil Perhitungan Kelas Tepat Waktu & Terlambat

					.1		
NO	GENDER	STATUS MAHASISWA	NILAI	BEASISWA	LAMA STUDI	TEPAT WAKTU	TERLAMBAT
1	PRIA	MAHASISWA	Α	YA	TEPAT WAKTU	8,77%	0,00%
2	WANITA	BEKERJA	В	YA	TEPAT WAKTU	5,03%	0,98%
4	PRIA	MAHASISWA	Α	YA	TEPAT WAKTU	8,77%	0,00%
5	WANITA	MAHASISWA	Α	YA	TEPAT WAKTU	9,78%	0,00%
7	WANITA	MAHASISWA	В	YA	TEPAT WAKTU	6,04%	1,02%
10	WANITA	MAHASISWA	Α	YA	TEPAT WAKTU	9,78%	0,00%
11	WANITA	BEKERJA	Α	YA	TEPAT WAKTU	8,15%	0,00%
13	WANITA	MAHASISWA	Α	YA	TEPAT WAKTU	9,78%	0,00%
14	PRIA	BEKERJA	В	YA	TEPAT WAKTU	4,51%	1,47%
16	WANITA	MAHASISWA	A	YA	TEPAT WAKTU	9,78%	0,00%

Kemudian, dibandingkan hasil per kelas yang telah didapatkan untuk mendapatkan prediksinya. Berdasarkan Tabel 7 bisa diketahui hasil dari prediksi yang didapatkan dari perbandingan kelas Tepat Waktu dan Terlambat.

	•	
TEPAT WAKTU	TERLAMBAT	PREDIKSI
8,77%	0,00%	TEPAT WAKTU
5,03%	0,98%	TEPAT WAKTU
8,77%	0,00%	TEPAT WAKTU
9,78%	0,00%	TEPAT WAKTU
6,04%	1,02%	TEPAT WAKTU
9,78%	0,00%	TEPAT WAKTU
8,15%	0,00%	TEPAT WAKTU
9,78%	0,00%	TEPAT WAKTU
4,51%	1,47%	TEPAT WAKTU

Tabel 7. Perbandingan Kelas Tepat Waktu & Terlambat

Perhitungan Menggunakan Metode Decision Tree

Berdasarkan Tabel 2, maka dilakukan proses perhitungan *entropy*(1) dan *gain*(2) untuk menentukan akar *(root)* dari pohon keputusan dalam membantu menentukan masa studi mahasiswa. Hasil dari perhitungan entropy dan gain pada Node 1 dapat ditunjukkan pada Tabel 8.

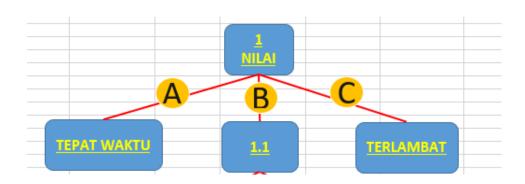
Node	Atribur	Kelas	Jumlah (S)	TEPAT WAKTU (Si)	Terlambat (Si)	Entrophy	Gain
1	Total		100	55	45	0,99277445	
	Gender						0,01165
		Pria	53	26	27	0,99974319	
		Wanita	47	29	18	0,96011866	
	Status Mahasiswa						0,00084
		Mahasiswa	53	30	23	0,98738002	
		Bekerja	47	25	22	0,99705906	
	Nilai						0,64002
		Α	34	34	0	0	
		В	36	21	15	0,97986876	
		С	30	0	30	0	
	Beasiswa						0,46805
		YA	70	55	15	0,74959526	
		Tidak	30	0	30	0	

Tabel 8. Hasil Perhitungan Gain dan Entrophy pada Node 1

Dari hasil Tabel 8, dapat diketahui bahwa atribut dengan gain tertinggi adalah "Nilai", yaitu 0,6400217. Dengan demikian atribut "Nilai" dapat dijadikan node akar. Dari hasil perhitungan nilai entropy dan gain pada node 1, maka dapat digambarkan decision tree sementara seperti terlihat pada Gambar 1.

	1 doct). Hasii i ciiii	ungan Gam da	in Entrophy pada	1 1 1 0 d C 1 . 2		
Node	Atribur	Kelas	Jumlah (S)	TEPAT WAKTU (Si)	Terlambat (Si)	Entrophy	Gain
1_1	Total		36	21	15	0,979869	
	Gender						0,045262
		Pria	14	6	8	0,985228	
		Wanita	22	15	7	0,902393	
	Status Mahasiswa						0,004135
		Mahasiswa	20	11	9	0,992774	
		Bekerja	16	10	6	0,954434	
	Beasiswa						0
		YA	36	21	15	0,979869	
		Tidak	0	0	0	0	

Tabel 9. Hasil Perhitungan Gain dan Entrophy pada Node 1.2

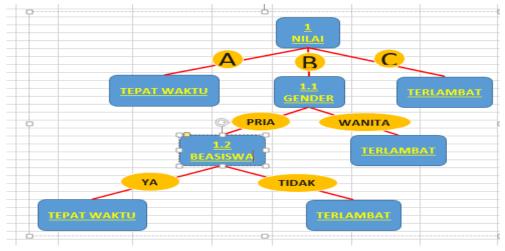


Gambar 1. Decision Tree Node 1

Ada 3 nilai atribut "Nilai" yaitu "A", "B", dan "C" dimana didapatkan nilai nilai "A" adalah "Tepat Waktu" nilai "C" adalah terlambat dan nilai "B" masih belum bisa diputuskan, oleh karena itu nilai B dihitung entrophy dan gain lebih lanjut, terlihat pada Tabel 9. Pada hasil Tabel 9 dapat diketahui bahwa atribut dengan gain tertinggi adalah "Gender" yaitu 0,045262. Dengan demikian "Gender" dapat menjadi node cabang dari "Nilai" yang bernilai "B". Dari Hasil perhitungan nilai entropy dan gain pada node 1.1, maka dapat digambarkan decision tree sementara seperti yang terlihat pada Gambar 2.



Gambar 1. Decision Tree Node 1.1



Gambar 2. Decision Tree Node 1.2

Pada Gambar 2 ada 2 nilai atribut "Gender" yaitu "Pria" dan "Wanita" dimana didapatkan nilai "Wanita" adalah "Terlambat" dan nilai "Pria" masih belum bisa diputuskan, oleh karena itu nilai "Pria" dihitung entrophy dan gain lebih lanjut, terlihat pada Tabel 10. Dari hasil Tabel 10 dapat diketahui bahwa atribut dengan gain tertinggi adalah "Beasiswa" yaitu 0,985228. Dengan demikian "Beasiswa" dapat menjadi node cabang dari "Nilai" yang bernilai "B" dan juga "Gender" yang bernilai pria. Dari Hasil perhitungan nilai entropy dan gain pada node 1.2, maka dapat digambarkan decision tree akhir seperti terlihat pada Gambar 3. Berdasarkan Gambar 3 maka didapatkan Decision Tree akhir dapat ditarik kesimpulan bahwa IF Nilai = B AND Gender = Pria AND Beasiswa = YA THEN Tepat Waktu.

Node	4.2	W-I	Levelah (5)	TER AT IMAKTIL (C)	Testander (57)	Formalis	Coin
Node	Atrībur	Kelas	Jumlah (S)	TEPAT WAKTU (Si)	Terlambat (Si)	Entrophy	Gain
1_2	Total		14	6	8	0,985228	
	Status Mahasiswa						0,00134
		Mahasiswa	5	2	3	0,970951	
		Bekerja	9	4	5	0,991076	
	Beasiswa						0,98522
		YA	9	9	0	0	
		Tidak	0	0	0	0	

Tabel 10. Perhitungan Gain dan Entrophy pada Node 1.2

Komparasi Akurasi Metode Naive Bayes dan Decision Tree

Menghitung tingkat akurasi metode naive bayes dengan cara membandingkan data yang sudah ada dan juga prediksi yang telah dihitung. Sedangkan Menghitung tingkat akurasi metode decision tree dengan cara membandingkan data yang sudah ada dan juga prediksi yang telah dihitung. Berdasarkan perhitungan yang telah ditentukan dengan menggunakan 100 data sample mahasiswa yang diambil secara acak sebagai bahan perhitungan maka didapatkan hasil akhir yaitu perhitungan menggunakan Metode Decision Tree mendapatkan hasil dengan tingkat akurasi 55%, dan perhitungan menggunakan Metode Naive Bayes mendapatkan tingkat akurasi sebesar 30%. Berdasarkan perhitungan tersebut dapat diketahui bahwa perhitungan menggunakan Metode Decision Tree lebih akurat.

KESIMPULAN

Berdasarkan perhitungan yang telah ditentukan dengan menggunakan 100 data sample yang sama dari masing-masing metode sebagai bahan perhitungan maka didapatkan hasil akhir yaitu perhitungan menggunakan Metode Decision Tree mendapatkan hasil dengan tingkat akurasi 55%, dan perhitungan menggunakan Metode Naive Bayes mendapatkan tingkat akurasi sebesar 30%. Dari perhitungan tersebut dapat disimpulkan bahwa tingkat akurasi Metode Decision Tree lebih tinggi dibandingkan tingkat akurasi Metode Naive Bayes, maka penulis memberi saran kepada penelitia data mining selanjutnya untuk lebih menggunakan Metode Decision Tree daripada Metode Naive Bayes.

REFERENSI

- [1] D. Marutho, "PERBANDINGAN METODE NAÏVE BAYES , KNN , DECISION TREE PADA LAPORAN WATER LEVEL JAKARTA," pp. 90–97, 2019.
- [2] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, 2013.
- [3] S. Defiyanti and D. L. Crispina Pardede, "Perbandingan kinerja algoritma id3 dan c4.5 dalam klasifikasi spam-mail," *ReCALL*, 2008.

- [4] K. Hastuti, "Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif," *Semin. Nas. Teknol. Inf. Komun. Terap.*, 2012.
- [5] P. G. S. C. Nugraha, I. W. Aribawa, I. P. O. Priyana, and G. Indrawan, "PENERAPAN METODE DECISION TREE(DATA MINING) UNTUK MEMPREDIKSI TINGKAT KELULUSAN SISWA SMPN1 KINTAMANI," *Semin. Nas. Vokasi dan Teknol.*, 2016.
- [6] D. Dicky Nofriansyah, "Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di," *Saintikom*, 2016.
- [7] A. Fathan Hidayatullah, M. Rifqi Ma, and arif Program Studi Manajemen Informatika STMIK Jenderal Achmad Yani Yogyakarta Jl Ringroad Barat, "Penerapan Text Mining dalam Klasifikasi Judul Skripsi," *Semin. Nas. Apl. Teknol. Inf. Agustus*, pp. 1907–5022, 2016.
- [8] Rismayanti, "Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus: Fakultas Teknik dan Komputer Universitas Harapan Medan)," *Query*, 2018.
- [9] S. Wahyuningsih, "Perbandingan Metode K-Nearest Neighbor, Naïve Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit," *Konf. Nas. Sist. Inf.* 2018, 2018.
- [10] "PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DENGAN METODE DECISION TREE (C4.5) UNTUK MENGANALISA KELANCARAN PEMBIAYAAN (Study Kasus: KSPPS / BMT AL-FADHILA," J. Teknol. Inf. Magister Darmajaya, 2016.