

BAB III

HASIL ANALISIS DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang dikumpulkan untuk penelitian ini terdiri dari 3 macam data yaitu :

1) Data yudisium mahasiswa angkatan 2017-2019

Data yudisium atau data kelulusan diperoleh dari arsip data yudisium Prodi Teknik Informatika dalam kondisi terpisah sesuai dengan tahun ajaran genap/ganjil. Data ini tersaji dengan NIM, Nama, Tempat, Tanggal Lahir, Tahun Masuk, IPK, dan Predikat Kelulusan dari mahasiswa angkatan 2017, 2018, dan 2019. Dari atribut tersebut dapat disimpulkan pula mana mahasiswa yang dikategorikan lulus tepat waktu dan terlambat. Berikut adalah contoh data yudisium mentah yang direkap untuk semester genap TA 2020/2021.

Tabel 3.1 Data Yudisium Prodi Teknik Informatika Semester Genap TA 2020/2021

NIM	NAMA	TEMPAT, TANGGAL LAHIR	TAHUN MASUK	IPK	PREDIKAT KELULUSAN
17111024410016	Normiatul Ilmiah	Pemangkih Seberang, 11 Nopember 1993	2017	3.81	Sangat Memuaskan
17111024410003	Andi Devina Safira Yasin	Muara Jawa, 7 Agustus 1999	2017	3.67	Sangat Memuaskan
17111024410001	Abdul Basir Salsabila	Samarinda, 16 September 1999	2017	3.57	Sangat Memuaskan
17111024400014	Muhammad Sandhyanur	Balikpapan, 118 April 1998	2017	3.56	Sangat Memuaskan
17111024410012	Muhammad Ansyori	Sebulu, 06 Juni 1997	2017	3.56	Sangat Memuaskan
17111024410007	Hidayatullah	Samarinda, 03 September 1999	2017	3.54	Sangat Memuaskan
17111024410009	Malik Fajar	Balikpapan, 13 Mei 1998	2017	3.49	Sangat Memuaskan
17111024410002	Agus Saswono	Samarinda, 19 Agustus 1997	2017	3.48	Sangat Memuaskan

2) Data akademik mahasiswa angkatan 2017- 2019

Untuk data akademik mahasiswa angkatan 2017-2019 terdiri dari No., Nama, NIM, Program Studi, Semester, SKS, IPS,SKS Kumulatif, IP Kumulatif, dan status mahasiswa. Data ini diperoleh secara langsung dari SIKAD Prodi Teknik Informatika Universitas Muhammadiyah Kalimantan Timur. Data ini diperoleh

dalam keadaan terpisah berdasarkan semesternya, yakni dari semester 1 hingga semester 7. Jumlah data akademik tahun 2017 adalah 21 data, 2018 sebanyak 119 dan 2019 sebanyak 181. Berikut adalah contoh gambaran data mentah yang didapat untuk data akademik 2017-2019.

Tabel 3.2 Data akademik 2017 Semester 1

No	NIM	NAMA	PRODI	Sem	SKS	IPS	SKSK	IPK	STATUS
1	17111...01	ABDUL BAS...	55201 - TI (S1)	20171	20	3.3250	20	3.3250	Lulus
2	17111...02	AGUS SAS...	55201 - TI (S1)	20171	20	3.4500	20	3.4500	Lulus
3	17111...03	ANDI DEV...	55201 - TI (S1)	20171	20	3.5500	20	3.5500	Lulus
4	17111...04	ANDI M. GIAS	55201 - TI (S1)	20171	20	3.0250	20	3.0250	Dikeluarkan
5	17111...05	AWANG AS...	55201 - TI (S1)	20171	20	3.2000	20	3.2000	Aktif
...
21	17111...21	YUSUF AR...	55201 - TI (S1)	20171	20	3.7250	20	3.7250	Lulus

Tabel 3.3 Data akademik 2018 Semester 1

No	NIM	NAMA	PRODI	Sem	SKS	IPS	SKSK	IPK	STATUS
1	18111...01	ACHMAD RUSS...	55201 - TI(S1)	20181	20	3.0250	20	3.0250	Dikeluarkan
2	18111...02	ADI ARIZALDY	55201 - TI(S1)	20181	20	2.8000	20	2.8000	Dikeluarkan
3	18111...03	AFIFAH FIRANTI	55201 - TI(S1)	20181	20	3.3000	20	3.3000	Aktif
4	18111...04	AGUS WIJAYA K.	55201 - TI(S1)	20181	20	1.0000	20	1.0000	Dikeluarkan
5	18111...05	AGUS FIRM...	55201 - TI(S1)	20181	20	2.9500	20	2.9500	Lulus
...
119	18111...1124	ARIEF FIRMANU..	55201 - TI(S1)	20181	19	3.3158	19	3.3158	Lulus

Tabel 3.4 Data akademik 2019 Semester 1

No	NIM	NAMA	PRODI	Sem	SKS	IPS	SKSK	IPK	STATUS
1	19111...01	AULIA RAHMAN	55201 - TI(S1)	20191	19	3.3947	19	3.3947	Dikeluarkan
2	19111...02	FATURRAHMAN	55201 - TI(S1)	20191	20	3.5750	20	3.5750	Aktif
3	19111...03	INDRIAN MUHAMMAD	55201 - TI(S1)	20191	20	3.7750	20	3.7750	Lulus
4	19111...04	SAP... FAHKRURI	55201 - TI(S1)	20191	20	3.4000	20	3.4000	Lulus
5	19111...05	MANG...	55201 - TI(S1)	20191	20	0.1000	20	0.1000	Dikeluarkan
...
181	19111...1188	MUHAMMAD ALI	55201 - TI(S1)	20191	19	3.8684	19	3.8684	Dikeluarkan

3) Data akademik mahasiswa angkatan 2020

Data ini kurang lebih mencakup informasi yang sama dengan data 2017-2019, dimana data ini juga diperoleh secara terpisah berdasarkan semesternya. Data ini

diperoleh dari SIKAD Program Studi Teknik Informatika dengan total sebanyak 241 data.

Tabel 3.5 Data akademik 2020 Semester 1

No	NIM	NAMA	PRODI	Sem	SKS	IPS	SKSK	IPK	STATUS
1	20111..01	AHLADA AL-QL..	55201 - TI (S1)	20201	20	3.3750	20	3.3750	Dikeluarkan
2	20111...02	SITI FA...	55201 - TI (S1)	20201	20	3.4750	20	3.4750	Dikeluarkan
3	20111...04	SURYADI	55201 - TI (S1)	20201	20	3.5000	20	3.5000	Dikeluarkan
4	20111...06	M. FAHRIZAL F. MUHAMMAD	55201 - TI (S1)	20201	20	2.2000	20	2.2000	Dikeluarkan
5	20111...07	RAJ...	55201 - TI (S1)	20201	20	3.1750	20	3.1750	Non-Aktif
...
241	20111...1250	VICTOR BENIEL C.	55201 - TI (S1)	20201	19	3.8684	19	3.8684	Aktif

3.2 Pra-proses data

Tahap pra-proses yang dilakukan untuk data pada penelitian ini terdiri dari penggabungan data dan *cleaning data*. Berikut adalah penjelasan lebih lanjut mengenai tahap pra-proses pada penelitian ini :

3.2.1 Penggabungan data

Pertama data akan melalui proses penggabungan data yang mana adalah menggabungkan data akademik angkatan 2017 – 2019 dari semester 1 sampai dengan 7 dan melakukan beberapa penyesuaian. Setelah itu data ini akan digabungkan dengan data yudisium, seperti yang telah diketahui bahwa data yudisium diperlukan untuk pelabelan data karena bentuk datanya adalah kategori yakni kategori lambat atau tepat. Jumlah data yang telah tergabung menjadi 321 data.

Tabel 3.6 Data gabungan angkatan 2017-2019

No	NIM	...	Se..2	SKS2	IPS 2	SKSK2	IPK 2	...	Stat..	LULUS
1	171...	...	20172	19	3.4474	39	3.3846	...	Lulus	TEPAT
2	171...	...	20172	19	3.6053	39	3.5256	...	Lulus	TEPAT
3	171...	...	20172	19	3.3947	39	3.4744	...	Lulus	TEPAT
...
22	181...	...	20182	19	3.0526	39	3.0385	...	Dike..	?
24	181...	...	20182	19	0.5263	39	1.6923	...	Dike	?

213	1911102441186	20	...	3.3	22	3.4545	LAMBAT
-----	---------------	----	-----	-----	----	--------	--------

Tabel 3.9 Data final 2020

No.	NIM	SKS 1	IPS 1	...	SKS 7	IPS 7
1	2011102441008	20	2.975	...	23	3.8696
2	2011102441009	20	2.975	...	23	3.8043
3	2011102441012	20	3.675	...	24	3.6667
...
165	2011102441250	19	3.8684	...	18	0.8333

Terdapat perbedaan pada masing masing tabel final, yakni pada atribut status kelulusan seperti pada tabel 3.8. sedangkan pada tabel 3.9 hanya berisi informasi mengenai SKS dan IPS. Atribut tersebut berguna untuk pelabelan. Dengan informasi ini, proses analisis dan model prediksi yang akan dilakukan pada dataset ini akan lebih terarah dan efisien.

3.3 Data Training dan Data Testing

Dalam tahap ini, dataset akan dibagi menjadi dua yaitu data *training* dan data *testing* menggunakan fungsi *train_test_split()* dari modul *scikit-learn*. Proses ini membagi data *training* 70% dan data *testing* sebanyak 30%.

Tabel 3.10 Data *training*

No.	SKS 1	IPS 1	SKS 2	IPS 2	...	SKS 6	IPS 6	SKS 7	IPS 7
1	20	3.45	19	3.8158	...	23	3.8696	17	3.5
2	20	3.425	19	3.8684	...	23	4	14	3.7857
3	20	2.9	19	2.8421	...	18	2.6944	17	2
4	20	3	19	2.5	...	24	3.7917	22	3.3182
5	20	3.175	19	3.6053	...	18	3.1667	18	3.4167
...
149	20	3.575	19	3.8158	...	21	3.3571	17	3.5588

Tabel 3.11 Data *testing*

No.	SKS 1	IPS 1	SKS 2	IPS 2	...	SKS 6	IPS 6	SKS 7	IPS 7
1	20	3.275	19	3.6579	...	21	3.7857	17	3.5882
2	20	3.65	19	3.9211	...	23	3.9348	17	3.5882
3	20	3.75	19	3.5526	...	23	3.8043	17	3.8529
4	19	3.7632	21	3.6905	...	17	4	15	4
5	20	3.375	19	2.7368	...	21	3.6429	19	3.2105
...
64	20	3.2	19	3.7368	...	23	3.6304	14	3.5714

Berdasarkan hasil distribusi dengan rasio yang sudah disebutkan telah menghasilkan data *training* sebanyak 149 entri dan data *testing* berjumlah 64 entri. Dimana data *training* digunakan untuk melatih model, model akan belajar dari pola dan hubungan antara SKS dan IPS dari data ini. Sedangkan data *testing* digunakan untuk mengevaluasi kemampuan model dalam memprediksi atau generalisasi ke data yang belum pernah dilihat sebelumnya.

3.4 Algoritme C4.5

Penelitian ini pada dasarnya menggunakan algoritme C4.5 untuk mengklasifikasi kelulusan mahasiswa. Namun, berdasarkan jumlah data yang telah dikumpulkan yaitu sebanyak 213 data, akan memakan waktu yang lama jika melakukan perhitungan secara manual. Untuk itu perhitungan yang melibatkan semua data pada penelitian ini diproses menggunakan python dengan menggunakan *library scikit-learn*. Dimana *library scikit-learn* akan mengimplementasi algoritme C4.5 sebagai kelas *DecisionTreeClassifier*, dan hasilnya dapat dilihat pada halaman lampiran. Jadi tahap ini akan menunjukkan proses perhitungan manual untuk mencari akar dari algoritme C4.5 menggunakan data sampel random yang tersaji dalam bentuk tabel, dimana data ini sudah disesuaikan dengan kebutuhan perhitungan itu sendiri. Tabel 3.12 adalah data sampel yang akan digunakan untuk perhitungan manual.

Tabel 3.12 Data Sampel

IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	KELULUSAN
3.9211	4	3.9375	3.6667	4	TEPAT
3.8	4	3.9	3.9	4	TEPAT
3.6	3.3	3.3	4	4	LAMBAT
2.9	3.4	3.5	4	3.4	LAMBAT
3	1.9211	2.5588	3.3611	2.0833	LAMBAT

Berdasarkan Tabel 3.12, diketahui bahwa :

Jumlah Kasus (S) = 5

Jumlah Kasus TEPAT(YES) = 2

Jumlah Kasus LAMBAT (NO)= 3

Setelah menganalisa data sampel, maka $Entropy(S)$ dapat dihitung menggunakan rumus

2.1.

$$Entropy(S) = -\left(\left(\frac{2}{5}\right) \times \log_2 \frac{2}{5}\right) + \left(\left(\frac{3}{5}\right) \times \log_2 \frac{3}{5}\right)$$

$$Entropy(S) = -\left(\frac{2}{5} \times \log_2(-1,3219)\right) + \left(\frac{3}{5} \times \log_2(-0,7369)\right) = 0,9717$$

Selanjutnya adalah mengitung $Entropy$ dari masing- masing atribut. Untuk mempersingkat perhitungan, IPS 1 akan dijadikan sebagai contoh dalam perhitungan ini.

Tabel 3.13 Frekuensi IPS 1

IPS 1	Frequency	probability	$\log_2(\text{probability})$
3.9211	1	1/5	-2,3219
3.8	1	1/5	-2,3219
3.6	1	1/5	-2,3219
2.9	1	1/5	-2,3219
3	1	1/5	-2,3219

$$Entropy (IPS 1) = -\sum (p \times \log_2 p) = -(-2,3219 - 2,3219 - 2,3219 - 2,3219 - 2,3219) = 2,3219$$

Selanjutnya adalah menghitung $Gain$, $Split Information$, dan $Gain Ratio$.

$$Gain IPS 1 = Entropy(S) - Entropy (IPS 1) = 0,9709 - 2,3219 = -1,3510$$

$$SI (IPS 1) = -\sum (p \times \log_2 p) = -\left(\left(\frac{1}{5}\right) \times \log_2 \left(\frac{1}{5}\right) + \left(\frac{1}{5}\right) \times \log_2 \left(\frac{1}{5}\right) + \left(\frac{1}{5}\right) \times \log_2 \left(\frac{1}{5}\right) + \left(\frac{1}{5}\right) \times \log_2 \left(\frac{1}{5}\right) + \left(\frac{1}{5}\right) \times \log_2 \left(\frac{1}{5}\right)\right)$$

$$= 2,3219$$

$$GR = \frac{IG (IPS 1)}{SI (IPS 1)} = \frac{-1,3510}{2,3219} = -0,5821$$

Hasil untuk tiap-tiap atribut yang telah dilakukan perhitungan yang sama seperti pada atribut IPS 1 dapat dilihat pada Tabel 3.14.

Tabel 3.14 $Entropy$ dan IG tiap atribut

Attribute	Entropy	Gain	Split Information	Gain Ratio
IPS 1	2,3219	-1,3510	2,3219	-0,5821

IPS 2	1,5219	-0,5510	1,5219	-0,3621
IPS 3	2,0119	-1,0410	2,0119	-0,5180
IPS 4	2,0119	-1,0410	2,0119	-0,5173
IPS 5	1,5219	-0,5510	1,5219	-0,3621

Berdasarkan Tabel 3.14, atribut yang memiliki *Gain Ratio* tertinggi adalah IPS 2 dan IPS 5. Karena kedua atribut ini memiliki nilai yang sama, jadi hanya perlu memilih salah satunya saja. Untuk itu IPS 2 akan dijadikan sebagai akar yang kemudian akan dipisahkan berdasarkan atribut IPS 2 itu sendiri. Setelah itu algoritme ini akan diterapkan secara rekursif ke tiap-tiap cabang hingga kriteria perhentian telah terpenuhi. Setelah perhitungan telah memenuhi kriteria, tahap selanjutnya adalah membuat pohon keputusan.

IP Semester 2

```

|-- IP Semester 2 <= 3.3
|
|-- IP Semester 1
|
|-- IP Semester 1 <= 3.5
|
|-- IP Semester 4
|
|-- IP Semester 4 <= 2.5
|
|-- TIDAK LULUS
|
|-- IP Semester 4 > 2.5
|
|-- IP Semester 3
|
|-- IP Semester 3 <= 3.2
|
|-- LULUS
|
|-- IP Semester 3 > 3.2
|
|-- TIDAK LULUS
|
|-- IP Semester 1 > 3.5
|
|-- LULUS
|-- IP Semester 2 > 3.3
|
|-- LULUS

```

Berikut adalah hasil aturan pohon keputusan yang dibuat menggunakan semua data dan dijalankan menggunakan *library sklearn* pada pemrograman *python*.

Aturan Pohon Keputusan:

```
|--- IPS 7 <= 3.33
| |--- class: LAMBAT
|--- IPS 7 > 3.33
| |--- IPS 4 <= 3.46
| | |--- SKS 7 <= 18.50
| | | |--- IPS 6 <= 3.83
| | | | |--- SKS 4 <= 20.50
| | | | | |--- IPS 5 <= 3.38
| | | | | | |--- class: LAMBAT
| | | | | |--- IPS 5 > 3.38
| | | | | | |--- IPS 7 <= 3.90
| | | | | | | |--- IPS 7 <= 3.63
| | | | | | | | |--- IPS 7 <= 3.53
| | | | | | | | | |--- class: LAMBAT
| | | | | | | | | |--- IPS 7 > 3.53
| | | | | | | | | | |--- class: TEPAT
| | | | | | | | | |--- IPS 7 > 3.63
| | | | | | | | | | |--- class: LAMBAT
| | | | | | | | | |--- IPS 7 > 3.90
| | | | | | | | | | |--- IPS 1 <= 2.91
| | | | | | | | | | | |--- class: LAMBAT
| | | | | | | | | | |--- IPS 1 > 2.91
| | | | | | | | | | | |--- class: TEPAT
| | | | | | | |--- SKS 4 > 20.50
| | | | | | | | |--- class: TEPAT
| | | | |--- IPS 6 > 3.83
| | | | |--- IPS 4 <= 3.33
| | | | | |--- class: TEPAT
| | | | |--- IPS 4 > 3.33
| | | | | |--- IPS 2 <= 3.36
| | | | | | |--- class: TEPAT
| | | | | |--- IPS 2 > 3.36
| | | | | | |--- class: LAMBAT
| | |--- SKS 7 > 18.50
| | | |--- class: LAMBAT
|--- IPS 4 > 3.46
| |--- SKS 6 <= 19.00
| | |--- IPS 7 <= 3.97
| | | |--- IPS 7 <= 3.80
| | | | |--- class: LAMBAT
| | | |--- IPS 7 > 3.80
| | | | |--- class: TEPAT
| | |--- IPS 7 > 3.97
| | | |--- class: LAMBAT
|--- SKS 6 > 19.00
| |--- IPS 7 <= 3.93
| | |--- IPS 1 <= 3.71
| | | |--- IPS 6 <= 3.88
| | | | |--- IPS 5 <= 3.91
| | | | | |--- IPS 6 <= 3.66
| | | | | | |--- SKS 5 <= 15.50
| | | | | | | |--- class: TEPAT
| | | | | | |--- SKS 5 > 15.50
| | | | | | | |--- IPS 4 <= 3.64
| | | | | | | | |--- IPS 6 <= 3.06
| | | | | | | | | |--- class: LAMBAT
| | | | | | | | | |--- IPS 6 > 3.06
| | | | | | | | | | |--- truncated branch of depth 2
| | | | | | | | | |--- IPS 4 > 3.64
| | | | | | | | | | |--- class: LAMBAT
| | | | | | | |--- IPS 6 > 3.66
| | | | | | | | |--- IPS 3 <= 3.78
| | | | | | | | | |--- class: TEPAT
| | | | | | | | |--- IPS 3 > 3.78
| | | | | | | | |--- SKS 5 <= 19.50
| | | | | | | | | |--- class: LAMBAT
```


<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
LAMBAT	0.72	0.78	0.75	37
TEPAT	0.67	0.59	0.63	27

Menurut informasi yang ditampilkan pada Tabel 3.16, hasil akurasi yang diperoleh adalah sebesar **70,31%**. Meskipun hasil akurasi yang ditunjukkan berada di angka yang rendah, namun nilai ketepatan (*Precision*) dalam memprediksi kelulusan mahasiswa yang terlambat dapat dikatakan cukup baik karena berada di angka **72%**.

3.6 Prediksi Kelulusan

Setelah mengetahui hasil evaluasi dari model algoritme C4.5, tahap selanjutnya yaitu menerapkan model algoritme ke dalam data akademik mahasiswa angkatan 2020. Hasil dari prediksi kelulusan terhadap mahasiswa Teknik Informatika angkatan 2020 telah tersaji dalam Tabel 3.17.

Tabel 3.17 Prediksi kelulusan

No.	IPS 1	SKS 1	IPS 7	SKS 7	PREDIKSI
1	2.975	20	3.8696	23	LAMBAT
2	2.975	20	3.8043	23	LAMBAT
3	3.675	20	3.6667	24	LAMBAT
4	3.325	20	3.7857	14	TEPAT
5	3.925	20	4	11	TEPAT
6	3.7	20	4	16	TEPAT
7	4	20	4	14	TEPAT
8	3.675	20	3.925	20	LAMBAT
9	4	20	4	16	TEPAT
10	3.55	20	3.8043	17	TEPAT
...
165	16	3.4524	0.8333	18	LAMBAT

Berdasarkan hasil prediksi kelulusan mahasiswa pada Tabel 3.17 menunjukkan bahwa model algoritme C4.5 pada penelitian ini telah memprediksi 118 mahasiswa yang kemungkinan akan lulus lambat, di sisi lain terdapat 47 mahasiswa diprediksi akan lulus tepat waktu.