

BAB III

METODE PENELITIAN

3.1. Objek Penelitian

Objek dalam penelitian ini merupakan data lulus tepat waktu yang diambil pada bulan september tahun 2020 dari salah satu *online repository* yang di akses pada situs www.kaggle.com, dengan menggunakan data yang dipakai 524 *dataset* dengan 14 atribut dan 1 label yaitu keterangan lulus. Dari variabel yang telah diketahui akan mengetahui apakah mahasiswa akan lulus tepat waktu atau tidak. Contoh dataset kelulusan tepat waktu yang akan digunakan terdapat pada Gambar 3. 1.

NAMA	JENIS KELAMIN	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK	STATUS KELULUSAN
STEVEN JULIAN	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	3.17	2.7	3.23	2.41	3	2.47	1.75	0	2.75	TEPAT
LEYLA TRIYANA PRATIWI	PEREMPUAN	MAHASISWA	26	BELUM MENIKAH	3.6	3.5	3.42	2.85	3.31	2.95	2.18		3.39	TEPAT
VERIS SOFIYAN PRAYOGA	LAKI - LAKI	MAHASISWA	29	BELUM MENIKAH	2.67	2.66	2.93	3.14	2.92	2.64	2.88	0.5	2.81	TEPAT
ADITYA AKBAR NUGRAHA	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	2.48	2.86	2.09	2.55	2.55	2.43	2.55	2.17	2.82	TEPAT
ERNA EKA RIYANTI	PEREMPUAN	MAHASISWA	25	BELUM MENIKAH	3.19	3.08	3.31	2.83	3.36	2.73	3.06	0	3.09	TEPAT
FARID DWI NORRYANTO	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	3.1	2.98	3.17	3.25	3.41	3.08	3.43	3	3.23	TEPAT
DAFIK HADI WNOTO	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	2.98	2.68	2.23	2.86	2.25	2.64	1.52	2.1	2.54	TEPAT
WAHYU FITRIYANTO	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	3.45	3.15	3.54	3.78	3.42	3.88	2.5	4	3.56	TEPAT
IMAM SURYO SUSANTO	LAKI - LAKI	MAHASISWA	26	BELUM MENIKAH	3.31	3.02	3.48	3.7	3.19	3.21	2.58	4	3.4	TEPAT
AHMAD SUTOPO	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	2.62	2.73	2.11	3.36	2.79	3.17	2.45	0	2.97	TEPAT
HERI SUFA'AT	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	3.24	3.06	2.79	2.73	3.02	2.59	3.08	0	3.08	TEPAT
MURYA AMIEN NUR PRABOWO	LAKI - LAKI	MAHASISWA	26	BELUM MENIKAH	2.83	3.02	3.02	3.63	3.21	3.09	2.57	3	3.33	TEPAT
DWI HESTYNA PRIHASTANTY	PEREMPUAN	MAHASISWA	32	BELUM MENIKAH	3	3.3	3.14	3.14	2.84	3.13	3.25	0	3.17	TERLAMBAT
MURYA ARIEF BASUKI	PEREMPUAN	BEKERJA	29	BELUM MENIKAH	3.5	3.3	3.7	3.29	3.53	3.72	3.73	0	3.54	TERLAMBAT
MARTHINA FLAVERIA	LAKI - LAKI	BEKERJA	25	MENIKAH	3.05	2.92	2.27	2.65	3	3.45	3.24	2	3.09	TERLAMBAT
MASLUH JAMIL	LAKI - LAKI	BEKERJA	30	BELUM MENIKAH	2.64	3	2.9	2.77	3.25	3.5	3.22	2	3.08	TERLAMBAT
MURDIANTORO	LAKI - LAKI	MAHASISWA	32	BELUM MENIKAH	2.45	2.45	2.14	1.95	1.82	2.66	1.61	0	2.32	TERLAMBAT
...
NIMATUL JANNAH	PEREMPUAN	MAHASISWA	23	BELUM MENIKAH	3.02	2.94	3.25	2.87	3	2.94	3.09	3	3.16	TEPAT
DINDU SETYO WICAKSONO	LAKI - LAKI	MAHASISWA	23	BELUM MENIKAH	3.1	3.06	3	3.23	2.79	3	2.41	3	2.16	TEPAT

Gambar 3. 1 *sample data*

Pada Gambar 3. 1 merupakan *sample data* permasalahan kelulusan tepat waktu sebelum dilakukannya transformasi data.

Tabel 3. 1 merupakan kegiatan dalam penelitian tugas akhir. Dalam pengumpulan data dilakukan pada bulan Desember, analisis data dilakukan pada bulan Desember sampai Januari, pengolahan data yang dilakukan pada bulan Januari sampai Februari, Pengujian yang dilakukan pada bulan Maret sampai April, evaluasi dilakukan pada bulan Juli, sedangkan keseluruhan dalam mengerjakan tugas akhir dilakukan pada bulan November sampai Mei.

Tabel 3. 1 Tabel penelitian

Kegiatan	Bulan							
	Des	Jan	Feb	Mar	Apr	Mei	Jun	Jul
Pengumpulan data	■							
Analisis data	■	■						
Pengolahan data			■	■				
Pengujian					■	■		
Evaluasi							■	■
Tugas Akhir	■	■	■	■	■	■	■	■

3.1.1. Peralatan Penelitian

1. Hardware yang digunakan dalam penelitian ini menggunakan laptop dengan spesifikasi:
 - a) Processor : Intel(R) Core (TM) i5-8250U CPU @ 1.60GHz 180GHz
 - b) Memory Size : 8.00 GB
2. *Software*
 - a) Windows 10
 - b) Microsoft Word 2020

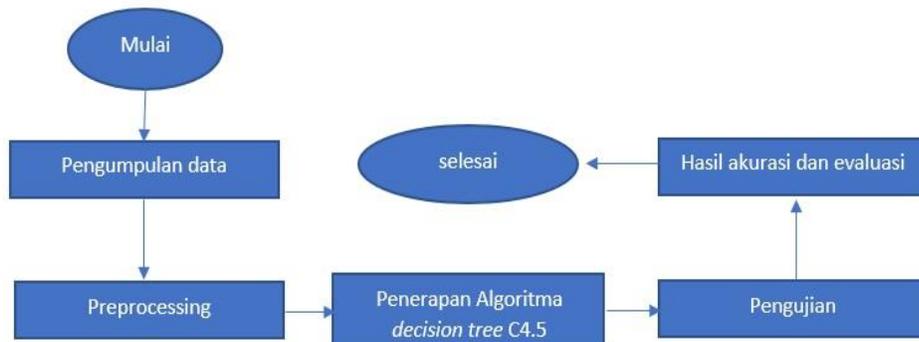
- c) Microsoft Excel 2020
- d) Weka
- e) Google Colaboratory (B)

3.1.2. Bahan penelitian

Bahan dari penelitian ini diambil dari *Website* Kaagle. Kaagle adalah komunitas yang bertujuan untuk belajar permasalahan data *science* dari *dataset*.

3.2. Prosedur Penelitian

Gambar 3.2 Alur Prosedur Penelitian merupakan langkah dalam melakukan penelitian yang dimulai dari pengumpulan data, *preprocessing*, pengaplikasian algoritma *decision tree* C4.5, pengujian, lalu mendapatkan hasil akurasi dan mengevaluasi.



Gambar 3. 2 Alur prosedur penelitian

3.2.1. Penerapan Algoritma *Decision Tree* C4.5



Pada Gambar 3.3 Tahapan Algoritma *decision Tree* menjelaskan tentang langkah-langkah pengaplikasian algoritma *decision tree* C4.5 yaitu :

- a) Melakukan pengujian atribut entropy dan gain.

Menghitung *entropy*

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (6)$$

Keterangan

S : Himpunan kasus.

N : Jumlah partisi dalam atribut.

Pi : Proporsi dari Si, pada S

Menghitung *gain*

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (7)$$

Keterangan

S : Himpunan kasus

A : Atribut

N : Jumlah partisi dalam atribut

SI : Jumlah kasus pada partisi ke-i

S : jumlah kasus S

- b) Mencari nilai gain tertinggi.
 c) Jika gain tertinggi tidak ditemukan maka kembali melakukan pengujian atribut ulang.
 d) Jika gain tertinggi ditemukan maka dapat diketahui bahwa partisi data tersebut sesuai dengan atribut.
 e) Gain tertinggi menjadi *root* atau *node* akar.
 f) Selanjutnya pembuatan *decision tree*.

3.2.2. Pengumpulan data

Penelitian ini diawali dengan pengumpulan data yang dilakukan secara *online* dan mendapatkan dataset tentang permasalahan kelulusan tepat waktu yang diakses pada situs <https://www.kaggle.com/datasets/stevenjulian/student-graduation->

dataset/code pada tahun 2021 yang terdiri dari 524 data dan 15 atribut, dan keterangan lulus mahasiswa.

3.2.3. *Preprocessing*

Preprocessing merupakan langkah yang penting karena tahapan ini bertujuan untuk mempermudah dalam menganalisa atau membaca data untuk menerapkan algoritma *decision tree* C4.5. Dalam tahapan *preprocessing* data yaitu Memilih atribut yang akan digunakan, melakukan transformasi data dari numerik pada kategorikal.

3.2.4. Pengujian

Dalam tahap pengujian ini bertujuan untuk mengetahui tingkat akurasi, presisi dan *recall* dari hasil klasifikasi yang akan diperoleh. Dalam pengujian ini terdapat tiga kali tahapan pengujian diantaranya yaitu :

- a) Melakukan perhitungan manual dengan Excel.
- b) Melakukan pengujian dengan menggunakan Weka.
- c) Mengimplementasikan algoritma *decision tree* C4.5 dengan menggunakan Google Colaboratory.

3.2.5. Hasil akurasi dan evaluasi.

Mengetahui hasil akurasi, presisi, dan *recall* dan mengevaluasi dari hasil klasifikasi yang didapatkan dari Algoritma *decision tree* C4.5.