

BAB 3

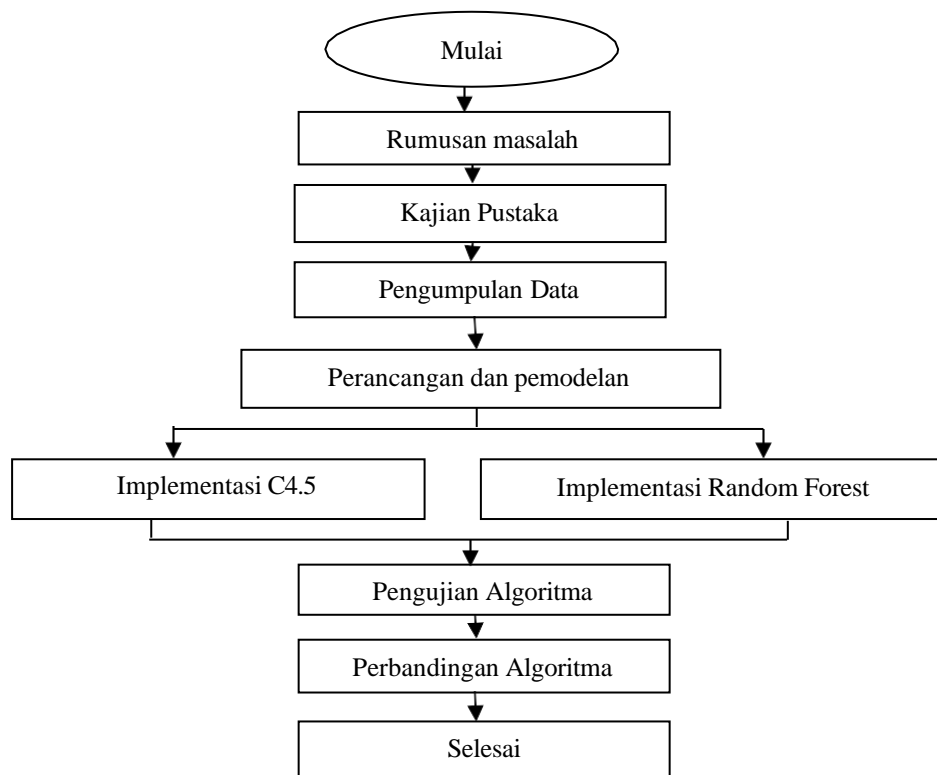
METODOLOGI PENELITIAN

3.1 Tempat dan Waktu Penelitian

Penelitian ini berlokasi di Fakultas Ekonomi dan Bisnis Islam IAIN Metro yang beralamat di Kampus 2 Jalan Ki Hajar Dewantara desa Banjar Rejo kecamatan Batanghari Kabupaten Lampung Timur. Pelaksanaan penelitian dimulai Bulan September 2022 – Maret 2023.

3.2 Tahapan penelitian

Proses penelitian dalam mengembangkan data mining mengikuti beberapa tahapan untuk melakukan analisa, mencari pola data kemudian dijadikan sebuah dataset sehingga memudahkan penelitian dan berjalan dengan sistematis serta memenuhi tujuan yang diinginkan maka dibuat bagan alur dalam tahapan penelitian yang akan dilakukan seperti pada gambar 3.1:



Gambar 3.1 Tahapan alur penelitian

Penjelasan tiap tahapan pada gambar 3.1 secara rinci dapat dijabarkan sebagai berikut:

3.2.1 Rumusan masalah

Pada tahap ini, penulis mencoba membuat klasifikasi dan menerapkan klasifikasi *data mining* pada data pendidikan untuk memprediksi masa kelulusan mahasiswa.

3.2.2 Kajian Pustaka

Pada tahap ini, penulis mencoba memberi informasi kepada pembaca hasil-hasil penelitian lain yang berkaitan erat dengan penelitian yang dilakukan peneliti, menghubungkan penelitian dengan literatur-literatur yang ada dan

mengisi celah-celah penelitian-penelitian sebelumnya yang menggunakan algoritma c4.5 dan Random Forest.

3.2.3 Pengumpulan data/eksplorasi data

Pada tahap ini merupakan tahap untuk mendapatkan data dari suatu informasi. Adapun langkah-langkah dalam proses pengumpulan data adalah sebagai berikut:

3.2.3.1 Sumber data

Pada penelitian ini menggunakan sumber data sekunder yaitu data alumni mahasiswa Fakultas Ekonomi Dan Bisnis IAIN Metro tahun Lulus 2021 dan 2022 sebanyak 646 record.

3.2.3.2 Metode pengumpulan data

Metode yang digunakan dalam proses pengumpulan data sekunder ini yaitu melalui metode dokumentasi. Metode dokumentasi ini mengambil sumber data penelitian atau objek dari dokumen atau catatan dari peristiwa yang sudah berlalu

3.2.3.3 Tahap pemahaman data

Data yang digunakan merupakan data sekunder yang berisi informasi data alumni mahasiswa Fakultas Ekonomi Dan Bisnis IAIN Metro tahun Lulus 2021 dan 2022. Data alumni tahun lulus 2021-2022 merupakan dataset yang masih harus diolah lagi menjadi data informasi yang dapat digunakan pada saat pengujian. Adapun atribut awal Mahasiswa Alumni terdiri dari 23 atribut predictor dan 1 atribut target. Atribut-atribut yang menjadi parameter terlihat pada tabel 3.1:

Tabel 3.1 Atribut Data

Atribut	Mahasiswa Alumni
	NPM
	Tahun Masuk
	Program Studi
	Nama Mahasiswa
	Tempat Tanggal Lahir
	Jenis Kelamin
	Agama
	Status Kerja
	Status Menikah
	Warga Negara
	Alamat
	Sekolah Asal
	Status Sekolah Asal
	Jalur Masuk
	Pendidikan Orang Tua
	Pekerjaan Orang Tua
	Penghasilan Orang Tua
	Indeks Prestasi (IP) Semester 1
	Indeks Prestasi (IP) Semester 2
	SKS-1
	SKS-2
	Tanggal Lulus

3.2.3.4 Persiapan data (*Data Preparation*)

Pada tahap ini berisi tentang tahap awal dari proses data mining. Data yang telah didapatkan akan diolah ke format yang dibutuhkan, pengelompokkan dan penentuan atribut atau variabel. Dalam proses pengolahan data awal akan dilakukan beberapa tahapan agar didapatkan data yang bisa digunakan untuk tahap selanjutnya.

1. Seleksi data (*data selection*)

Pada Tahap data selection atau pemilihan data merupakan tahap pemilihan atribut dari data yang akan dianalisis. Hal ini disebabkan karena tidak semua data yang terdapat dalam data mentah akan digunakan sehingga didapat beberapa atribut yang akan digunakan. Dasar Pemilihan atribut yang dilakukan dalam

penelitian ini dengan melihat kajian pustaka yang penulis gunakan pada bab II sebagai berikut :

Tabel 3.2 Seleksi Atribut Data

No	Atribut	Indikator Penilaian atribut
1	Kondisi demografis mahasiswa (Yazici, Seyis and Altun, 2011); (Oluwagbenga Abiodun and Isaiah, 2015); (Dayioğlu and Türüt-Aşik, 2007); (Kikas <i>et al.</i> , 2009); (Crawford, 2014).	1. Gender 2. Usia 3. Kota asal 4. Asal sekolah 5. Status Sekolah asal 6. Jalur masuk
2	Kondisi sosial ekonomi keluarga (Kikas <i>et al.</i> , 2009); (Crawford, 2014); (Jacobs and Harvey, 2005); (Li and Qiu, 2018); (Yazici, Seyis and Altun, 2011), (Kormos and Kiddle, 2013), mereka (American Pyschological Association, 2017).	1. Pendidikan orag tua 2. Pendapatan orang tua 3. Pekerjaan orang tua
3	Kinerja akademik mahasiswa (Dian Indriana TL, Amerti Irvin Widowati, 2017), (Shahiri, Husain and Nur'Aini Abdul Rashid, 2015).	1. IP-Semester 1 2. IP-Semster 2 3. SKS-1 4. SKS-2

Sementara untuk atribut NPM, Nama mahasiswa dan Program setudi hanya dijadikan identitas data sehingga dipilih salah satu yaitu NPM, sedangkan untuk atribut agama, Status kerja, status menikah, warga negara dan tahun lulus tidak digunakan dengan pertimbangan dianggap tidak memiliki pengaruh terhadap lamanya masa studi mahasiswa. Adapun hasil data selection seperti pada tabel 3.3:

Tabel 3.3 Hasil Seleksi Atribut Data

Atribut	Indikator	Detail Selection
NPM	√	Digunakan
Tahun Masuk	√	Digunakan
Program Studi	X	Tidak Digunakan
Nama Mahasiswa	X	Tidak Digunakan
TTL	√	Digunakan
Jenis Kelamin	√	Digunakan
Agama	√	Tidak Digunakan
Status Kerja	√	Tidak Digunakan
Status Menikah	√	Tidak Digunakan

Warga Negara	√	Tidak Digunakan
Alamat	√	Digunakan
Sekolah Asal	√	Digunakan
Status Sekolah Asal	√	Digunakan
Jalur Masuk	√	Digunakan
Pendidikan Orang Tua	√	Digunakan
Pekerjaan Orang Tua	√	Digunakan
Penghasilan Orang Tua	√	Digunakan
Indeks Prestasi (IP) Semester 1	√	Digunakan
Indeks Prestasi (IP) Semester 2	√	Digunakan
SKS-1	√	Digunakan
SKS-2	√	Digunakan
Tanggal Lulus	√	Digunakan

Secara rinci variabel prediktor atau atribut prediktor yang digunakan untuk membentuk model klasifikasi adalah sebagai berikut :

Tabel 3.4 Identifikasi Variabel

Variabel/atribut	Deskripsi Variabel/atribut	Tipe Data
X1	TTL	Ordinal
X2	Jenis Kelamin	Nominal
X3	Alamat	Nominal
X4	Sekolah Asal	Ordinal
X5	Status Sekolah Asal	Ordinal
X6	Jalur Masuk	Ordinal
X7	Pendidikan Orang Tua	Ordinal
X8	Pekerjaan Orang Tua	Ordinal
X9	Penghasilan Orang Tua	Ordinal
X10	IP-Semester 1	Ordinal
X11	IP-Semester 2	Ordinal
X12	SKS-1	Ordinal
X13	SKS-2	Ordinal
Y	Tanggal Lulus	Ordinal

2. Tahap Preprocessing Data

Tahap pre-processing merupakan tahap dilakukannya pembersihan terhadap data missing value yaitu data yang tidak konsisten atau kosong, dan juga dilakukannya pemilihan variabel terhadap data yang ingin digunakan pada proses data mining. Klasifikasi Kinerja akademik Mahasiswa akan menjadi keputusan lulus tepat waktu atau lulus tidak tepat waktu.

a. *Data Cleaning*

Pada tahap ini merupakan kelanjutan dari data selection. Pada tahap cleaning ini proses menghilangkan data yang memiliki informasi tidak lengkap atau missing data. Dalam proses data cleaning, peneliti menggunakan aplikasi microsoft excel dengan cara memfilter setiap kolom data yang kosong atau missing kemudian dilakukan proses penghapusan data yang kosong tersebut. Pada penelitian ini, file data alumni mahasiswa tahun lulus 2021-2022 tidak terdapat terdapat data missing sehingga jumlah data sesuai jumlah data asli yaitu sebanyak 646 record.

b. *Data transformation*

Variabel cenderung memiliki rentang yang sangat bervariasi satu sama lain. Untuk beberapa algoritme data mining, seperti perbedaan dalam rentang akan menyebabkan kecenderungan variabel dengan rentang yang lebih besar memiliki pengaruh yang tidak semestinya pada hasil (Larose, 2005). Pada penelitian ini dilakukan proses transformasi atribut asli menjadi atribut baru yaitu pada atribut Tempat tanggal lahir menjadi atribut Usia masuk, Alamat menjadi Kota Asal dan Atribut tanggal lulus menjadi Atribut Status Kelulusan. Sementara pada atribut lainnya hanya mengubah distribusi nilai asli menjadi satu set nilai baru dengan properti yang diinginkan. Berikut data transformasi pada masing-masing atribut dapat dilihat pada tabel 3.5 :

Tabel 3.5 Data Transformasi

Variabel/Atribut	Deskripsi Variabel/Atribut	Tipe Data	Nilai Data
X1	Usia Masuk	Ordinal	17 Tahun 18 Tahun 19 Tahun

			20 Tahun
X2	Jenis Kelamin	Nominal	Laki- Laki Perempuan
X3	Kota Asal	Nominal	Metro Luar Metro
X4	Sekolah Asal	Ordinal	SMA SMK MA
X5	Status Sekolah Asal	Ordinal	Negeri Swasta
X6	Jalur Masuk	Ordinal	SPAN-PTKIN UM-PTKIN Jalur Mandiri-PTKIN
X7	Pendidikan Orang Tua	Ordinal	Tidak Sekolah ≤SMA Diploma S1 S2
X8	Pekerjaan Orang Tua	Ordinal	Tidak bekerja Pensiun PNS Dosen Guru Pegawai Swasta Pengusaha/Wiraswasta Petani/Peternak Buruh Pedagang Nelayan Sopir Lainnya
X9	Penghasilan Orang Tua	Ordinal	≤Rp. 1.000.000 ≤Rp. 2.000.000 ≤Rp. 4.000.000 >Rp. 4.000.000
X10	IP-Semester 1	Ordinal	≤2.00 >2.00 - ≤ 3.00 >3.00 – 4.00
X11	IP-Semester 2	Ordinal	≤2.00 >2.00 - ≤ 3.00 >3.00 – 4.00
X12	SKS-1	Ordinal	<24 =24
X13	SKS-2	Ordinal	<24 =24
Y	Status Kelulusan	Nominal	Lulus tepat waktu Lulus tidak tepat waktu

c. Normalisasi Data

Pada Data yang dikumpulkan dalam kumpulan data terdapat kemungkinan tidak cukup berguna untuk algoritme Data Mining atau Biasanya atribut asli ini tidak cukup baik untuk mendapatkan model prediksi yang akurat, sehingga umum untuk melakukan serangkaian langkah manipulasi untuk mengubah atribut asli atau untuk menghasilkan atribut baru dengan properti yang lebih baik yang akan membantu kekuatan prediksi model (García, Luengo and Herrera, 2015). Pada penelitian ini, fokus pada transformasi yang tidak menghasilkan atribut baru, tetapi mengubah distribusi nilai asli menjadi satu set nilai baru dengan properti yang diinginkan.

Kemungkinan lain adalah Variabel yang cenderung memiliki rentang yang sangat bervariasi satu sama lain, sehingga untuk beberapa algoritma data mining dengan perbedaan dalam rentang yang lebih besar memiliki pengaruh yang tidak semestinya pada hasil (Larose and Larose, 2015). Oleh karena itu, Nilai-nilai atribut yang berbeda-beda rentangnya seringkali perlu dinormalisasi atau distandarisasikan agar proses data mining tidak terjadi bias. Adapun tiga teknik normalisasi yang sederhana dan efektif adalah sebagai berikut (Kantardics, 2020):

1) Normalisasi Min-Max

Pada metode ini menggunakan nilai minimum dan maksimum untuk melakukan konversi data secara linear. Normalisasi min-maks bekerja dengan melihat seberapa besar nilai bidang daripada nilai minimum $\min(X)$, dan menskalakan perbedaan ini dengan rentang (Larose and Larose, 2015). Berikut formula min-max:

$$X_{mm}^* \frac{X - \min(X)}{\text{range}(X)} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

Normalisasi data min-max banyak digunakan secara praktis. Namun, metode ini memiliki kelemahan pada dua sisi yaitu jika terdapat nilai pencilan yang mendominasi dan menjadi min-A (yang jauh lebih besar atau lebih kecil dibandingkan nilai-nilai lainnya) atau jika suatu saat ada nilai-nilai atribut yang lebih besar daripada maks-A atau lebih lebih daripada min-A maka nilai-nilai tersebut akan berada diluar rentang (minbaru-A, maxbaru-A)

2) Normalisasi Z-Score

Standardisasi Z-score yang sangat luas dalam dunia analisis statistik, bekerja dengan mengambil selisih antara nilai bidang dan nilai rata-rata bidang, dan menskalakan perbedaan ini dengan SD dari nilai bidang (Larose and Larose, 2015). Normalisasi dengan deviasi standar sering bekerja dengan baik dengan pengukuran jarak, tetapi mengubah data menjadi bentuk yang tidak dapat dikenali dari data aslinya (Kantardzic, 2020). Berikut formula Standardisasi Z-score (Larose and Larose, 2015):

$$Z - score = \frac{X - \text{mean}(X)}{SD(X)}$$

Pada metode normalisasi Z-Score lebih stabil terhadap pencilan maupun adanya nilainilai yang baru yang lebih besar daripada maks-A atau lebih kecil daripada min-A. Pada metode ini normalisasi nilai x_i pada atribut A menjadi nilai baru x_i^1 berdasarkan nilai rata-rata \bar{A} dan deviasi standar σA menggunakan formula $x_i^1 = \frac{x_i - \bar{A}}{\sigma A}$ dengan memanfaatkan nilai rata-rata dan deviasi standar.

3) Normalisasi dengan penskalaan Decimal

Cara sederhana untuk mengurangi nilai absolut dari atribut numerik adalah dengan menormalkan nilainya dengan menggeser titik desimal menggunakan pangkat sepuluh sehingga nilai absolut maksimum selalu lebih rendah dari 1 setelah transformasi (García, Luengo and Herrera, 2015). Penskalaan desimal memastikan bahwa setiap nilai yang dinormalisasi berada di antara -1 dan 1 (Larose and Larose, 2015). Berikut formula Normalisasi dengan penskalaan Decimal:

$$X_{decimal}^* = \frac{X}{10^d}$$

Di mana d mewakili jumlah digit dalam nilai data dengan nilai absolut terbesar.

3.2.4 Perancangan dan pemodelan Algoritma

Pada tahap ini dilakukan pemilihan teknik pemodelan yang tepat sesuai dataset yang telah dimiliki dengan jumlah 646 record data alumni mahasiswa. Pada penelitian ini, menggunakan model atau algoritma algoritma Decision Tree C4.5 dan Algoritma Random Forest Clasifier dalam RapidMiner sebagai solusi untuk melakukan analisis terhadap data mining, klasifikasi dan analisis prediksi.

Setelah melewati tahap persiapan pengolahan data maka dataset ini kemudian dibagi menjadi 2 bagian yaitu data training yang digunakan oleh algoritma klasifikasi dengan algoritma Decision Tree C4.5 dan Algoritma Random Forest Clasifier untuk membentuk sebuah model klasifikasi dan data

testing digunakan untuk mengukur sejauh mana model classifier berhasil melakukan klasifikasi dengan benar.

3.2.5 Implementasi dan Pengujian Algoritma

Pada tahap ini, data yang sudah melalui tahap *preprocessing* data kemudian diimplementasikan dan diuji menggunakan algoritma algoritma C4.5 dan Random Forest Classifier. Berikut sebagian contoh dataset pada penelitian ini yang diambil secara random sebanyak 14 record dataset sebagai ilustrasi perhitungan manual untuk mengetahui cara kerja algoritma C4.5 seperti pada tabel 3.6:

Tabel 3.6 Contoh Dataset Penelitian

No	NPM	Usia Masuk	JK	Kota Asal	Sekolah Asal	Status Sekolah	Jalur Masuk	Pendidikan Orang Tua	Pekerjaan Orang Tua	Penghasilan Orang Tua	IPS-1	IPS-2	SKS-1	SKS-2	Status Kelulusan
1	1804101075	17 Tahun	L	Luar Metro	SMK	Negeri	SPAN-PTKIN	≤SMA	Lainnya	≤1.000.000	3,58	3,33	24	24	ttw
2	1804101013	18 Tahun	L	Luar Metro	SMK	Swasta	SPAN-PTKIN	≤SMA	Wiraswasta	≤1.000.000	3,17	3,36	24	22	ttw
3	1804102045	19 Tahun	L	Metro	MA	Negeri	SPAN-PTKIN	Tidak Sekolah	Buruh	≤1.000.000	3,33	2,67	24	24	tw
4	1804100045	19 Tahun	L	Luar Metro	SMA	Negeri	UM-PTKIN	≤SMA	Buruh	≤1.000.000	3,17	3,32	24	22	tw
5	1804100060	18 Tahun	P	Luar Metro	SMA	Negeri	SELEKSI MANDIRI	≤SMA	Petani/Peternak	≤1.000.000	3,54	3,13	24	24	tw
6	1804101049	19 Tahun	P	Luar Metro	SMA	Swasta	SELEKSI MANDIRI	≤SMA	Petani/Peternak	≤1.000.000	3,5	3,54	24	24	ttw
7	1804022013	18 Tahun	P	Luar Metro	MA	Swasta	SELEKSI MANDIRI	≤SMA	Petani/Peternak	≤1.000.000	3,21	3,41	24	22	tw
8	1804100059	18 Tahun	L	Metro	SMK	Negeri	UM-PTKIN	≤SMA	Buruh	≤1.000.000	3,33	3,04	24	24	ttw
9	1804100052	18 Tahun	P	Metro	SMK	Negeri	SELEKSI MANDIRI	≤SMA	Buruh	≤1.000.000	3,42	3,54	24	24	tw
10	1804101004	19 Tahun	P	Luar Metro	SMA	Negeri	UM-PTKIN	≤SMA	Petani/Peternak	≤1.000.000	3,54	3,63	24	24	tw
11	1802040200	17 Tahun	P	Metro	SMK	Swasta	UM-PTKIN	≤SMA	Buruh	≤1.000.000	3,54	3,67	24	24	tw
12	1804011013	17 Tahun	L	Metro	MA	Swasta	UM-PTKIN	≤SMA	Wiraswasta	≤1.000.000	3,5	3,54	24	24	tw
13	1804020017	18 Tahun	P	Luar Metro	MA	Negeri	SPAN-PTKIN	≤SMA	Petani/Peternak	≤1.000.000	3,21	3,41	24	24	tw
14	1804101008	17 Tahun	L	Metro	SMA	Swasta	UM-PTKIN	≤SMA	Petani/Peternak	≤1.000.000	3,33	3,04	24	24	ttw

3.2.5.1 Implementasi Algoritma C4.5

Tahap proses perhitungan untuk memprediksi tingkat kelulusan mahasiswa menggunakan algoritma C4.5 adalah sebagai berikut :

Tahap 1 – Menghitung nilai Entropy total seluruh data berdasarkan kategori label kelas

Adapun perhitungan nilai Entropy pada contoh dalam penelitian ini adalah sebagai berikut:

Jumlah kasus (S) 14

Kasus tw (S₁) : 9

Kasus ttw (S₂) : 5

Maka nilai Entropy dapat dihitung dengan menggunakan persamaan berikut:

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

$$Entropy (Total) = (- \frac{9}{14} * (\log_2 \frac{9}{14})) + (- \frac{5}{14} * (\log_2 \frac{5}{14}))$$

$$Entropy (Total) = 0,94029$$

Tahap 2 – Menghitung Nilai Entropy, info gain, Split Information dan Gain Ratio atribut berdasarkan kelasnya

Perhitungan nilai Entropy, info gain, Split Information dan Gain Ratio pada contoh dalam penelitian ini adalah sebagai berikut:

a. Nilai Entropy atribut Usia Masuk

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

$$Entropy (17 Tahun) = (- \frac{2}{4} * (\log_2 \frac{2}{4})) + (- \frac{2}{4} * (\log_2 \frac{2}{4})) = 1$$

$$\text{Entropy (18 Tahun)} = \left(-\frac{4}{6} * (\log_2 \frac{4}{6})\right) + \left(-\frac{2}{4} * (\log_2 \frac{2}{4})\right) = 0,9183$$

$$\text{Entropy (19 Tahun)} = \left(-\frac{3}{4} * (\log_2 \frac{3}{4})\right) + \left(-\frac{1}{4} * (\log_2 \frac{1}{4})\right) = 0,81128$$

b. Nilai Info Gain atribut Usia Masuk

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

$$\begin{aligned} \text{Gain (Total, Usia Masuk)} &= \text{Entropy (S)} - \sum_{i=1}^n \frac{|S_i|}{S} * \text{Entropy (S}_i) \\ &= 0,94029 - \left(\frac{4}{14} * 1\right) + \left(\frac{6}{14} * 0,9183\right) + \left(\frac{4}{14} * 0,81128\right) = 0,02922257 \end{aligned}$$

c. Nilai Split Info atribut Usia Masuk

$$\begin{aligned} \text{Split Info (S,A)} &= -\sum_{j=1}^k \frac{S_j}{S} * \log_2 \frac{S_j}{S} \\ &= \left(-\frac{4}{14} * \log_2 \left(\frac{4}{14}\right)\right) - \left(\frac{6}{14} * \log_2 \left(\frac{6}{14}\right)\right) - \left(\frac{4}{14} * \log_2 \left(\frac{4}{14}\right)\right) = 1,37453 \end{aligned}$$

d. Nilai Gain Ratio Atribut Usia Masuk

$$\text{Gain Ratio (S,A)} = \frac{\text{Gain (S,A)}}{\text{Split (S,A)}}$$

$$\text{Gain Ratio (Total, Prodi)} = \frac{0,02922257}{1,37453} = 0,02126$$

Sementara untuk menghitung nilai atribut numerik yaitu IPS-1, IPS2, SKS-1 dan SKS-2 dengan cari nilai rata-ratanya seperti pada tabel 3.7:

Tabel 3.7 Nilai Rata-Rata Atribut Numerik

Atribut	Rata-Rata
IPS-1	3,38
IPS-2	3,33
SKS-1	24
SKS-2	23,57

Berdasarkan hasil perhitungan nilai rata-rata data numerik di atas, maka dapat dicari nilai Entropy, info gain, Split Information dan Gain Ratio atribut berdasarkan kelasnya dengan perhitungan yang sama.

Tabel 3.8 Hasil perhitungan nilai Entropy, info gain, Split Info dan Gain Ratio

Atribut	Nilai Atribut	Jumlah	TW	TTW	Entropy	Info Gain	Split Info	Gain Ratio
Usia Masuk						0,029222566	1,374534	0,02126
	17 Tahun	4	2	2	1			
	18 Tahun	6	4	2	0,918296			
	19 Tahun	4	3	1	0,811278			
JK						0,151835501	1	0,151836
	L	7	3	4	0,985228			
	P	7	6	1	0,591673			
Kota Asal						0,001339742	0,985228	0,00136
	Metro	6	4	2	0,918296			
	Luar Metro	8	5	3	0,954434			
Sekolah Asal						0,24674982	1,577406	0,156428
	MA	4	4	0	0			
	SMA	5	3	2	0,970951			
	SMK	5	2	3	0,970951			
Status Sekolah						0,04812703	0,985228	0,048849
	Negeri	8	6	2	0,811278			
	Swasta	6	3	3	1			
Jalur Masuk						0,029222566	1,556657	0,018773
	SPAN-PTKIN	4	2	2	1			
	UM-PTKIN	6	4	2	0,918296			
	SELEKSI MANDIRI	4	3	1	0,811278			
pendidikan Ayah						0,047709111	0,371232	0,128516
	Tidak Sekolah	1	1	0	0			
	<=SMA	13	8	5	0,961237			
Pekerjaan Ayah						0,146041996	1,727397	0,084545
	Buruh	5	4	1	0,721928			
	Petani/Peternak	6	4	2	0,918296			
	Wiraswasta	2	1	1	1			
	Lainnya	1	0	1	0			
Penghasilan Ayah						2,22045E-16	0	0
	>1.000.000	0	0	0	0			
	≤1.000.000	14	9	5	0,940286			
IPS-1						-0,08021093	0,961346	-0,08344
	>3,38	8	5	2	0,923795			
	≤3,38	7	4	3	0,985228			
IPS-2						-0,04897297	0,933659	-0,05245
	>3,33	9	6	2	0,872181			
	≤3,33	6	3	3	1			
SKS-1						-0,03021423	0	0
	24	15	9	5	0,9705			
	<24	0	0	0	0			
SKS-2						-0,09814194	0,666849	-0,14717
	>23,57	12	7	4	0,981925			
	≤23,57	3	2	1	0,918296			

Tahap 3 – Mencari Gain Ratio tertinggi dari pada masing-masing atribut

Berdasarkan hasil perhitungan gain ratio pada masing-masing atribut, maka

diperoleh gain ratio tertinggi pada atribut Sekolah Asal yaitu sebesar **0,156428**,

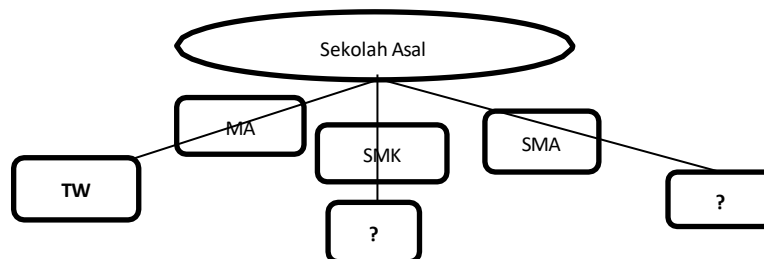
sehingga atribut Kota Asal dijadikan Node akar dari pohon keputusan. Atribut

Sekolah Asal memiliki 3 nilai atribut yaitu MA, SMA dan SMK dengan perbandingan gain ratio tertinggi seperti pada tabel 3.9:

Tabel 3.9 perbandingan gain ratio tertinggi

Gain Ratio Max	0,156428	Sekolah Asal	Jumlah	TW	TTW	Perbandingan		selisih
		MA	4	4	0	100%	0%	100%
		SMA	5	3	2	60%	40%	20%
		SMK	5	2	3	40%	60%	-20%

Dari hasil perbandingan dari kedua nilai atribut tersebut, nilai atribut MA sudah membentuk kelas yaitu tepat waktu maka tidak dilakukan perhitungan lebih lanjut. Sedangkan nilai atribut SMA dan SMK belum membentuk kelas sehingga perlu dilakukan perhitungan kembali dengan langkah dan cara yang sama. Berikut pohon keputusan Node 1 untuk atribut Sekolah Asal dapat digambarkan seperti pada gambar 3.2:



Gambar 3.2 Pohon Keputusan Node 1

➤ Atribut Sekolah Asal = SMK

Tahap 1 : Menghitung Nilai Entropy Total Seluruh Data Berdasarkan Kategori Label Kelasnya

Berikut hasil perhitungan nilai nilai entropy Total Seluruh Data Berdasarkan Sekolah Asal = SMK seperti pada tabel 3.10:

Tabel 3.10 Hasil Perhitungan Nilai Entropy Atribut Sekola Asal = SMK

Atribut	Jumlah	TW	TTW	Entropy
Total	5	2	3	0,970951

Tahap 2 – Menghitung Nilai Entropy Berdasarkan Sekolah Asal = SMK, Info Gain, Split Information dan Gain Ratio Pada Masing-Masing Atribut Berikut hasil perhitungan nilai entropy, info gain, split info dan gain ratio atribut data diskret berdasarkan Sekolah Asal = SMK dan untuk atribut numerik dengan nilai rata-rata seperti pada tabel 3.11:

Tabel 3.11 Nilai entropy, info gain, split info dan gain ratio Sekolah Asal = SMK

Atribut	Nilai Atribut	Jumlah	TW	TTW	Entropy	Info Gain	Split Info	Gain Ratio
Usia Masuk						0,019973	0	0
	17 Tahun	2	1	1	1			
	18 Tahun	3	1	2	0,918296			
	19 Tahun	0	0	0	0			
	20 Tahun	0	0	0	0			
JK						0,970951	0,970951	1
	L	3	0	3	0			
	P	2	2	0	0			
Kota Asal						0,419973	0,970951	0,43253807
	Metro	3	2	1	0,918296			
	Luar Metro	2	0	2	0			
Status Sekolah						0,019973	0,970951	0,02057066
	Negeri	3	1	2	0,918296			
	Swasta	2	1	1	1			
Jalur Masuk						0,570951	1,521928	0,37514952
	SPAN-PTKIN	2	0	2	0			
	UM-PTKIN	2	1	1	1			
	SELEKSI MANDIRI	1	1	0	0			
pendidikan Ayah						1,11E-15	0	0
	Tidak Sekolah	0	0	0	0			
	<=SMA	5	2	3	0,970951			
Pekerjaan Ayah						0,419973	0	0
	Buruh	3	2	1	0,918296			
	Petani/Peternak	0	0	0	0			
	Wiraswasta	1	0	1	0			
	Lainnya	1	0	1	0			
Penghasilan Ayah						1,11E-15	0	0
	>1.000.000	0	0	0	0			
	≤1.000.000	5	2	3	0,970951			
IPS-1						0,419973	0,970951	0,43253807
	>3,40	3	2	1	0,918296			
	≤3,40	2	0	2	0			
IPS-2						0,970951	0,970951	1
	>3,39	2	2	0	0			
	≤3,39	3	0	3	0			
SKS-1						1,11E-15	0	0
	24	5	2	3	0,970951			
	<24	0	0	0	0			
SKS-2						0,170951	0,721928	0,23679726
	>23,60	4	2	2	1			
	≤23,60	1	0	1	0			

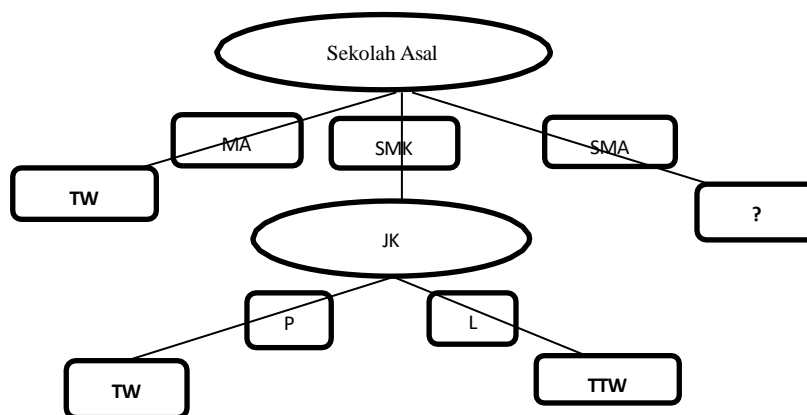
Tahap 3 – Mencari Gain Ratio tertinggi dari pada masing-masing atribut

Berdasarkan hasil perhitungan gain ratio pada masing-masing atribut, maka diperoleh gain ratio tertinggi pada atribut Jenis Kelamin yaitu sebesar **1**, sehingga atribut Jenis Kelamin dijadikan Node akar dari pohon keputusan. Atribut Jenis Kelamin memiliki 2 nilai atribut Laki-Laki dan Perempuan dengan perbandingan gain ratio tertinggi seperti pada tabel 3.12:

Tabel 3.12 perbandingan gain ratio tertinggi

Gain Ratio Max	1	Nilai Atribut	Jumlah	TW	TTW	Perbandingan		selisih
		L	3	0	3	0%	100%	-100%
		P	2	2	0	100%	0%	100%

Dari hasil perbandingan dari kedua nilai atribut tersebut, nilai atribut Laki-Laki sudah membentuk kelas yaitu tidak tepat waktu dan nilai Atribut Perempuan membentuk kelas tepat waktu sehingga tidak perlu dilakukan perhitungan kembali. Berikut pohon keputusan Node 1.1 untuk atribut Jenis Kelamin dapat digambarkan seperti pada gambar 3.3:



Gambar 3.3 Pohon Keputusan Node 1.1

➤ Atribut Sekolah Asal = SMA

Tahap 1 : Menghitung Nilai Entropy Total Seluruh Data Berdasarkan Kategori

Label Kelasnya

Berikut hasil perhitungan nilai entropy Total Seluruh Data Berdasarkan Sekolah Asal = SMA seperti pada tabel 3.13:

Tabel 3.13 nilai entropy Seluruh Data Berdasarkan Sekolah Asal = SMA

Atribut	Jumlah	TW	TTW	Entropy
Total	5	3	2	0,970951

Tahap 2 – Menghitung Nilai Entropy Berdasarkan Sekolah Asal = SMA, Info Gain, Split Information dan Gain Ratio Pada Masing-Masing Atribut Berikut hasil perhitungan nilai entropy, info gain, split info dan gain ratio atribut data diskret berdasarkan Sekolah Asal = SMA dan untuk atribut numerik dengan nilai rata-rata seperti pada tabel 3.14:

Tabel 3.14 Hasil Perhitungan Nilai Rata-Rata Atribut Numerik

Atribut	Rata-Rata
IPS-1	3,41
IPS-2	3,33
sks-1	24,00
sks-2	23,60

Berdasarkan hasil perhitungan nilai rata-rata data numerik di atas, maka dapat dicari nilai Entropy, info gain, Split Info dan Gain Ratio atribut berdasarkan kelasnya dengan perhitungan yang sama.

Tabel 3.15 nilai entropy, info gain, split info dan gain ratio Sekolah Asal = SMA

Atribut	Nilai Atribut	Jumlah	TW	TTW	Entropy	Info Gain	Split Info	Gain Ratio
Usia Masuk						0,419973	0	0
	17 Tahun	1	0	1	0			
	18 Tahun	1	1	0	0			
	19 Tahun	3	2	1	0,918296			
	20 Tahun	0	0	0	0			
JK						0,019973	0,970951	0,020571
	L	2	1	1	1			
	P	3	2	1	0,918296			
Kota Asal						0,321928	0,721928	0,445928
	Metro	1	0	1	0			
	Luar Metro	4	3	1	0,811278			
Status Sekolah						0,970951	0,970951	1
	Negeri	3	3	0	0			
	Swasta	2	0	2	0			
Jalur Masuk						0,019973	0	0
	SPAN-PTKIN	0	0	0	0			

	UM-PTKIN	3	2	1	0,918296			
	SELEKSI MANDIRI	2	1	1	1			
pendidikan Ayah						1,11E-15	0	0
	Tidak Sekolah	0	0	0	0			
	<=SMA	5	3	2	0,970951			
Pekerjaan Ayah						0,170951	0	0
	Buruh	1	1	0	0			
	Petani/Peternak	4	2	2	1			
	Wiraswasta	0	0	0	0			
	Lainnya	0	0	0	0			
Penghasilan Ayah						1,11E-15	0	0
	>1.000.000	0	0	0	0			
	≤1.000.000	5	3	2	0,970951			
IPS-1						0,019973	0,970951	0,020571
	>3,41	3	2	1	0,918296			
	≤3,41	2	1	1	1			
IPS-2						0,019973	0,970951	0,020571
	>3,33	2	1	1	1			
	≤3,33	3	2	1	0,918296			
SKS-1						1,11E-15	0	0
	24	5	3	2	0,970951			
	<24	0	0	0	0			
SKS-1						0,170951	0,721928	0,236797
	>23,60	4	2	2	1			
	≤23,60	1	1	0	0			

Tahap 3 – Mencari Gain Ratio tertinggi dari pada masing-masing atribut

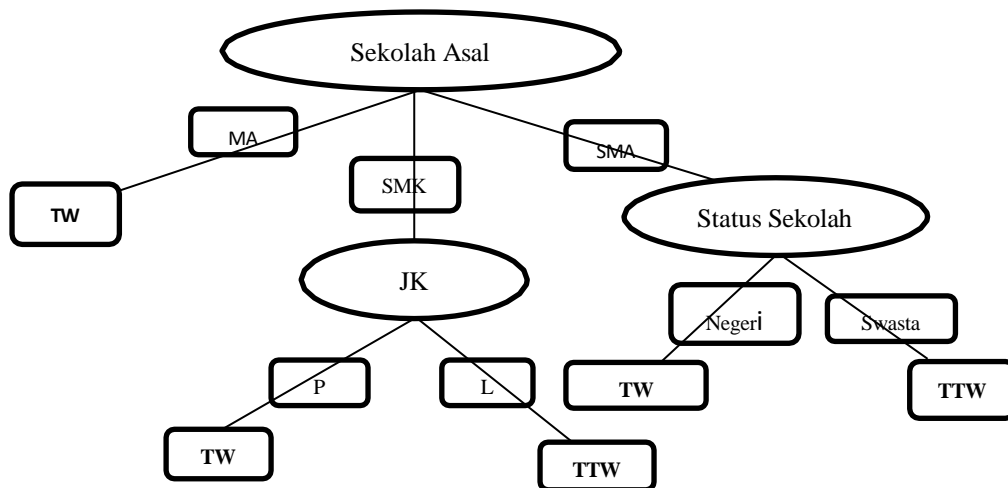
Berdasarkan hasil perhitungan gain ratio pada masing-masing atribut, maka diperoleh gain ratio tertinggi pada atribut Status Sekolah Asal yaitu sebesar **1**, sehingga atribut Status Sekolah Asal dijadikan Node akar dari pohon keputusan. Atribut Status Sekolah Asal memiliki 2 nilai atribut yaitu Negeri dan Swasta dengan perbandingan gain ratio tertinggi seperti pada tabel 3.16:

Tabel 3.16 Perbandingan Gain Ratio Tertinggi

Gain Ratio Max	1	Status Sekolah Asal	Jumlah	TW	TTW	Perbandingan		Selisih
		Negeri	3	3	0	100%	0%	100%
		Swasta	2	0	2	0%	100%	-100%

Dari hasil perbandingan dari kedua nilai atribut tersebut, nilai atribut Negeri dan Swasta sudah membentuk kelas maka tidak dilakukan perhitungan lebih lanjut.

Berikut pohon keputusan Node 1.2 untuk atribut Status Sekolah Asal dapat digambarkan seperti pada gambar 3.4:



Gambar 3.4 Pohon Keputusan Akhir

Role/aturan pohon keputusan:

- Jika sekolah asal MA maka Lulus Tepat Waktu
- Jika sekolah asal SMK dan jenis kelamin perempuan maka lulus tepat waktu
- Jika sekolah asal SMK dan jenis kelamin Laki-Laki maka lulus tidak tepat waktu
- Jika sekolah asal SMA dan status sekolah Negeri maka lulus tepat waktu
- Jika sekolah asal SMA dan status sekolah swasta maka lulus tidak tepat waktu

Berdasarkan role/pohon keputusan yang sudah dibuat maka data mahasiswa ke 15 dengan nilai atribut membentuk kelas Tepat Waktu

Tabel 3.17 Hasil Validasi data Testing

No	NPM	Usia Masuk	JK	Kota Asal	Sekolah Asal	Status Sekolah	Jalur Masuk	Pendidikan Orang Tua	Pekerjaan Orang Tua	Penghasilan Orang Tua	IPS-1	IPS-2	SKS-1	SKS-2	Status Kelulusan
15	2103010002	19 Tahun	P	Metro	SMK	Negeri	SPAN-PTKIN	≤SMA	Petani/Peternak	<1.000.000	4	3,58	24	24	TW

3.2.5.2 Implementasi Algoritma Random Forest

Mengklasifikasikan data masa studi mahasiswa menggunakan Random Forest Classifier diperoleh dengan cara membuat kombinasi dari beberapa tree yang kemudian dikombinasikan ke dalam satu model. Adapun langkah-langkah pengerjaan Random Forest dapat dilakukan sebagai berikut:

- a. Mengambil n sampel bootstrap dengan pengembalian dari data training
- b. Menentukan jumlah variabel prediktor yang akan dilakukan pengambilan secara acak dalam proses penentuan pemilah saat pembentukan pohon klasifikasi dengan rumus $\frac{1}{2}\sqrt{p}$
- c. Membentuk pohon klasifikasi dimana pemilihan node terbaik dilakukan berdasarkan variabel-variabel predictor yang diambil secara acak
- d. Mengulangi tahap b sampai tahap d hingga K kali replikasi

e. Melakukan voting mayoritas (*majority vote*) hasil prediksi klasifikasi dari K kali replikasi pembentukan pohon klasifikasi

Contoh perhitungan:

Tabel 3.18 Data Training Asli

No	NPM	Usia Masuk	JK	Kota Asal	Sekolah Asal	Status Sekolah	Jalur Masuk	Pendidikan Orang Tua	Pekerjaan Orang Tua	Penghasilan Orang Tua	IPS-1	IPS-2	SKS-1	SKS-2	Status lulus
1	1804101075	17 Tahun	L	Luar Metro	SMK	Negeri	SPAN-PTKIN	≤SMA	Lainnya	<1.000.000	3,58	3,33	24	24	ttw
2	1804101013	18 Tahun	L	Luar Metro	SMK	Swasta	SPAN-PTKIN	≤SMA	Wiraswasta	<1.000.000	3,17	3,36	24	22	ttw
3	1804102045	19 Tahun	L	Metro	MA	Negeri	SPAN-PTKIN	Tidak Sekolah	Buruh	<1.000.000	3,33	2,67	24	24	tw
4	1804100045	19 Tahun	L	Luar Metro	SMA	Negeri	UM-PTKIN	≤SMA	Buruh	<1.000.000	3,17	3,32	24	22	tw
5	1804100060	18 Tahun	P	Luar Metro	SMA	Negeri	SELEKSI MANDIRI	≤SMA	Petani/Peternak	<1.000.00	3,54	3,13	24	24	tw
6	1804101049	19 Tahun	P	Luar Metro	SMA	Swasta	SELEKSI MANDIRI	≤SMA	Petani/Peternak	<1.000.000	3,5	3,54	24	24	ttw
7	1804022013	18 Tahun	P	Luar Metro	MA	Swasta	SELEKSI MANDIRI	≤SMA	Petani/Peternak	<1.000.00	3,21	3,41	24	22	tw
8	1804100059	18 Tahun	L	Metro	SMK	Negeri	UM-PTKIN	≤SMA	Buruh	<1.000.000	3,33	3,04	24	24	ttw
9	1804100052	18 Tahun	P	Metro	SMK	Negeri	SELEKSI MANDIRI	≤SMA	Buruh	<1.000.000	3,42	3,54	24	24	tw
10	1804101004	19 Tahun	P	Luar Metro	SMA	Negeri	UM-PTKIN	≤SMA	Petani/Peternak	<1.000.000	3,54	3,63	24	24	tw
11	1802040200	17 Tahun	P	Metro	SMK	Swasta	UM-PTKIN	≤SMA	Buruh	<1.000.000	3,54	3,67	24	24	tw
12	1804011013	17 Tahun	L	Metro	MA	Swasta	UM-PTKIN	≤SMA	Wiraswasta	<1.000.000	3,5	3,54	24	24	tw
13	1804020017	18 Tahun	P	Luar Metro	MA	Negeri	SPAN-PTKIN	≤SMA	Petani/Peternak	<1.000.000	3,21	3,41	24	24	tw
14	1804101008	17 Tahun	L	Metro	SMA	Swasta	UM-PTKIN	≤SMA	Petani/Peternak	<1.000.000	3,33	3,04	24	24	ttw

Berdasarkan dataset di atas, akan dibuat pohon keputusan sebanyak 3 pohon (n=14) dengan 13 atribut prediktor dan 1 atribut

kelas/label. Jika terdapat mahasiswa ke 15 dengan nilai atribut, Maka masuk kelas yang mana?

Tabel 3.20 Data Testing

No	NPM	Usia Masuk	JK	Kota Asal	Sekolah Asal	Status Sekolah	Jalur Masuk	Pendidikan Orang Tua	Pekerjaan Orang Tua	Penghasilan Orang Tua	IPS-1	IPS-2	SKS-1	SKS-2	Status Kelulusan
15	2103010002	19 Tahun	P	Metro	SMK	Negeri	SPAN-PTKIN	≤SMA	Petani/Peternak	<1.000.000	4	3,58	24	24	?

Tahap proses perhitungan untuk memprediksi tingkat kelulusan mahasiswa menggunakan Rnadam Forest adalah sebagai berikut :

- a. Melakukan Bootstrap Sampling untuk training data sebanyak n yaitu 14 record data yang akan digunakan untuk membuat pohon keputusan

Tabel 3.20 Data Training Bootstrap Sampling

Sampel Bootstrap	NPM	Usia Masuk	JK	Kota Asal	Sekolah Asal	Status sekolah asal	Jalur masuk	Pendidikan Ayah	Pekerjaan Ayah	Penghasilan Ayah	IPS-1	IPS-2	sks-1	sks-2	Status
12	1804011013	17 Tahun	L	Metro	MA	Swasta	UM-PTKIN	≤SMA	Wiraswasta	<1.000.000	3,5	3,54	24	24	tw
3	1804102045	19 Tahun	L	Metro	MA	Negeri	SPAN-PTKIN	Tidak Sekolah	Buruh	<1.000.000	3,33	2,67	24	24	tw
9	1804100052	18 Tahun	P	Metro	SMK	Negeri	SELEKSI MANDIRI	≤SMA	Buruh	<1.000.000	3,42	3,54	24	24	tw
13	1804020017	18 Tahun	P	Luar Metro	MA	Negeri	SPAN-PTKIN	≤SMA	Petani/Peternak	<1.000.000	3,21	3,41	24	24	tw
2	1804101013	18 Tahun	L	Luar Metro	SMK	Swasta	SPAN-PTKIN	≤SMA	Wiraswasta	<1.000.000	3,17	3,36	24	22	ttw
8	1804100059	18 Tahun	L	Metro	SMK	Negeri	UM-PTKIN	≤SMA	Buruh	<1.000.000	3,33	3,04	24	24	ttw
3	1804102045	19 Tahun	L	Metro	MA	Negeri	SPAN-PTKIN	Tidak Sekolah	Buruh	<1.000.000	3,33	2,67	24	24	tw
9	1804100052	18 Tahun	P	Metro	SMK	Negeri	SELEKSI MANDIRI	≤SMA	Buruh	<1.000.000	3,42	3,54	24	24	tw
5	1804100060	18 Tahun	P	Luar Metro	SMA	Negeri	SELEKSI MANDIRI	≤SMA	Petani/Peternak	<1.000.00	3,54	3,13	24	24	tw
12	1804011013	17 Tahun	L	Metro	MA	Swasta	UM-PTKIN	≤SMA	Wiraswasta	<1.000.000	3,5	3,54	24	24	tw
14	1804101008	17 Tahun	L	Metro	SMA	Swasta	UM-PTKIN	≤SMA	Petani/Peternak	<1.000.000	3,33	3,04	24	24	ttw
6	1804101049	19 Tahun	P	Luar Metro	SMA	Swasta	SELEKSI MANDIRI	≤SMA	Petani/Peternak	<1.000.000	3,5	3,54	24	24	ttw
7	1804022013	18 Tahun	P	Luar Metro	MA	Swasta	SELEKSI MANDIRI	≤SMA	Petani/Peternak	<1.000.00	3,21	3,41	24	22	tw
2	1804101013	18 Tahun	L	Luar Metro	SMK	Swasta	SPAN-PTKIN	≤SMA	Wiraswasta	<1.000.000	3,17	3,36	24	22	ttw

- b. Selanjutnya Menentukan jumlah variabel prediktor yang akan dilakukan pengambilan secara acak dalam proses penentuan

pemilah saat pembentukan pohon klasifikasi dengan rumus $\frac{1}{2}\sqrt{p}$, variable predictor dalam contoh sebanyak 13 sehingga nilai m

$$= \frac{1}{2}\sqrt{13} = 2$$

- c. Membentuk pohon klasifikasi dimana pemilihan node terbaik dilakukan berdasarkan variabel-variabel predictor yang diambil

secara acak dengan m = 2

Misalkan variable predictor yang dipilih untuk pembuatan node 1 adalah Jenis Kelamin dan Jalur Masuk. Maka Langkah selanjutnya dalah menghitung nilai dari Gini dan $Gini_{split}$ dengan rumus sebagai berikut :

- Menghitung Gini Index

$$Gini(S_i) = 1 - \sum_{i=0}^{c-1} P_i^2$$

- Menghitung Gini Index Split

$$Gini_{split} = \sum_{i=0}^{k-1} \left(\frac{n_i}{n}\right) Gini(S_i)$$

Maka dari contoh kasus di atas, dapat dihitung nilai Gini Index sebagai berikut:

Tabel 3.21 Perhitungan nilai Gini Index

Atribut	Nilai Atribut	Jumlah	tw	ttw
Total		14	9	5
JK	L	8	4	4
	p	6	5	1
Jalur Masuk	SPAN-PTKIN	5	3	2
	UM-PTKIN	4	2	2
	Mandiri-PTKIN	5	4	1

Gini Index Jenis Kelamin Laki-laki

$$Gini(JK:L) = 1 - \left(\left(\frac{4}{8}\right)^2 + \left(\frac{4}{8}\right)^2\right) = 0,5$$

Gini Index Jenis Kelamin Laki-laki

$$Gini(JK:P) = 1 - \left(\left(\frac{5}{6}\right)^2 + \left(\frac{1}{6}\right)^2\right) = 0,27778$$

Untuk menghitung Gini Split Atribut Jenis Kelamin

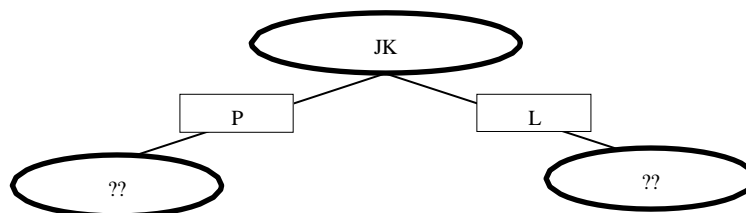
$$Gini_{split} = \left(\frac{8}{14}\right) * 0,5 + \left(\frac{6}{14}\right) * 0,27778 = 0,40476$$

Berdasarkan contoh perhitungan nilai Gini dan Gini Split pada atribut Jenis Kelamin di atas, maka dapat diketahui hasil perhitungan Gini dan Gini Split pada atribut lain dengan perhitungan yang sama.

Tabel 3.22 Node 1 (Root)

Node	Atribut		Jumlah (S)	tw	ttw	Gini	Gini Split
1	Total		14	9	5		
	JK						0,40476
		P	6	5	1	0,27778	
		L	8	4	4	0,5	
	Jalur Masuk						0,42857
		SPAN	5	3	2	0,48	
		UM	4	2	2	0,5	
		Mandiri	5	4	1	0,32	

Dari perhitungan gain antara atribut Jenis kelamin dan Jalur masuk diperoleh nilai Gini split paling kecil pada atribut Jenis kelamin sehingga Jenis kelamin dijadikan Node 1. Berikut pohon keputusan hasil perhitungan Node 1 untuk atribut JK dapat digambarkan seperti pada gambar 3.5 dibawah ini:



Gambar 3.5 Node 1

Atribut Jenis Kelamin memiliki 2 nilai atribut yaitu P dan L yang belum mengklasifikasikan kasus. Oleh karena itu nilai atribut P dan L masih perlu dilakukan perhitungan nilai Gini dan Gini split untuk masing- masing nilai atribut dengan cara perhitungan yang sama seperti pada contoh di atas sampai menghasilkan keputusan.

➤ Node 1.1 (JK = Perempuan)

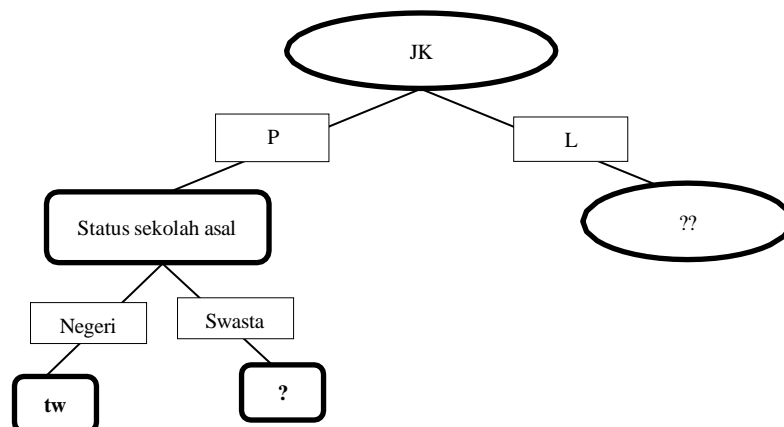
Misalkan untuk membentuk Node 1.1 dipilih variable Jalur Masuk dan Status Sekolah Asal. Maka dihitung kembali nilai Gini dan Gini Split dengan cara

perhitungan yang sama seperti pada perhitungan sebelumnya. Berikut hasil perhitungan nilai Gini dan Gini Split pada atribut jalur masuk dan status sekolah:

Tabel 3.23 Perhitungan Node 1.1

Node	Atribut	Jumlah (S)	tw	ttw	Gini	Gini Split
1.1	Total	6	5	1		
	Jalur Masuk					0,26667
	SPAN	1	1	0	0	
	Mandiri	5	4	1	0,32	
	Status Sekolah Asal					0,16667
	Negeri	4	4	0	0	
	Swasta	2	1	1	0,5	

Dari perhitungan gain antara atribut Jalur Masuk dan Status Sekolah Asal diperoleh nilai Gini split paling kecil pada atribut Status Sekolah Asal sehingga Status Sekolah Asal dijadikan Node 1.1. Karena nilai atribut sekolah asal Negeri telah mengklasifikasi menjadi satu bagian yaitu lulus tepat waktu maka tidak perlu melakukan perhitungan lebih lanjut. Berikut pohon keputusan hasil perhitungan Node 1.1 untuk atribut Status sekolah asal dapat digambarkan seperti pada gambar 3.6:



Gambar 3.6 Node 1.1

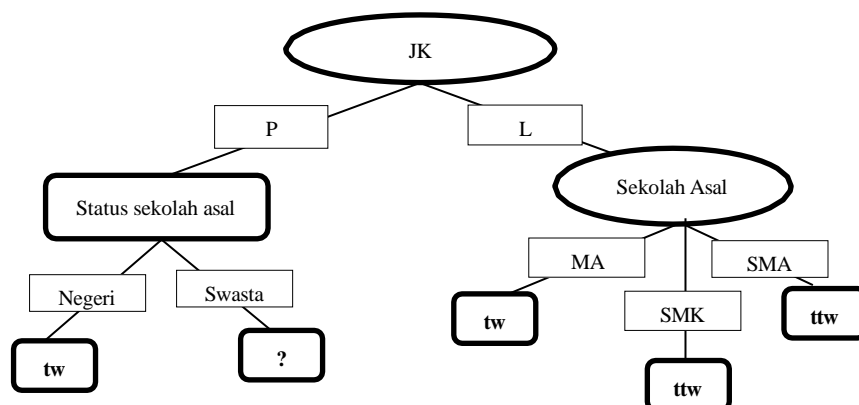
➤ Node 1.2 (JK = Laki-Laki)

Misalkan untuk membentuk Node 1.2 dipilih variable sekolah asal dan Status Sekolah Asal. Maka dihitung kembali nilai Gini dan Gini Split dengan cara perhitungan yang sama seperti pada perhitungan sebelumnya. Berikut hasil perhitungan nilai entropy dan nilai Gain pada atribut sekolah asal dan Status Sekolah Asal:

Tabel 3.24 Perhitungan Node 1.2

Node	Atribut	Jumlah (S)	tw	ttw	Gini	Gini Split
1.2	Total	8	4	4		
	Sekolah Asal					0
	MA	4	4	0	0	
	SMK	3	0	3	0	
	SMA	1	0	1	0	
	Status Sekolah Asal					0,34166
	Negeri	3	2	1	0,1111	
	Swasta	5	2	3	0,48	

Dari perhitungan gini split antara atribut sekolah asal dan Status Sekolah Asal diperoleh nilai Gini split paling kecil pada atribut Sekolah Asal sehingga Sekolah Asal dijadikan Node 1.2. Karena semua nilai atribut sekolah asal telah mengklasifikasi menjadi satu bagian kelas maka tidak perlu melakukan perhitungan lebih lanjut. Berikut pohon keputusan hasil perhitungan Node 1.2 untuk atribut Status sekolah asal dapat digambarkan seperti pada gambar 3.7:



Gambar 3.7 Node 1.2

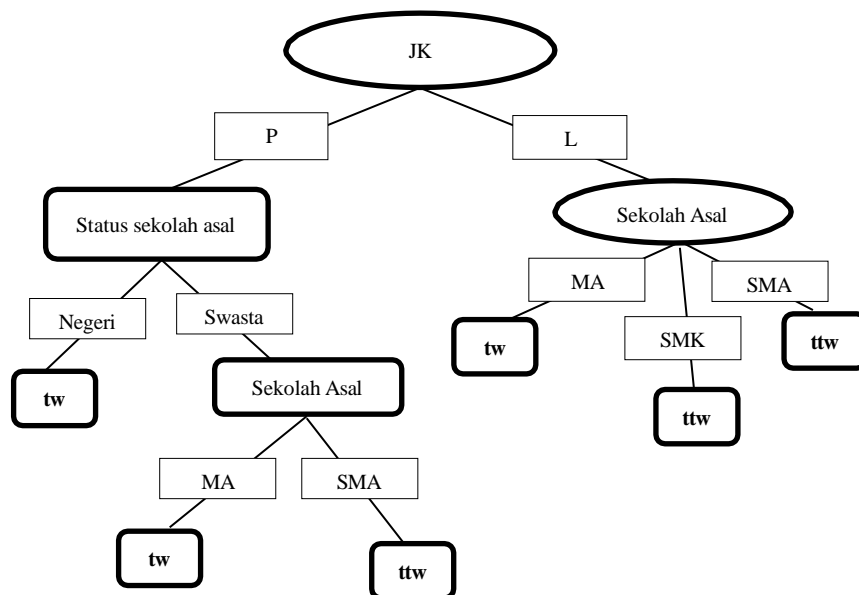
➤ Node 1.1.2 (Status Sekolah Asal = Swasta)

Misalkan untuk membentuk Node 1.2 dipilih variable Sekolah Asal dan Jalur Masuk untuk membentuk Node 1.1.2. Maka dihitung kembali nilai gini dan gini split dengan cara perhitungan yang sama seperti pada perhitungan sebelumnya. Berikut hasil perhitungan nilai gini dan gini split pada atribut Sekolah Asal dan Jalur Masuk:

Tabel 3.25 Perhitungan Node 1.1.2

No	NPM	Jenis Kelamin	Sekolah Asal	Status Sekolah	Jalur Masuk	Status Kelulusan
6	1804101049	P	SMA	Swasta	SELEKSI MANDIRI	ttw
7	1804022013	P	MA	Swasta	SELEKSI MANDIRI	tw

Berikut pohon keputusan hasil perhitungan Node 1.1.2 untuk atribut Status sekolah asal dapat digambarkan seperti pada gambar 3.8:



Gambar 3.8 Pohon Keputusan pertama

d. Mengulangi tahap b sampai tahap d hingga K kali replikasi

Dengan melakukan Langkah yang sama dalam membuat pohon keputusan pertama, maka pohon keputusan kedua dapat dibentuk.

Misalkan variabel prediktor yang dipilih untuk pembuatan node 1 pada pohon kedua adalah Usia Masuk dan Status Sekolah. Maka Langkah selanjutnya adalah menghitung nilai dari gini da gini split dengan rumus sebagai berikut :

- Menghitung Gini Index

$$Gini(S_i) = 1 - \sum_{i=0}^{c-1} P_i^2$$

- Menghitung Gini Index Split

$$Gini_{split} = \sum_{i=0}^{k-1} \left(\frac{n_i}{n}\right) Gini(S_i)$$

Maka dari contoh kasus di atas, dapat dihitung nilai Gini Index sebagai berikut:

Tabel 3.26 Nilai gini index pohon keputusan kedua

Atribut	Nilai Atribut	Jumlah	TW	TTW	Gini	Gini Split
Usia Masuk	17 Tahun	3	2	1	0,44444	0,45833
	18 Tahun	8	5	3	0,46875	
	19 Tahun	3	2	1	0,44444	
	20 Tahun	0	0	0		
Status sekolah	Negeri	7	6	1	0,2449	0,36735
	Swasta	7	3	4	0,4898	

Gini Index Usia Masuk

$$Gini(17\ Tahun) = 1 - \left(\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right) = 0,44444$$

$$Gini(18\ Tahun) = 1 - \left(\left(\frac{5}{8}\right)^2 + \left(\frac{3}{8}\right)^2\right) = 0,46875$$

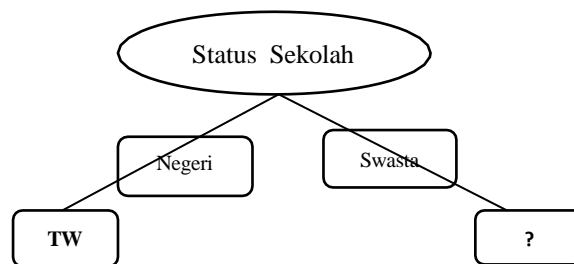
$$Gini(19\ Tahun) = 1 - \left(\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right) = 0,44444$$

Untuk menghitung Gini Split Atribut Jenis Kelamin

$$Gini_{split} = \left(\frac{3}{14}\right) * 0,44444 + \left(\frac{8}{14}\right) * 0,46875 + \left(\frac{3}{14}\right) * 0,44444 = 0,45833$$

Berdasarkan contoh perhitungan gini dan gini split pada atribut Jenis Kelamin di atas, maka dapat diketahui hasil perhitungan Gini dan Gini Split pada status sekolah asal dengan perhitungan yang sama.

Dari perhitungan gini split antara atribut Usia Masuk dan Status Sekolah Asal diperoleh nilai gini split paling kecil pada atribut Sekolah Asal sehingga Sekolah Asal dijadikan Node 1. Berikut pohon keputusan hasil perhitungan Node 1 untuk atribut Status Sekolah dapat digambarkan seperti pada gambar 3.9:



Gambar 3.9 Node 1 pohon keputusan kedua

Atribut Status Sekolah memiliki 2 nilai atribut yaitu Negeri dan Swasta. Hasil perhitungan nilai gini pada atribut status sekolah menunjukkan nilai atribut status sekolah Negeri Sudah mengklasifikasi kelas tepat waktu sedangkan nilai status sekolah swasta belum mengklasifikasi kelasnya sehingga masih perlu dilakukan perhitungan nilai Gini dan Gini split lagi untuk masing- masing nilai atribut dengan cara perhitungan yang sama seperti pada contoh di atas sampai menghasilkan keputusan.

➤ Node 1.2 (Status Sekolah = Swasta)

Misalkan variable predictor yang dipilih untuk pembuatan node 1.2 adalah Sekolah Asal dan IPS-1. Maka Langkah selanjutnya dalah menghitung nilai dari Gini dan Gini Split dengan rumus sebagai berikut :

- Menghitung Gini Index

$$Gini(S_i) = 1 - \sum_{i=0}^{c-1} P_i^2$$

- Menghitung Gini Index Split

$$Gini_{split} = \sum_{i=0}^{k-1} \left(\frac{n_i}{n}\right) Gini(S_i)$$

Maka dari contoh kasus di atas, dapat dihitung nilai Gini Index sebagai berikut:

Tabel 2.27 Nilai gini index Node 1.2 pohon keputusan kedua

Atribut		Jumlah	TW	TTW	Gini	Gini Split
Sekolah Asal						0
	MA	3	3	0		
	SMA	2	0	2		
	SMK	2	0	2		
IPS-1						0,40476
	>3,34	3	2	1	0,44444	
	<=3,34	4	1	3	0,375	

Gini Index Sekolah Asal

$$Gini(MA) = 1 - \left(\left(\frac{3}{3}\right)^2 + \left(\frac{0}{3}\right)^2\right) = 0$$

$$Gini(SMA) = 1 - \left(\left(\frac{0}{2}\right)^2 + \left(\frac{2}{2}\right)^2\right) = 0$$

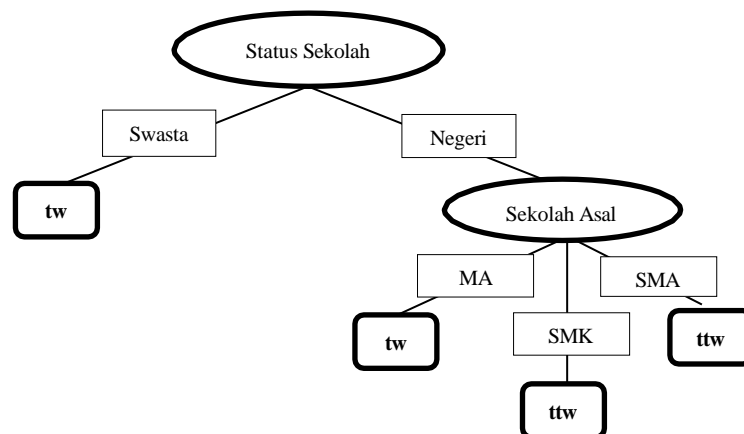
$$Gini(SMK) = 1 - \left(\left(\frac{0}{2}\right)^2 + \left(\frac{2}{2}\right)^2\right) = 0$$

Untuk menghitung Gini Split Atribut Jenis Kelamin

$$Gini_{split} = \left(\frac{3}{7}\right) * 0 + \left(\frac{2}{7}\right) * 0 + \left(\frac{2}{7}\right) * 0 = 0$$

Berdasarkan contoh perhitungan nilai Ginidan Gini Split pada atribut Sekolah Asal di atas, maka dapat diketahui hasil perhitungan Gini dan Gini Split pada atribut IPS-1 dengan perhitungan yang sama. Dari perhitungan gini split antara atribut Sekolah Asal dan IPS-1 diperoleh nilai gin split paling kecil pada atribut Sekolah Asal sehingga Sekolah Asal dijadikan Node 1.2.

Karena semua nilai atribut sekolah asal telah mengklasifikasi menjadi satu bagian kelas maka tidak perlu melakukan perhitungan lebih lanjut. Berikut pohon keputusan hasil perhitungan Node 1.2 untuk atribut Status sekolah asal dapat digambarkan seperti pada gambar 3.10:



Gambar 3.10 pohon keputusan kedua

Dengan melakukan Langkah yang sama dalam membuat pohon keputusan pertama dan kedua, maka pohon keputusan ketiga dapat dibentuk.

Misalkan variable predictor yang dipilih untuk pembuatan node 1 pada pohon ketiga adalah Sekolah Asal dan Kota Asal. Maka Langkah selanjutnya adalah menghitung nilai dari Gini dan Gini Split dengan rumus sebagai berikut :

- Menghitung Gini Index

$$Gini(S_i) = 1 - \sum_{i=0}^{c-1} P_i^2$$

- Menghitung Gini Index Split

$$Gini_{split} = \sum_{i=0}^{k-1} \left(\frac{n_i}{n} \right) Gini(S_i)$$

Maka dari contoh kasus di atas, dapat dihitung nilai Gini Index sebagai berikut:

Tabel 3.28 Perhitungan Gini dan Gini Split node 1 pohon keputusan ketiga

Atribut		Jumlah	TW	TTW	Gini	Gini Split
Kota Asal	Metro	8	6	2	0,375	0,42857
	Luar Metro	6	3	3	0,5	
Sekolah Asal	MA	6	6	0	0	0,26667
	SMA	3	1	2	0,44444	
	SMK	5	2	3	0,48	

Gini Index Kota Asal

$$Gini(Metro) = 1 - \left(\left(\frac{6}{8} \right)^2 + \left(\frac{2}{8} \right)^2 \right) = 0,375$$

$$Gini(Luar Metro) = 1 - \left(\left(\frac{3}{6} \right)^2 + \left(\frac{3}{6} \right)^2 \right) = 0,5$$

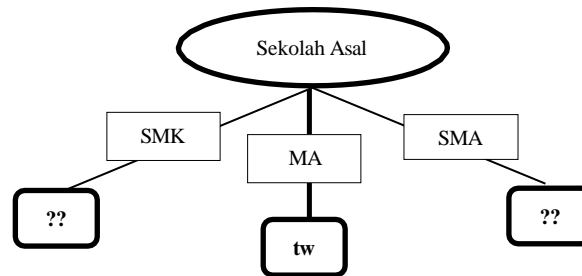
Untuk menghitung Gini Split Atribut Jenis Kelamin

$$Gini_{split} = \left(\frac{8}{14} \right) * 0,375 + \left(\frac{6}{14} \right) * 0,5 = 0,42857$$

Berdasarkan contoh perhitungan nilai Gini Split pada atribut Kota Asal di atas, maka dapat diketahui hasil perhitungan Gini dan Gini Split pada atribut Sekolah Asal dengan perhitungan yang sama.

Dari perhitungan gain antara atribut Kota Asal dan Sekolah Asal diperoleh nilai GiniSplit paling kecil pada atribut Sekolah Asal sehingga Sekolah Asal

dijadikan Node 1. Berikut pohon keputusan hasil perhitungan Node 1 untuk atribut Sekolah Asal dapat digambarkan seperti pada gambar 3.11:



Gambar 3.11 Node 1 pohon keputusan ketiga

Atribut Sekolah Asal memiliki 3 nilai atribut yaitu MA, SMK dan SMA. Hasil perhitungan nilai gini pada atribut Sekolah Asal menunjukkan nilai atribut MA sudah mengklasifikasi kelas tepat waktu sedangkan nilai sekolah Asal SMK dan SMA belum mengklasifikasi kelasnya sehingga masih perlu dilakukan perhitungan nilai Gini dan Gini Split lagi untuk masing- masing nilai atribut dengan cara perhitungan yang sama seperti pada contoh di atas sampai menghasilkan keputusan.

➤ Node 1.1 (Sekolah Asal = SMK)

Misalkan variable predictor yang dipilih untuk pembuatan node 1.1 adalah Jenis kelamin dan Jalur Masuk. Maka Langkah selanjutnya dalah menghitung nilai dari Gini dan Gini Split dengan rumus sebagai berikut :

- Menghitung Gini Index

$$Gini(S_i) = 1 - \sum_{i=0}^{c-1} P_i^2$$

- Menghitung Gini Index Split

$$Gini_{split} = \sum_{i=0}^{k-1} \left(\frac{n_i}{n}\right) Gini(S_i)$$

Maka dari contoh kasus di atas, dapat dihitung nilai Gini Index sebagai berikut:

Tabel 3.29 Perhitungan Gini dan Gini Split Node 1.1 pohon keputusan ketiga

Atribut		Jumlah	TW	TTW	Gini	Gini Split
Jenis Kelamin						0
	L	3	0	3	0	
	P	2	2	0	0	
Jalur Masuk						0
	Seleksi Mandiri	2	2	0	0	
	SPAN-PTKIN	2	0	2	0	
	UM-PTKIN	1	0	1	0	

Gini Index Jenis Kelamin

$$Gini(L) = 1 - \left(\left(\frac{0}{3} \right)^2 + \left(\frac{3}{3} \right)^2 \right) = 0$$

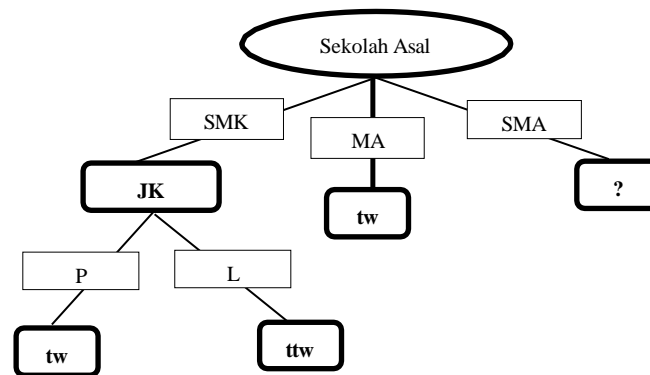
$$Gini(P) = 1 - \left(\left(\frac{2}{2} \right)^2 + \left(\frac{0}{2} \right)^2 \right) = 0$$

Untuk menghitung Gini Split Atribut Jenis Kelamin

$$Gini_{Split} = \left(\frac{3}{5} \right) * 0 + \left(\frac{2}{5} \right) * 0 = 0$$

Berdasarkan contoh perhitungan nilai Gini dan Gini Split pada atribut Jenis kelamin di atas, maka dapat diketahui hasil perhitungan Gini dan Gini Split pada atribut Jalur Masuk dengan perhitungan yang sama.

Dari perhitungan gain antara atribut Jenis Kelamin dan Jalur Masuk diperoleh nilai Gini Split yang sama kecil, maka dipilih salah satu untuk dipilih sebagai node 1.1 misalkan dipilih atribut Jenis Kelamin yang dijadikan Node 1.1. Berikut pohon keputusan hasil perhitungan Node 1.1 untuk atribut Sekolah Asal SMK dapat digambarkan seperti pada gambar 3.12:



Gambar 3.12 Node 1.1 pohon keputusan ketiga

Atribut jenis kelamin memiliki 2 nilai atribut yaitu Perempuan dan Laki-Laki. Hasil perhitungan nilai Gini pada atribut jenis kelamin menunjukkan nilai atribut perempuan sudah mengklasifikasi kelas tepat waktu sedangkan nilai jenis kelamin Laki-Laki mengklasifikasi kelas tidak tepat waktu sehingga proses perhitungan nilai Gini dan Gini Split tidak dilanjutkan.

➤ Node 1.2 (Sekolah Asal = SMA)

Misalkan variable predictor yang dipilih untuk pembuatan node 1.2 adalah Status Sekolah dan IPS-2. Maka Langkah selanjutnya adalah menghitung nilai dari Gini dan Gini Split dengan rumus sebagai berikut :

- Menghitung Gini Index

$$Gini(S_i) = 1 - \sum_{i=0}^{c-1} P_i^2$$

- Menghitung Gini Index Split

$$Gini_{split} = \sum_{i=0}^{k-1} \left(\frac{n_i}{n}\right) Gini(S_i)$$

Maka dari contoh kasus di atas, dapat dihitung nilai Gini Index sebagai berikut:

Tabel 3.30 Perhitungan Gini dan Gini Split node 1.2 pohon keputusan ketiga

Atribut		Jumlah	TW	TTW	Gini	Gini Split
Status sekolah	Negeri	1	1	0	0	0
	swasta	2	0	2	0	
IPS-2	>3,24	1	0	1	0	0,3333
	<=3,24	2	1	1	0,5	

Gini Index Status Sekolah

$$Gini (Negeri) = 1 - \left(\left(\frac{1}{1} \right)^2 + \left(\frac{0}{1} \right)^2 \right) = 0$$

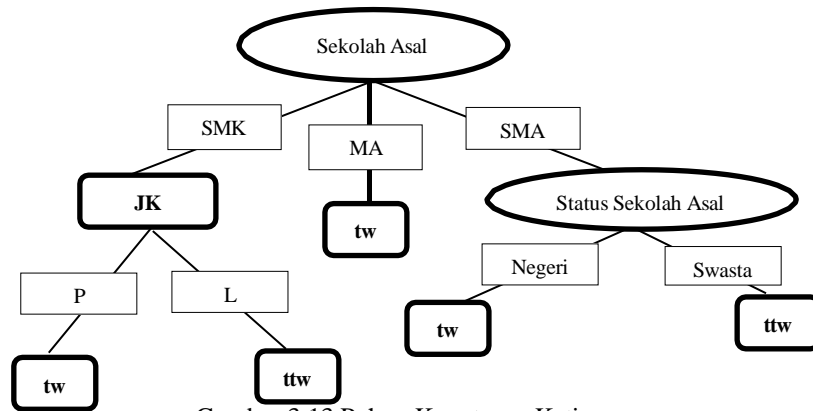
$$Gini (Swasta) = 1 - \left(\left(\frac{0}{2} \right)^2 + \left(\frac{2}{2} \right)^2 \right) = 0$$

Untuk menghitung Gini Split Atribut Jenis Kelamin

$$Gini_{Split} = \left(\frac{1}{3} \right) * 0 + \left(\frac{2}{3} \right) * 0 = 0$$

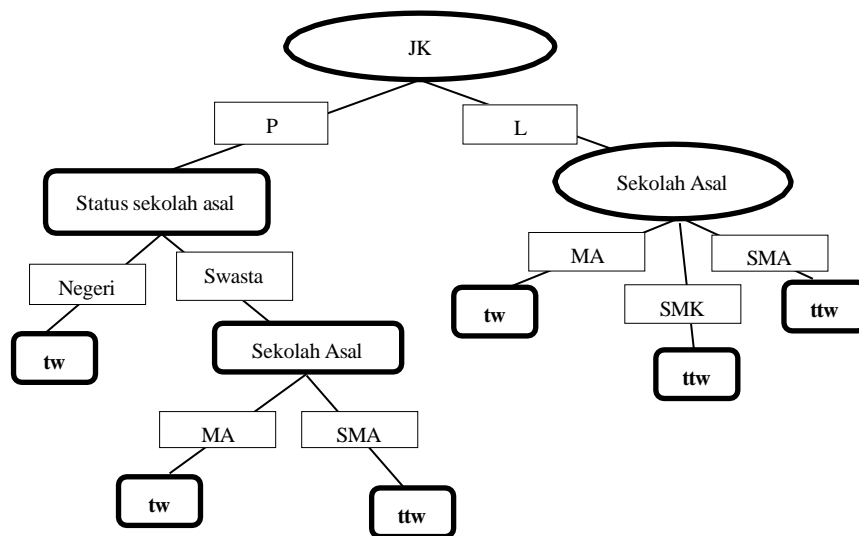
Berdasarkan contoh perhitungan nilai Gini dan Gini Split pada atribut Status Sekolah di atas, maka dapat diketahui hasil perhitungan Gini dan Gini Split pada atribut IPS-2 dengan perhitungan yang sama.

Dari perhitungan Gini Split antara atribut Status Sekolah dan IPS-2 diperoleh nilai Gini Split yang kecil pada atribut Status Sekolah sehingga Status Sekolah dijadikan Node 1.2. Karena semua nilai atribut Status Sekolah asal telah mengklasifikasi menjadi satu bagian kelas maka tidak perlu melakukan perhitungan lebih lanjut. Berikut pohon keputusan ketiga hasil perhitungan Node 1.2 untuk atribut Status sekolah asal dapat digambarkan seperti pada gambar 3.13:

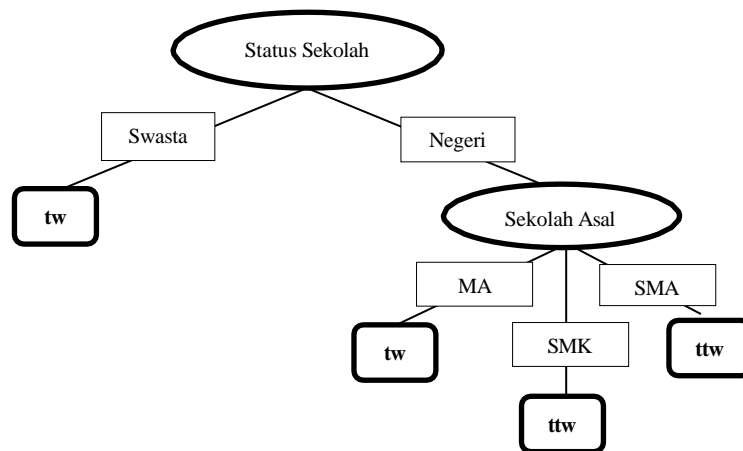


Gambar 3.13 Pohon Keputusan Ketiga

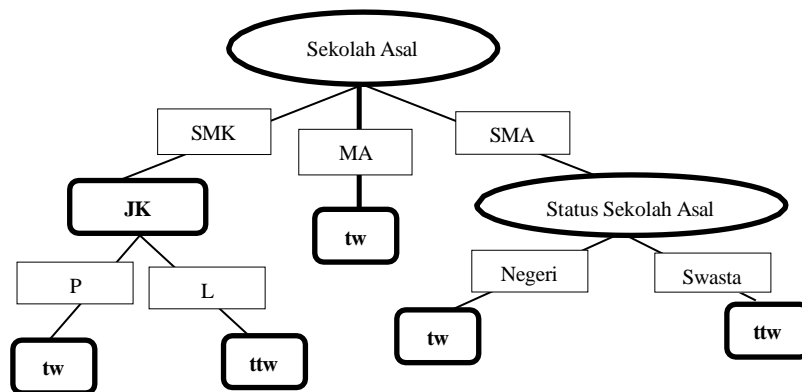
- e. Melakukan voting mayoritas (*majority vote*) hasil prediksi klasifikasi dari K kali replikasi pembentukan pohon klasifikasi
- Berikut adalah 3 pohon keputusan yang selanjutnya akan diklasifikasi dengan random forest



Gambar 3.14 Hasil Pohon keputusan 1



Gambar 3.15 Hasil Pohon keputusan 2



Gambar 3.16 Hasil Pohon keputusan 3

Berdasarkan ketiga pohon yang telah dibentuk Klasifikasikan data baru berdasarkan mayoritas vote dengan kriteria lulus tepat waktu dan lulus tidak tepat waktu. Maka didapatkan data sebagai berikut:

Tabel 3.31 Klasifikasi Random Forest

Status kelulusan	Tree 1	Tree 2	Tree 3	jumlah
tw	3	2	3	8
ttw	3	2	2	7

Dari tiga pohon keputusan di atas, diperoleh hasil 8 tepat waktu dan 7 menghasilkan kelas tidak tepat waktu.

Oleh karena itu, data mahasiswa ke 15 dengan nilai atribut Jenis kelamin = Perempuan, Sekolah Asal = SMK dan Status sekolah = Negeri, diprediksi akan Lulus Tepat Waktu.

Tabel 3.32 Hasil Validasi data Testing Random Forest

No	NPM	Usia Masuk	JK	Kota Asal	Sekolah Asal	Status Sekolah	Jalur Masuk	Pendidikan Orang Tua	Pekerjaan Orang Tua	Penghasilan Orang Tua	IPS-1	IPS-2	SKS-1	SKS-2	Status Kelulusan
15	2103010002	19 Tahun	P	Metro	SMK	Negeri	SPAN-PTKIN	≤SMA	Petani/Peternak	<1.000.000	4	3,58	24	24	TW

3.2.6 Perbandingan Algoritma

Pada tahap ini berisi hasil performance tertinggi dari masing-masing algoritma dalam memprediksi tingkat kelulusan mahasiswa yang terdiri dari Akurasi, precision dan Recall.

3.2.7 Proses Selesai

Proses uji dan analisis sudah selesai maka tahap terakhir adalah membuat kesimpulan dari serangkaian proses pengujian kedua algoritma pada penelitian ini.

3.3 Alat dan bahan

Alat dan bahan yang digunakan dalam penelitian ini antara lain perangkat keras berupa laptop Asus X455L core i5-1135GT GPU NVIDIA MX 350 penyimpanan SSD 512 GB Layar LED FHD 15.6 OS Windows 10. Sementara untuk perangkat lunak yang dipakai adalah *Ms. Excel* digunakan untuk pengolahan data mentah atau data awal dan *Rapidminer Studio* adalah aplikasi yang digunakan untuk pengolahan data.