

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **A. Analisa Hasil**

##### **1. Identifikasi Masalah**

Dalam tahap ini penulis melakukan identifikasi masalah dengan melakukan wawancara dengan bagian kemahasiswaan STMIK Amikom Purwokerto tentang proses penerimaan beasiswa dan beberapa studi pendahuluan dengan mempelajari literatur-literatur yang berkaitan dengan penelitian ini.

Berdasarkan hasil wawancara yang diperoleh, proses penyeleksian yang dilakukan oleh STMIK Amikom Purwokerto menggunakan *Microsoft Excel*, tetapi setelah menggunakan *Microsoft Excel* data tersebut masih harus diseleksi dan dipilih secara manual dikarenakan kriteria yang cukup banyak. Selain itu, data harus diseleksi secara hati-hati serta waktu yang diberi hanya 1 minggu dari dibukanya pendaftaran beasiswa.

Berdasarkan hasil analisis wawancara diatas, dapat disimpulkan bahwa permasalahan yang terdapat pada penerimaan beasiswa di STMIK Amikom Purwokerto yaitu setelah proses penyeleksian yang dilakukan dengan menggunakan *Microsoft Excel* masih harus diseleksi dan pilih kembali secara manual, sementara waktu yang diberikan hanya 1 minggu

serta kriteria yang ditentukan cukup banyak. Jadi dapat disimpulkan bahwa penerimaan beasiswa di STMIK Amikom Purwokerto memerlukan suatu metode atau algoritme untuk menganalisis penerimaan beasiswa agar menghasilkan suatu atribut yang paling berpengaruh serta informasi yang bermanfaat untuk proses penyeleksian penerimaan beasiswa selanjutnya yang diharapkan nantinya analisis tersebut akan menjadi pertimbangan untuk bagian kemahasiswaan STMIK Amikom Purwokerto.

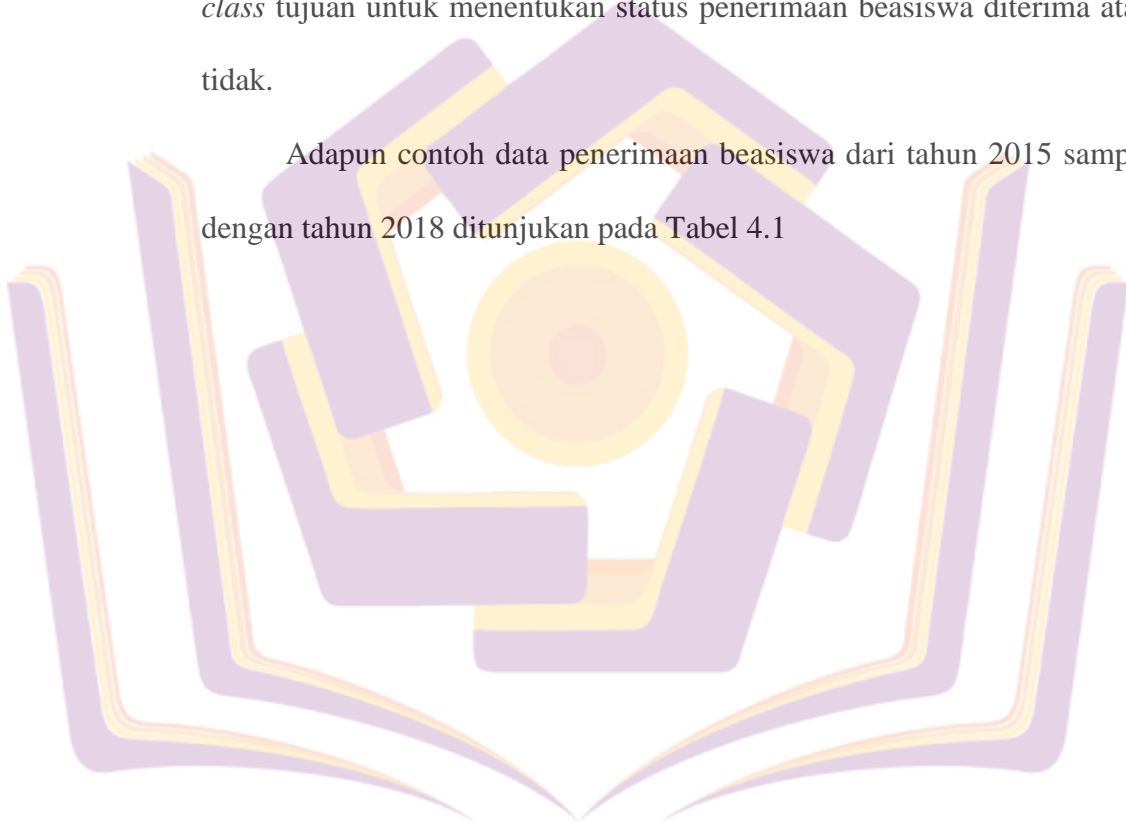
Tahapan selanjutnya dalam identifikasi masalah yaitu mengumpulkan beberapa studi literatur dari penelitian sebelumnya. Dari beberapa hasil penelitian, metode yang dapat digunakan dalam proses penyeleksian data penerimaan beasiswa adalah dengan menggunakan teknik *data mining* menggunakan algoritme *C4.5* sehingga dari beberapa hasil analisis wawancara dan studi literature tersebut permasalahan yang akan diangkat adalah apakah algoritme *C4.5* memiliki nilai performa yang baik agar dapat digunakan dalam proses penerimaan beasiswa di STMIK Amikom Purwokerto.

## 2. Pengumpulan Data

Data yang akan digunakan penulis dalam penelitian ini adalah data penerimaan beasiswa dari tahun 2015 sampai dengan tahun 2018. Data yang digunakan merupakan data primer yang diperoleh secara langsung dari bagian kemahasiswaan STMIK Amikom Purwokerto pada saat

melakukan tahap wawancara. Data yang digunakan sebanyak 279 *record* yang terdiri dari 120 orang penerima beasiswa dengan status diterima dan 159 orang penerima beasiswa dengan status tidak diterima. Serta terdapat 10 atribut yang digunakan untuk penyeleksian calon penerima beasiswa. 10 atribut ini terdiri dari 9 atribut dan 1 atribut yang digunakan sebagai *class* tujuan untuk menentukan status penerimaan beasiswa diterima atau tidak.

Adapun contoh data penerimaan beasiswa dari tahun 2015 sampai dengan tahun 2018 ditunjukkan pada Tabel 4.1



Tabel 4.1 Sampel *dataset* penerima beasiswa

No	JK	SMT	...	...	...	...	...	...	...	...	...	...	...	Tanggal Masuk Persyaratan	Kelengkapan	Status
1	P	4	...	...	...	...	...	...	...	...	...	...	...	4-Mar-2015	?	Ya
2	P	4	...	...	...	...	...	...	...	...	...	...	...	4-Mar-2015	?	Ya
3	L	6	...	...	...	...	...	...	...	...	...	...	...	4-Mar-2015	?	Ya
4	P	6	...	...	...	...	...	...	...	...	...	...	...	4-Mar-2015	?	Ya
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
277	L	4	...	...	...	...	...	...	...	...	...	...	...	18-Mei-2018	Belum Lengkap	Ya
278	L	6	...	...	...	...	...	...	...	...	...	...	...	18-Mei-2018	Lengkap	Ya
279	L	6	...	...	...	...	...	...	...	...	...	...	...	18-Mei-2018	Lengkap	Ya

Gambar 4.1 diatas adalah data primer asli yang belum diolah. Data primer harus diolah terlebih dahulu melalui tahap *preprocessing* data sebelum diproses pada tahap selanjutnya. Pada penelitian ini, penulis memilih atribut yang akan digunakan dengan proses manual dikarenakan beberapa atribut ada yang memiliki maksud yang sama dan ada atribut yang tidak berpengaruh. Dibawah ini adalah atribut-atribut yang sudah dipilih melalui proses manual, berikut penjelasannya:

1. SMT (Semester)

Semester adalah satuan waktu kegiatan yang terdiri dari 16 minggu perkuliahan atau kegiatan terjadwal lainnya yang ditempuh mahasiswa.

2. IPK (Indeks Prestasi Kumulatif)

IPK adalah ukuran kemampuan mahasiswa dalam hal akademik sampai pada periode tertentu berdasarkan pada jumlah sks yang ditempuh dan menjadi salah satu syarat utama untuk penerimaan beasiswa.

Tabel 4.2 Kategori IPK

IPK	Predikat
$2.00 \leq \text{IPK} \leq 3.00$	Memuaskan
$3.00 \leq \text{IPK} \leq 3.50$	Sangat Memuaskan
$3.50 \leq \text{IPK} \leq 4.00$	Pujian

### 3. Usia

Usia yang ditetapkan untuk penerimaan beasiswa maksimal 23 tahun.

### 4. Jumlah Anggota Keluarga

Jumlah Anggota Keluarga adalah jumlah anggota yang harus ditanggung oleh keluarga dari calon penerima beasiswa.

Tabel 4.3 Kategori Jumlah Anggota Keluarga

Jumlah Anggota Keluarga	Keterangan
1-2 Orang	Sedikit
3-4 Orang	Sedang
≥ 5 Orang	Banyak

### 5. Penghasilan

Penghasilan adalah kemampuan ekonomi dari setiap keluarga calon penerima beasiswa.

Tabel 4.4 Kategori Penghasilan

IPK	Keterangan
< 1.500.000	Rendah
≥ 1.500.000	Sedang
≥ 2.500.000	Tinggi
≥ 3.500.000	Sangat Tinggi

### 6. Prestasi

Prestasi adalah prestasi yang telah didapat oleh mahasiswa baik dilingkungan kampus atau diluar kampus.

#### 7. Keaktifan

Keaktifan adalah keaktifan mahasiswa dalam menjalani kegiatan di lingkungan kampus baik kegiatan di dalam kelas maupun dalam berorganisasi.

#### 8. Mendapat Beasiswa

Mendapat beasiswa adalah mahasiswa yang pada semester sebelumnya sudah pernah atau belum pernah mendapat beasiswa.

#### 9. Kelengkapan

Kelengkapan adalah lengkap atau tidaknya mahasiswa saat mengumpulkan berkas persyaratan penerimaan beasiswa.

#### 10. Status

Status merupakan class tujuan untuk menyatakan pendaftar beasiswa tersebut diterima atau tidak. Terdapat dua nilai kemungkinan yaitu nilai iya dan tidak. Jika iya maka pendaftar beasiswa dinyatakan diterima, jika tidak maka pendaftar beasiswa dinyatakan tidak diterima.

### 3. Preprocessing Data

#### a. Persiapan Data

Pada tahapan *preprocessing* data yaitu dengan mengubah data dari bentuk CSV (*Comma-Separated Value*) atau *Microsoft Excel* menjadi *file* dengan ekstensi *.arff* (dot *arff*) agar data dengan mudah dapat dibuka oleh *software weka*. Adapun sampel data penerimaan beasiswa yang sudah mengalami *preprocessing* data dapat dilihat pada table 4.5

Tabel 4.5 Sampel *dataset* setelah tahap *preprocessing*

No.	S M T	IPK	Usia	Jumlah Anggota Keluarga	Penghasi lan	Presta si	Keakti fan	Mendapat Beasiswa	Kelengk apan	Sta tus
1	4	3.82	?	3	1300000	Tidak	?	?	?	Ya
2	4	3.53	?	4	1500000	Tidak	?	?	?	Ya
3	6	3.51	?	4	2000000	Tidak	?	?	?	Ya
4	6	3.66	?	4	1500000	Tidak	?	?	?	Ya
5	6	3.6	?	7	650000	Ya	?	?	?	Ya
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
278	6	3.82	21	?	3000000	Ya	Ya	Belum	Lengkap	Ya
279	6	3.25	21	?	4500000	Ya	Tidak	Belum	Lengkap	Ya

Tabel 4.5 menampilkan sampel *dataset* penerimaan beasiswa dengan jumlah data sebanyak 279 data. Pada table tersebut terdapat 10 atribut. 9 atribut sebagai atribut penilaian dan 1 atribut sebagai *class* tujuan. 9 atribut tersebut diantaranya yaitu atribut semester dari semester 2 sampai semester 8, atribut IPK memiliki nilai *minimum* 3.00 dan *maximum* 4.00, atribut usia memiliki nilai *minimum* (?) atau kosong



dan nilai *maximum* 23 tahun, atribut jumlah anggota keluarga memiliki nilai *minimum* (?) atau kosong dan nilai *maximum* 9 orang, atribut penghasilan memiliki nilai nilai *minimum* (?) atau kosong dan nilai *maximum* Rp. 8.000.000, sedangkan pada atribut prestasi yang merupakan atribut kategorikal memiliki nilai ya serta nilai tidak, atribut keaktifan memiliki memiliki nilai kosong atau (?) dan nilai ya serta nilai tidak. Adapun atribut mendapat beasiswa juga merupakan atribut kategorikal memiliki nilai kosong atau (?) dan nilai pernah serta nilai belum, atribut kelengkapan yang merupakan atribut kategorikal memiliki nilai kosong atau (?) dan nilai lengkap serta nilai belum.

b. Penanganan *Missing values*

*Missing values* adalah keadaan dimana beberapa nilai atribut dalam *dataset* kosong atau tidak ada nilainya. Penangan *missing values* dapat dilakukan dengan mencari nilai mean, median dan modus. Nilai mean dan median digunakan untuk data bernilai numerik sedangkan nilai modus digunakan untuk data bernilai kategorikal, tetapi pada penelitian ini penanganan *missing values* hanya menggunakan perhitungan nilai mean dan modus. Data yang digunakan sebanyak 279 data dimana terdapat data dengan *missing values* pada atribut usia, jumlah anggota, penghasilan, keaktifan, mendapat beasiswa dan kelengkapan. Dibawah ini merupakan penanganan *missing values* pada data dengan menggunakan perhitungan nilai mean dan modus yaitu :

## 1. Mean

Menurut Izzatul (2013) Mean melakukan *imputasi missing* data dengan menghitung nilai rata-rata dari seluruh nilai data yang diketahui. Metode ini hanya digunakan untuk data yang bersifat numerik. Kriteria usia, jumlah anggota keluarga dan penghasilan menggunakan mean sebagai cara untuk *imputasi* data karena kriteria tersebut merupakan data yang bersifat numerik. Berikut merupakan rumus perhitungan nilai mean ;

$$Mean = \frac{\sum \text{nilai yang diketahui}}{279}$$

Tabel 4.6 Sampel *dataset* nilai mean

No	Usia	Jumlah Anggota Keluarga	Penghasilan (Rp)	Status
1	20	3	1300000	Ya
2	20	4	1500000	Ya
...	...	...	...	...
...	...	...	...	...
278	21	4	3000000	Ya
279	21	4	4500000	Ya

Tabel 4.6 menjelaskan penanganan *missing values* menggunakan teknik mean pada data penerimaan beasiswa yang bersifat numerik, atribut tersebut diantaranya yaitu kriteria usia, jumlah anggota keluarga dan penghasilan. Berdasarkan tabel diatas dapat disimpulkan bahwa pada kolom usia diisi dengan nilai rata-rata usia yaitu 20 tahun dan pada kolom jumlah anggota keluarga

diisi dengan nilai rata-rata jumlah anggota keluarga sebanyak 4 anggota serta pada kolom penghasilan diisi dengan nilai rata-rata penghasilan yaitu sebesar Rp.1.961.113.

## 2. Modus

Menurut Izzatul (2013) Modus hanya melakukan *imputasi* data untuk data yang bersifat kategorikal. Modus merupakan nilai yang memiliki frekuensi terbesar atau nilai yang paling banyak muncul. Kriteria keaktifan, mendapat beasiswa dan kelengkapan menggunakan modus sebagai cara untuk *imputasi* data karena kriteria tersebut merupakan data yang bersifat kategorikal. Dibawah ini merupakan sampel *dataset* pada kriteria keaktifan, mendapat beasiswa dan kelengkapan yang sudah dilakukan penanganan *missing values* dengan menggunakan nilai modus, disajikan pada tabel 4.7 sebagai berikut

Tabel 4.7 Sampel *dataset* nilai modus

No	Keaktifan	Mendapat Beasiswa	Kelengkapan	Status
1	Ya	Belum	Lengkap	Ya
2	Ya	Belum	Lengkap	Ya
...	...	...	...	...
...	...	...	...	...
278	Ya	Belum	Lengkap	Ya
279	Tidak	Belum	Lengkap	Ya

c. Proses Diskritisasi

Setelah data melalui tahap penanganan *missing values* maka selanjutnya data akan dikelompokan atau dikategorikan sesuai dengan kategori yang sudah ada yang disebut diskritisasi data. Diskritisasi data adalah proses mentransformasikan nilai-nilai atribut bertipe numerik menjadi nominal, ordinal atau bahkan biner. Dibawah ini merupakan tabel kategori yang disajikan pada tabel 4.8 sebagai berikut :

Tabel 4.8 Kategori IPK, Jumlah anggota keluarga dan Penghasilan

Kategori	Keterangan
IPK	$2.00 \leq \text{IPK} \leq 3.00$ = Memuaskan $3.00 \leq \text{IPK} \leq 3.50$ = Sangat Memuaskan $3.50 \leq \text{IPK} \leq 4.00$ = Pujian
Jumlah Anggota Keluarga	1-2 Orang = Sedikit 3-4 Orang = Sedang $\geq 5$ Orang = Banyak
Penghasilan	$< 1.500.000$ = Rendah $\geq 1.500.000$ = Sedang $\geq 2.500.000$ = Tinggi $\geq 3.500.000$ = Sangat Tinggi

Berikut merupakan tabel dari *dataset* yang sudah melalui proses diskritisasi

Tabel 4.9 Sampel *dataset* hasil diskritisasi

No	IPK	Jumlah Anggota Keluarga	Penghasilan (Rp)	Status
1	Pujian	Sedang	Rendah	Ya
2	Pujian	Sedang	Sedang	Ya
...	...	...	...	...
...	...	...	...	...
278	Pujian	Sedang	Tinggi	Ya
279	Sangat Memuaskan	Sedang	Sangat Tinggi	Ya

d. Konversi data CSV ke dalam. arff

Pada tahap ini data disimpan dalam bentuk CSV (*Comma-Separated Value*) kemudian dikonversikan ke dalam ekstensi. arff, tujuannya agar data tersebut dapat dengan mudah dibuka melalui *software weka*. Dibawah merupakan hasil sampel dari *dataset* yang melalui tahap penanganan *missing values* dengan nilai mean untuk data berkategori numerik dan nilai modus untuk data yang bersifat kategorikal serta sudah melalui proses diskritisasi data, berikut adalah data dalam bentuk CSV yang akan disajikan dalam gambar 4.1

```
4,Pujian,20,Sedang,Rendah,Tidak,Ya,Belum,Lengkap,Ya
4,Pujian,20,Sedang,Sedang,Tidak,Ya,Belum,Lengkap,Ya
6,Pujian,20,Sedang,Sedang,Tidak,Ya,Belum,Lengkap,Ya
6,Pujian,20,Sedang,Sedang,Tidak,Ya,Belum,Lengkap,Ya
6,Pujian,20,Banyak,Rendah,Ya,Ya,Belum,Lengkap,Ya
6,Sangat Memuaskan,20,Sedang,Rendah,Ya,Ya,Belum,Lengkap,Tidak
Dst.....
```

Gambar 4.1 Sampel data dalam bentuk CSV

Gambar 4.1 merupakan gambar sampel data dalam bentuk CSV, agar data tersebut bisa langsung digunakan dalam *software weka*, maka data harus diubah ekstensinya ke dalam bentuk *.arff*. Di bawah ini merupakan data yang disajikan dalam bentuk *.arff* pada gambar 4.2

```

@relation beasiswa

@attribute SMT {2,4,6,8}
@attribute IPK {Memuaskan, Sangat_Memuaskan, Pujian}
@attribute USIA {18,19,20,21,22,23}
@attribute JA {Sedikit, Sedang, Banyak}
@attribute PENGHASILAN {Rendah, Sedang, Tinggi, Sangat_Tinggi}
@attribute PRESTASI {Ya, Tidak}
@attribute KEAKTIFAN {Ya, Tidak}
@attribute MB {Pernah,Belum}
@attribute KELENGKAPAN {Lengkap,Belum}
@attribute STATUS {Ya,Tidak}

@data

4,Pujian,20,Sedang,Rendah,Tidak,Ya,Belum,Lengkap,Ya
4,Pujian,20,Sedang,Sedang,Tidak,Ya,Belum,Lengkap,Ya
6,Pujian,20,Sedang,Sedang,Tidak,Ya,Belum,Lengkap,Ya
6,Pujian,20,Sedang,Sedang,Tidak,Ya,Belum,Lengkap,Ya
6,Pujian,20,Banyak,Rendah,Ya,Ya,Belum,Lengkap,Ya
6,Sangat Memuaskan,20,Sedang,Rendah,Ya,Ya,Belum,Lengkap,Tidak
Dst.....

```

Gambar 4.2 Konversi dari data CSV ke bentuk.arff



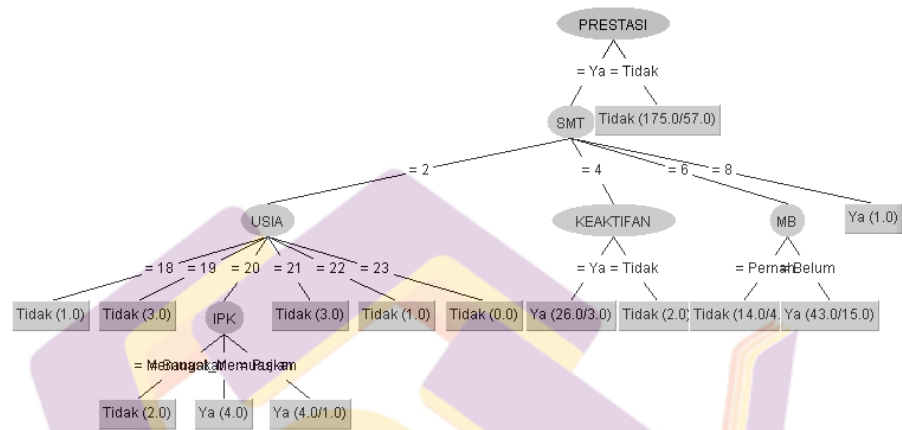
Pada gambar 4.2 merupakan data dengan format .arff. Didalam data tersebut terdapat *@relation* yang berfungsi untuk memberikan penamaan pada file, *@attribute* yang berfungsi untuk menjelaskan atribut yang akan digunakan untuk pendeklarasian atribut pada data tersebut, *@data* digunakan untuk mendeklarasikan data secara keseluruhan sehingga data sebanyak 279 dideklarasikan pada *@data*. *Rule* yang ada pada data pertama yaitu 4 = Semester, Pujian = IPK, 20 = Usia, Sedang = Jumlah anggota keluarga, Rendah = Penghasilan, Tidak = Prestasi, Ya = Keaktifan, Belum = Mendapat beasiswa, Lengkap = Kelengkapan dan yang terakhir yaitu Ya = Status atau *class* tujuan yang digunakan untuk menentukan apakah calon penerima beasiswa tersebut diterima atau tidak.

#### 4. *Decision Tree* (Pohon Keputusan)

Setelah *dataset* sudah melalui tahap *preprocessing* data, tahap selanjutnya yaitu tahapan untuk pengujian terhadap algoritme *C4.5* yang nantinya akan menghasilkan sebuah model berupa pohon keputusan atau *rule*. Pohon keputusan atau *rule* tersebut dihasilkan dengan cara menghitung nilai atribut menggunakan metode *gain ratio*. Atribut dengan nilai *gain ratio* terbesar akan digunakan sebagai *node* akar pertama dan menjadi atribut yang paling berpengaruh.



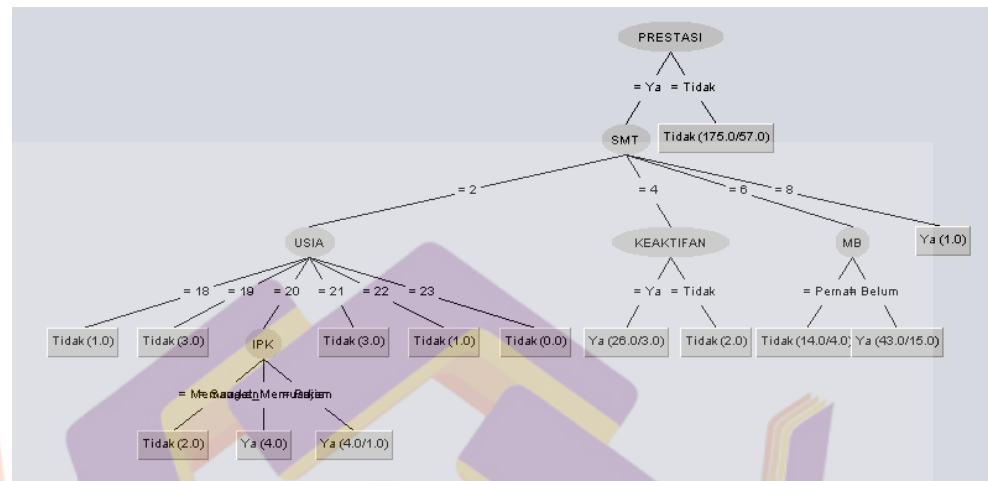
Dibawah ini merupakan pohon keputusan yang dihasilkan oleh *software weka*



Gambar 4.3 Tampilan *Decision Tree* Algoritma J48

Gambar 4.3 merupakan pohon keputusan yang diperoleh dari hasil pengujian dengan menggunakan algoritme *J48* atau *C4.5*. Pengujian dilakukan dengan data sebanyak 279 data serta menggunakan *test option* berupa *cross-validation folds 10*. Atribut yang digunakan pada pengujian ini sebanyak 10 atribut yaitu atribut Semester, IPK, Usia, Jumlah Anggota Keluarga, Penghasilan, Prestasi, Keaktifan, Mendapat Beasiswa., Kelengkapan dan Status. Setelah diuji menggunakan pohon keputusan terdapat atribut yang tidak terbaca pada *weka*, atribut tersebut antara lain jumlah anggota keluarga, penghasilan dan kelengkapan. Oleh karena itu, dilakukan pengujian kembali dengan menggunakan atribut yang hanya terbaca pada *weka*. Berikut merupakan pohon keputusan yang diperoleh

dari hasil penghilangan 3 atribut yaitu penghilangan atribut jumlah anggota keluarga, penghasilan dan kelengkapan.



Gambar 4.4 Tampilan *Decision tree* penghilangan 3 atribut

Gambar 4.4 merupakan tampilan pohon keputusan setelah menghilangkan 3 atribut yang tidak berpengaruh pada tampilan pohon keputusan Gambar 4.3. Pohon keputusan yang dihasilkan pada Gambar 4.3 ataupun 4.4 sama, tidak terdapat perbedaan dikarenakan atribut jumlah anggota keluarga, penghasilan dan kelengkapan adalah atribut yang tidak berpengaruh sehingga tidak terbaca pada pohon keputusan aplikasi *weka*. Hasil akurasi yang diperoleh dengan menggunakan 10 atribut dan 3 atribut yang dihapus berbeda, akurasi dengan menghilangkan 3 atribut yang tidak berpengaruh lebih besar dibanding dengan akurasi menggunakan 10 atribut.

Berikut merupakan perbandingan dari 2 akurasi tersebut

```

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      177          63.4409 %
Incorrectly Classified Instances    102          36.5591 %
Kappa statistic                    0.2141
Mean absolute error                 0.44
Root mean squared error             0.4939
Relative absolute error             89.7316 %
Root relative squared error         99.7642 %
Total Number of Instances          279

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.367   0.164   0.629     0.367   0.463     0.232   0.607    0.537    Ya
                0.836   0.633   0.636     0.836   0.723     0.232   0.607    0.636    Tidak
Weighted Avg.   0.634   0.431   0.633     0.634   0.611     0.232   0.607    0.593

=== Confusion Matrix ===
 a  b  <-- classified as
44 76 |  a = Ya
26 133 | b = Tidak

```

Gambar 4.5 Tampilan akurasi 10 atribut

Gambar 4.5 merupakan tampilan hasil akurasi dengan menggunakan 10 atribut. Hasil akurasi yang diperoleh sebesar 63,44%.

```

Time taken to build model: 0.12 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      188          67.3835 %
Incorrectly Classified Instances    91          32.6165 %
Kappa statistic                    0.3041
Mean absolute error                 0.4248
Root mean squared error             0.4737
Relative absolute error             86.6454 %
Root relative squared error         95.6792 %
Total Number of Instances          279

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.442   0.151   0.688     0.442   0.538     0.322   0.624    0.586    Ya
                0.849   0.558   0.668     0.849   0.748     0.322   0.624    0.644    Tidak
Weighted Avg.   0.674   0.383   0.677     0.674   0.658     0.322   0.624    0.619

=== Confusion Matrix ===
 a  b  <-- classified as
53 67 |  a = Ya
24 135 | b = Tidak

```

Gambar 4.6 Tampilan akurasi penghilangan 3 atribut

Gambar 4.6 merupakan tampilan hasil akurasi dengan menghapus 3 atribut yang tidak berpengaruh. Atribut tersebut antara lain jumlah anggota keluarga, penghasilan dan kelengkapan. Hasil akurasi yang diperoleh sebesar 67,38%. Dilihat dari Gambar 4.5 dan 4.6 hasil akurasi dengan menghapus 3 atribut yang tidak berpengaruh lebih besar dibandingkan dengan tetap menggunakan atribut yang tidak berpengaruh.

Pohon keputusan pada gambar 4.3 dapat dijabarkan secara lebih rinci pada gambar 4.7 dibawah ini :

```

=== Classifier model (full training set) ===

J48 pruned tree
-----
PRESTASI = Ya
|  SMI = 2
|  |  USIA = 18: Tidak (1.0)
|  |  USIA = 19: Tidak (3.0)
|  |  USIA = 20
|  |  |  IPK = Memuaskan: Tidak (2.0)
|  |  |  IPK = Sangat_Memuaskan: Ya (4.0)
|  |  |  IPK = Pujian: Ya (4.0/1.0)
|  |  USIA = 21: Tidak (3.0)
|  |  USIA = 22: Tidak (1.0)
|  |  USIA = 23: Tidak (0.0)
|  SMI = 4
|  |  KEAKTIFAN = Ya: Ya (26.0/3.0)
|  |  KEAKTIFAN = Tidak: Tidak (2.0)
|  SMI = 6
|  |  MB = Pernah: Tidak (14.0/4.0)
|  |  MB = Belum: Ya (43.0/15.0)
|  SMI = 8: Ya (1.0)
PRESTASI = Tidak: Tidak (175.0/57.0)

```

Gambar 4.7 Tampilan rule *IF THEN* Algoritma J48

Gambar 4.7 merupakan *rule* atau aturan *IF THEN* yang dihasilkan dari proses klasifikasi dengan menggunakan algoritma C4.5. *Rule* yang

dihasilkan nantinya dapat digunakan dalam proses penyeleksian penerimaan beasiswa.

Tabel 4.10 Penjabaran rinci hasil *decision tree*

No	Uraian
1.	Jika prestasi = ya dan semester = 2 dan usia = 18 maka tidak diterima
2	Jika prestasi = ya dan semester = 2 dan usia = 19 maka tidak diterima
3	Jika prestasi = ya dan semester = 2 dan usia = 20 dan ipk = memuaskan maka tidak diterima
4	Jika prestasi = ya dan semester = 2 dan usia = 20 dan ipk = sangat memuaskan maka diterima
5	Jika prestasi = ya dan semester = 2 dan usia = 20 dan ipk = pujian maka diterima
6	Jika prestasi = ya dan semester = 2 dan usia = 21 maka tidak diterima
7	Jika prestasi = ya dan semester = 2 dan usia = 22 maka tidak diterima
8	Jika prestasi = ya dan semester = 2 dan usia = 23 maka tidak diterima
9	Jika prestasi = ya dan semester = 4 dan keaktifan = ya maka diterima
10	Jika prestasi = ya dan semester = 4 dan keaktifan = tidak maka tidak diterima
11	Jika prestasi = ya dan semester = 6 dan mendapat beasiswa = pernah maka tidak diterima
12	Jika prestasi = ya dan semester = 6 dan mendapat beasiswa = belum maka diterima
13	Jika prestasi = ya dan semester = 8 maka diterima
14	Jika prestasi = tidak maka tidak diterima

(Sumber : *dataset* yang sudah diuji)

## 5. Evaluasi

Setelah proses *decision tree* selesai, maka tahap selanjutnya yaitu evaluasi. Evaluasi dilakukan untuk menganalisa hasil pengujian yang dilakukan dengan menggunakan algoritme *C4.5*. Tujuan dari evaluasi yaitu untuk mengetahui tingkat akurasi yang dihasilkan algoritme *C4.5* terhadap *dataset* penerimaan beasiswa sebanyak 279 data. Berdasarkan hasil pengujian yang telah dilakukan diperoleh nilai TP atau *true positive*, TN atau *true negative*, FP atau *false positive* dan FN atau *false negative*.

### a) Pengujian algoritme *C4.5* dengan *imputasi missing values*

Dibawah ini adalah *confusion matrix* hasil dari pengujian dengan *imputasi missing values*

Tabel 4.11 Hasil *confusion matrix* dengan *imputasi missing value*

44 ( <i>True Positive</i> )	76 ( <i>False Negative</i> )
26 ( <i>False Positive</i> )	133 ( <i>True Negative</i> )

(Sumber : *dataset* yang sudah diuji)

Pada tabel 4.11 merupakan hasil *confusion matrix* yang diperoleh dari algoritme *C4.5* dengan *imputasi missing values*. Nilai 44 sebagai *true positive*, 76 sebagai *false negative*, 26 sebagai *false positive*, dan 133 sebagai *true negative*. Selanjutnya akan dilakukan perhitungan untuk mencari nilai performa yaitu nilai akurasi yang diperoleh dari algoritme *C4.5* sebagai berikut :

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \times 100 \% \\
 &= \frac{44+133}{44+133+26+76} \times 100 \% \\
 &= \frac{177}{279} \times 100 \% \\
 &= 0,6344 \times 100\% \\
 &= 63,44 \%
 \end{aligned}$$

Berdasarkan hasil perhitungan *confusion matrix* diatas, pengujian menggunakan algoritme *C4.5* dengan *imputasi missing values* memiliki performa sebagai berikut :

Tabel 4.12 Hasil Performa dengan *imputasi missing value*

Performa	Nilai
Accuracy	63,44 %

(Sumber : Hasil Perhitungan)

Tabel 4.12 merupakan hasil perhitungan dari *accuracy* yang dihitung secara manual dan memiliki hasil yang sama dengan output yang dihasilkan pada software *weka*. Hasil *accuracy* yang diperoleh dengan *imputasi missing values* sebesar 63,44%.



b) Pengujian algoritme *C4.5* tanpa *imputasi missing values*

Dibawah ini adalah *confusion matrix* hasil dari pengujian tanpa *imputasi missing values*

Tabel 4.13 Hasil *confusion matrix* tanpa *imputasi missing value*

52 ( <i>True Positive</i> )	68 ( <i>False Negative</i> )
20 ( <i>False Positive</i> )	139 ( <i>True Negative</i> )

(Sumber : *dataset* yang sudah diuji)

Pada tabel 4.13 merupakan hasil *confusion matrix* yang diperoleh dari algoritme *C4.5* tanpa *imputasi missing values*. Nilai 52 sebagai *true positive*, 68 sebagai *false negative*, 20 sebagai *false positive*, dan 139 sebagai *true negative*. Selanjutnya akan dilakukan perhitungan untuk mencari nilai performa yaitu nilai akurasi yang diperoleh dari algoritme *C4.5* sebagai berikut :

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \times 100 \% \\
 &= \frac{52+139}{52+139+20+68} \times 100 \% \\
 &= \frac{191}{279} \times 100 \% \\
 &= 0,6845 \times 100\% \\
 &= 68,45 \%
 \end{aligned}$$

Berdasarkan hasil perhitungan *confusion matrix* diatas, pengujian menggunakan algoritme *C4.5* tanpa *imputasi missing values* memiliki performa sebagai berikut :



Tabel 4.14 Hasil Performa tanpa *imputasi missing value*

Performa	Nilai
<i>Accuracy</i>	68,45 %

(Sumber : Hasil Perhitungan)

Tabel 4.14 merupakan hasil perhitungan dari *accuracy* yang dihitung secara manual dan memiliki hasil yang sama dengan output yang dihasilkan pada software *weka*. Hasil *accuracy* yang diperoleh tanpa *imputasi missing values* sebesar 68,45%.

- c) Pengujian algoritme *C4.5* dengan menghilangkan atribut yang tidak berpengaruh

Dibawah ini adalah *confusion matrix* hasil dari pengujian dengan menghilangkan atribut yang tidak berpengaruh.

Tabel 4.15 *Confusion matrix* menghilangkan atribut tidak berpengaruh

53 ( <i>True Positive</i> )	67 ( <i>False Negative</i> )
24 ( <i>False Positive</i> )	135 ( <i>True Negative</i> )

(Sumber : *dataset* yang sudah diuji)

Pada tabel 4.15 merupakan hasil *confusion matrix* yang diperoleh dari algoritme *C4.5* dengan dengan menghilangkan atribut yang tidak berpengaruh. Nilai 53 sebagai *true positive*, 67 sebagai *false negative*, 24 sebagai *false positive*, dan 135 sebagai *true*

*negative*. Selanjutnya akan dilakukan perhitungan untuk mencari nilai performa yaitu nilai akurasi yang diperoleh dari algoritme *C4.5* sebagai berikut :

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \times 100 \% \\
 &= \frac{53+135}{53+135+24+67} \times 100 \% \\
 &= \frac{188}{279} \times 100 \% \\
 &= 0,6738 \times 100\% \\
 &= 67,38 \%
 \end{aligned}$$

Berdasarkan hasil perhitungan *confusion matrix* diatas, pengujian menggunakan algoritme *C4.5* dengan menghilangkan atribut yang tidak berpengaruh memiliki performa sebagai berikut :

Tabel 4.16 Hasil Performa menghilangkan atribut tidak berpengaruh

Performa	Nilai
Accuracy	67,38 %

(Sumber : Hasil Perhitungan)

Tabel 4.16 merupakan hasil perhitungan dari *accuracy* yang dihitung secara manual dan memiliki hasil yang sama dengan output yang dihasilkan pada software *weka*. Hasil *accuracy* yang diperoleh dengan menghilangkan atribut yang tidak berpengaruh sebesar 67,38%.

## 5. Penarikan Kesimpulan

Dari penggunaan algoritme *C4.5* yang telah dilakukan menghasilkan nilai performa dengan tiga tahapan yang berbeda, tahap pertama yaitu pengujian dengan *imputasi missing values* dan yang kedua pengujian tanpa *imputasi missing values* serta yang ketiga yaitu pengujian dengan menghilangkan atribut yang tidak berpengaruh. Berdasarkan pengujian diatas diperoleh masing-masing hasil akurasi sebesar 63,44% pengujian dengan *imputasi missing values* dan akurasi sebesar 68,45% tanpa *imputasi missing values* serta pengujian dengan menghilangkan atribut yang tidak berpengaruh sebesar 67,38%. Dilihat dari ketiga hasil akurasi tersebut algoritme *C4.5* mempunyai nilai akurasi yang cukup baik.

