

BAB IV HASIL DAN PEMBAHASAN

4.1 Bussiness Understanding atau Pemahaman Bisnis

Tujuan dari penelitian ini adalah untuk membandingkan model algoritma yang memiliki akurasi sangat tinggi menggunakan metode algoritma *Decision Tree C4.5* dan *Support Vector Machine*.

4.2 Data Understanding atau Pemahaman Data

Dataset yang digunakan memiliki 18 atribut terdiri dari data nominal dan numerik. Kelas data sasaran adalah tidak menerima beasiswa dan menerima beasiswa. Untuk atribut yang digunakan dapat dilihat pada tabel 4.1 dibawah ini :

Tabel 4.1 Atribut Dataset







No	Atribut	Skala Pengukuran
1	Tahun	[2019,2020,2021]
2	Nama Siswa	ID
3	Status DTKS	[0,1] 0. Belum Terdata, 1. Terdata
4	Pekerjaan Ayah	[1,2,3,4,5,6] 1. Tidak Bekerja, 2. Buruh, 3. Petani, 4. Wirausaha, 5. Peg. Swasta, 6. PNS
5	Penghasilan Ayah	[0-10000000]
6	Status Ayah	[1,2] 1. Meninggal, 2. Hidup
7	Pekerjaan Ibu	[1,2,3,4,5,6] 1. Tidak Bekerja, 2. Buruh, 3. Petani, 4. Wirausaha, 5. Peg. Swasta, 6. PNS

Tabel 4.1 Atribut Dataset

No	Atribut	Skala Pengukuran
8	Penghasilan Ibu	[0-10000000]
9	Status Ibu	[0,1] 1. Hidup, 0. Meninggal
10	Jumlah Tanggungan	[1,2,3,4,5,6]
11	Kepemilikan Rumah	[1,2]1. Kontrakan, 2. Sendiri
12	Sumber Listrik	[1,2]1. Numpang, 2. PLN
13	Luas Tanah	[0-1000]
14	Luas Bangunan	[0-250]
15	Sumber Air	[1,2]1. Kemasan, 2. Sumur
16	MCK	[1,2]1. Berbagi Pakai, 2. Kepemilikan Sendiri Didalam
17	Prestasi	[0,1]0. Tidak Ada, 1. Ada
18	BEASISWA	[Tidak Diterima, Diterima]

4.3 Data Preparation atau Persiapan Data

Pengolahan data awal yang dilakukan pada data mining untuk mengubah data mentah (raw data) yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih yang kemudian digunakan untuk proses pengolahan data selanjutnya. Berdasarkan sebaran kuisioner yang dilakukan diperoleh hasil raw data sebanyak 1090 data. Potongan dataset bisa dilihat pada gambar 4.1 berikut:

Luas Tanah  ▼ <i>integer</i>	Luas Bang...  ▼ <i>integer</i>	Sumber Air  ▼ <i>integer</i>	MCK  ▼ <i>integer</i>	Prestasi  ▼ <i>integer</i>	BEASISWA  ▼ <i>binominal label</i>
50	25	2	2	0	DITERIMA
0	0	2	1	0	DITERIMA
220	180	2	2	0	DITERIMA
220	180	2	1	0	DITERIMA
25	25	2	2	0	DITERIMA
100	50	2	2	0	DITERIMA
0	0	2	1	0	DITERIMA
0	0	2	1	1	DITERIMA
220	180	2	2	0	TIDAK DITERIMA
220	180	2	2	0	TIDAK DITERIMA
99	50	1	2	1	DITERIMA
99	50	1	2	0	DITERIMA

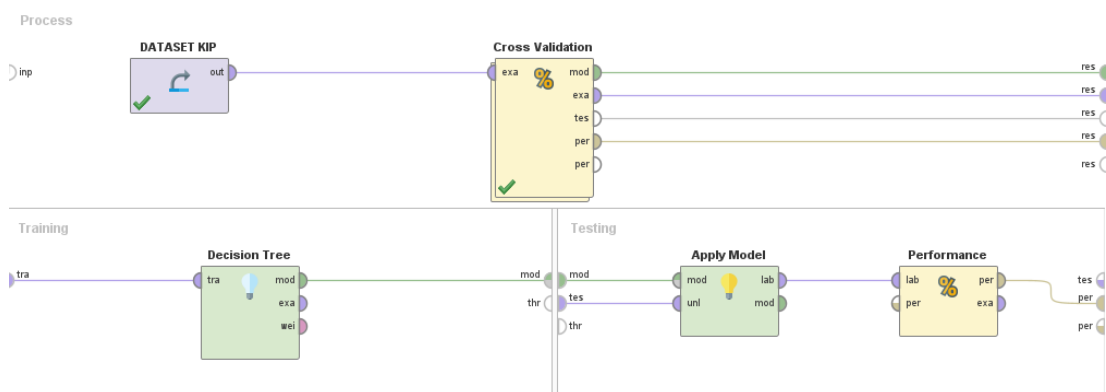
Gambar 4.1 Potongan Dataset

Pada tahap ini dilakukan pengubahan *type* variabel dan perubahan target/label

4.4 Pemodelan

4.4.1 Penelitian Menggunakan Algoritma *Decision Tree* C4.5

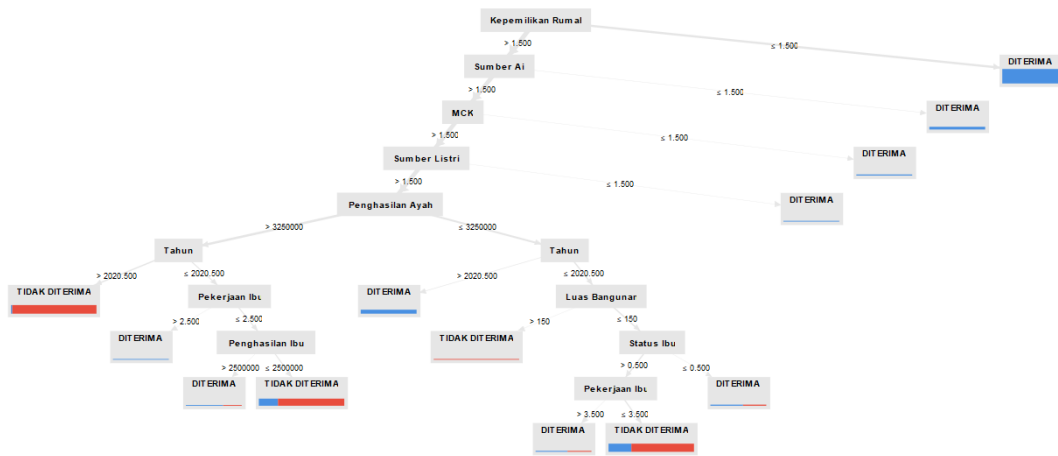
Pada algoritma *Decision Tree*, record yang sudah di import ke *Rapid Miner* digunakan untuk menentukan pola pohon keputusan. Penerapan data pada *Rapid Miner* digunakan untuk Klasifikasi penerima beasiswa menggunakan algoritma *Decision Tree* ditunjukkan pada gambar 4.2 dibawah ini:



Gambar 4.2 Model Klasifikasi Algoritma C4.5

Gambar 4.3 menunjukkan proses penerapan klasifikasi menggunakan *tool rapidminer*. Data yang sudah disiapkan diimport ke *rapidminer* untuk melakukan pengujian menggunakan *Cross Validation* untuk membagi data menjadi data training dan data testing karena data yang digunakan adalah *supervised* dan algoritma yang digunakan adalah algoritma *Decision Tree*. Tahapan tersebut merupakan tahapan pengolahan data tahap akhir dalam proses klasifikasi menggunakan model algoritma C4.5, pada tahap penerapan model data yang diolah dapat menjadi informasi berupa model yang terbentuk dari proses pengklasifikasian.

Setelah melakukan beberapa langkah diatas dalam proses klasifikasi metode algoritma C4.5 maka akan diperoleh model yang terbentuk dari proses pengklasifikasian algoritma C4.5 berupa pohon keputusan seperti gambar 4.3 dibawah ini.



Gambar 4.3 Pohon Keputusan

Gambar 4.3 merupakan gambar pohon keputusan yang merupakan output dari proses klasifikasi menggunakan algoritma C4.5. Pohon keputusan terbentuk berdasarkan node. Node dalam pohon keputusan merupakan variabel-variabel yang digunakan dalam penelitian. Berdasarkan gambar 4.3 maka faktor dari variabel yang paling dominan dalam mempengaruhi klasifikasi penerima beasiswa yaitu kepemilikan rumah. Selain pohon keputusan seperti pada gambar 4.3 juga terdapat *text view* yang menjelaskan berdasarkan pohon keputusan seperti pada gambar 4.4 dibawah ini.

```

Tree
Kepemilikan Rumah > 1.500
| Sumber Air > 1.500
| | MCK > 1.500
| | | Sumber Listrik > 1.500
| | | | Penghasilan Ayah > 3250000
| | | | | Tahun > 2020.500: TIDAK DITERIMA {DITERIMA=3, TIDAK DITERIMA=200}
| | | | | Tahun <= 2020.500
| | | | | | Pekerjaan Ibu > 2.500: DITERIMA {DITERIMA=2, TIDAK DITERIMA=0}
| | | | | | Pekerjaan Ibu <= 2.500
| | | | | | | Penghasilan Ibu > 2500000: DITERIMA {DITERIMA=2, TIDAK DITERIMA=1}
| | | | | | | Penghasilan Ibu <= 2500000: TIDAK DITERIMA {DITERIMA=37, TIDAK DITERIMA=128}
| | | | | | Penghasilan Ayah <= 3250000
| | | | | | | Tahun > 2020.500: DITERIMA {DITERIMA=85, TIDAK DITERIMA=0}
| | | | | | | Tahun <= 2020.500
| | | | | | | | Luas Bangunan > 150: TIDAK DITERIMA {DITERIMA=0, TIDAK DITERIMA=2}
| | | | | | | | Luas Bangunan <= 150
| | | | | | | | | Status Ibu > 0.500
| | | | | | | | | | Pekerjaan Ibu > 3.500: DITERIMA {DITERIMA=4, TIDAK DITERIMA=3}
| | | | | | | | | | Pekerjaan Ibu <= 3.500: TIDAK DITERIMA {DITERIMA=51, TIDAK DITERIMA=140}
| | | | | | | | | | Status Ibu <= 0.500: DITERIMA {DITERIMA=7, TIDAK DITERIMA=5}
| | | | | | Sumber Listrik <= 1.500: DITERIMA {DITERIMA=3, TIDAK DITERIMA=0}
| | | MCK <= 1.500: DITERIMA {DITERIMA=11, TIDAK DITERIMA=0}
| | Sumber Air <= 1.500: DITERIMA {DITERIMA=53, TIDAK DITERIMA=0}
Kepemilikan Rumah <= 1.500: DITERIMA {DITERIMA=353, TIDAK DITERIMA=0}

```

Gambar 4.4 Text View pohon keputusan

Gambar 4.4 merupakan *text view* pohon keputusan merupakan aturan (*rule*) yang didapatkan berdasarkan perhitungan dan pengujian data pada setiap atribut menggunakan algoritma C4.5. Selain pohon keputusan seperti pada gambar

4.3 dan *Text View* pohon keputusan seperti gambar 4.4 pengujian data menggunakan algoritma C4.5 juga diperoleh tabel hasil akurasi seperti pada gambar 4.5 dibawah ini Hasil pengujian dapat kita lihat pada gambar 4.5 dibawah ini.

accuracy: 90.37% +/- 2.17% (micro average: 90.37%)			
	true DITERIMA	true TIDAK DITERIMA	class precision
pred. DITERIMA	514	8	98.47%
pred. TIDAK DITERIMA	97	471	82.92%
class recall	84.12%	98.33%	
precision: 83.03% +/- 3.10% (micro average: 82.92%) (positive class: TIDAK DITERIMA)			
	true DITERIMA	true TIDAK DITERIMA	class precision
pred. DITERIMA	514	8	98.47%
pred. TIDAK DITERIMA	97	471	82.92%
class recall	84.12%	98.33%	
recall: 98.33% +/- 3.07% (micro average: 98.33%) (positive class: TIDAK DITERIMA)			
	true DITERIMA	true TIDAK DITERIMA	class precision
pred. DITERIMA	514	8	98.47%
pred. TIDAK DITERIMA	97	471	82.92%
class recall	84.12%	98.33%	

Gambar 4.5 Confusion matrix Hasil klasifikasi Algoritma C4.5

Gambar 4.5 adalah *confusion matrix* yang menunjukkan hasil eksperimen, didalam *confusion matrix* kita dapat melihat hasil *accuracy*, *precision*, dan *recall*. *accuracy* yang dihasilkan adalah 90.37%, *precision* 83.03% dan *recall* 98.33%. Hasil akurasi pada penelitian sebelumnya dengan eksperimen penelitian ini dengan menggunakan algoritma *Decision Tree* untuk klasifikasi penerima beasiswa ditunjukkan pada tabel 4.2 dibawah ini :

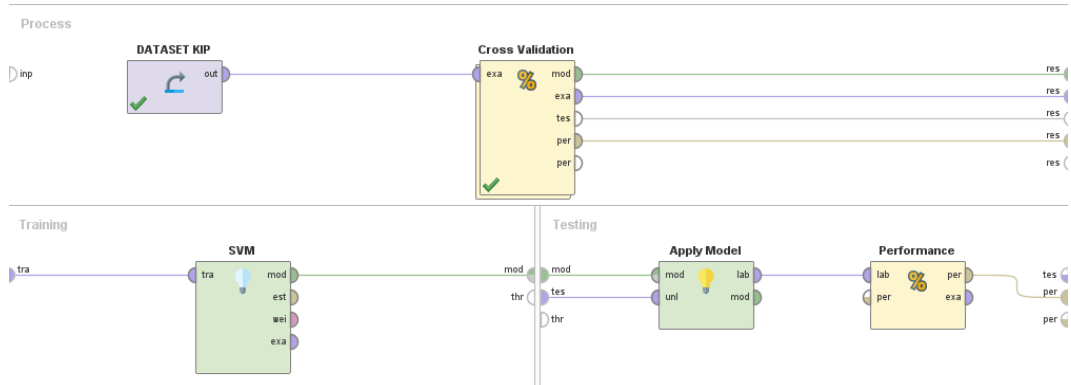
Tabel 4.2 Perbandingan Hasil Akurasi Algoritma C4.5

Penelitian	Akurasi
Penelitian Sebelumnya (Dita Noviana 2019)	88.3%
Penelitian saat ini	90,37%

4.4.3 Penelitian Menggunakan Algoritma *Support Vector Machine*

Proses klasifikasi menggunakan model *Support Vector Machine* digunakan untuk menggambarkan atau memprediksi peluang berdasarkan masing-masing kondisi. Penerapan data pada *Rapid Miner* digunakan untuk Klasifikasi penerima

beasiswa menggunakan algoritma *Support Vector Machine* ditunjukkan pada gambar 4.6 dibawah ini:



Gambar 4.6 Model Klasifikasi Algoritma SVM

Gambar 4.6 menunjukkan proses penerapan klasifikasi menggunakan *tool rapidminer*. Data yang sudah disiapkan diimport ke rapidminer untuk melakukan pengujian menggunakan *Cross Validation* untuk membagi data menjadi data training dan data testing karena data yang digunakan adalah *supervised* dan algoritma yang digunakan adalah algoritma *Support Vector Machine*. Berdasarkan gambar 4.6 yang telah dibangun pada aplikasi *Rapid Miner* maka diperoleh hasil sebagai berikut:

Attribute	Weight
Tahun	-0.141
Status DTKS	0.973
Pekerjaan Ayah	0.249
Penghasilan Ayah	0.537
Status Ayah	-0.026
Pekerjaan Ibu	-0.053
Penghasilan Ibu	0.053
Status Ibu	0.078
Jumlah Tanggungan	-0.070
Kepemilikan Rumah	1.045
Sumber Listrik	0.176
Luas Tanah	-0.132
Luas Bangunan	0.349
Sumber Air	0.626
MCK	0.831
Prestasi	-0.109

Gambar 4.7 Hasil Pengujian *Support Vector Machine*

Gambar 4.7 merupakan tabel distribusi hasil pengujian dari model *Support Vector Machine*. Pada tabel tersebut diuraikan prediksi berdasarkan *weight* dari masing-masing atribut. Selain tabel seperti pada gambar 4.7 data menggunakan model *Support Vector Machine* juga memperoleh hasil akurasi seperti pada gambar 4.8 dibawah ini.

accuracy: 86.42% +/- 2.36% (micro average: 86.42%)			
	true DITERIMA	true TIDAK DITERIMA	class precision
pred. DITERIMA	532	69	88.52%
pred. TIDAK DITERIMA	79	410	83.84%
class recall	87.07%	85.59%	
precision: 84.14% +/- 4.30% (micro average: 83.84%) (positive class: TIDAK DITERIMA)			
	true DITERIMA	true TIDAK DITERIMA	class precision
pred. DITERIMA	532	69	88.52%
pred. TIDAK DITERIMA	79	410	83.84%
class recall	87.07%	85.59%	
recall: 85.62% +/- 5.58% (micro average: 85.59%) (positive class: TIDAK DITERIMA)			
	true DITERIMA	true TIDAK DITERIMA	class precision
pred. DITERIMA	532	69	88.52%
pred. TIDAK DITERIMA	79	410	83.84%
class recall	87.07%	85.59%	

Gambar 4.8 Confusion matrix Hasil klasifikasi Algoritma SVM

Gambar 4.8 adalah *confusion matrix* yang menunjukkan hasil eksperimen, didalam *confusion matrix* kita dapat melihat hasil *accuracy*, *precision*, dan *recall*. *Accuracy* yang dihasilkan adalah 86.42% , *precision* 84.14% dan *recall* 85.62%.

Hasil akurasi pada penelitian sebelumnya dengan eksperimen penelitian ini dengan menggunakan algoritma *Support Vector Machine* untuk klasifikasi penerima beasiswa ditunjukkan pada tabel 4.3 dibawah ini :

Tabel 4.3 Perbandingan Hasil Akurasi Algoritma SVM

Penelitian	Akurasi
Penelitian Sebelumnya (Lukman 2017)	83.98%
Peneltian saat ini	86.42%

Untuk gambaran *Kernel Model* dari metode *Support Vector Machine* dapat dilihat pada gambar 4.11 dibawah ini.

Kernel Model

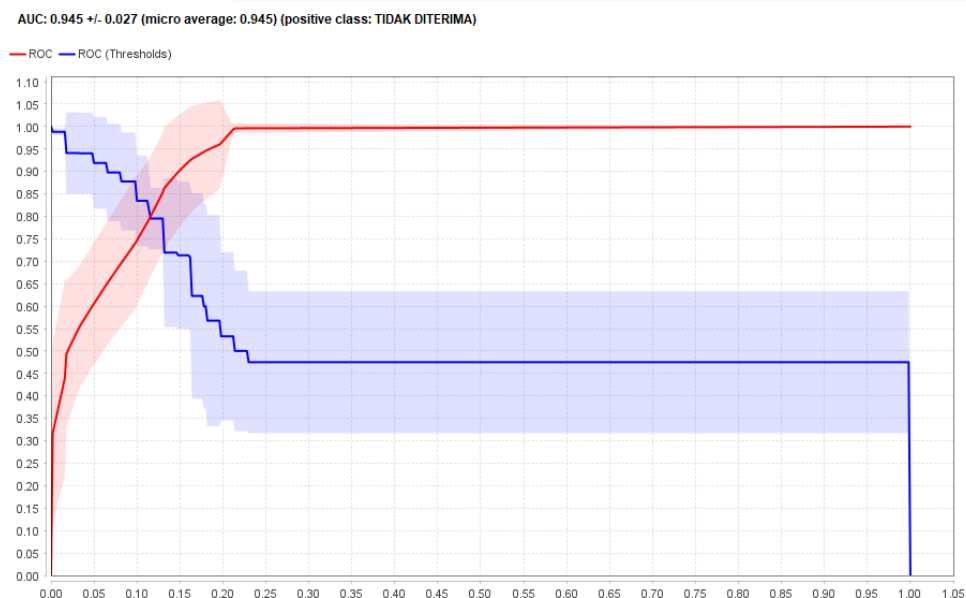
Total number of Support Vectors: 1090
Bias (offset): -1.368

```
w[Tahun] = -0.141
w[Status DTKS] = 0.973
w[Pekerjaan Ayah] = 0.249
w[ Penghasilan Ayah ] = 0.537
w[Status Ayah] = -0.026
w[Pekerjaan Ibu] = -0.053
w[ Penghasilan Ibu ] = 0.053
w[Status Ibu] = 0.078
w[Jumlah Tanggungan] = -0.070
w[Kepemilikan Rumah] = 1.045
w[Sumber Listrik] = 0.176
w[Luas Tanah] = -0.132
w[Luas Bangunan] = 0.349
w[Sumber Air] = 0.626
w[MCK] = 0.831
w[Prestasi] = -0.109
```

Gambar 4.9 Kernel Model Support Vector Machine

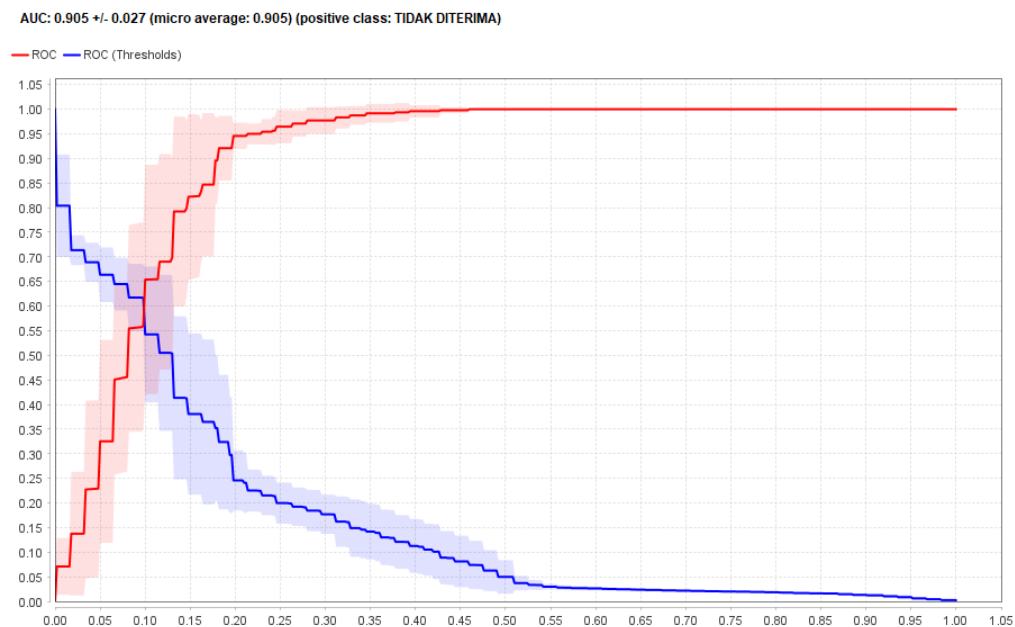
Hasil dari klasifikasi data penerimaan beasiswa yang diuji menggunakan *Support Vector Machine* dapat kita lihat bahwa atribut kepemilikan rumah berkontribusi lebih besar dari pada atribut lainnya.

Selain *Confusion matrix* untuk mengetahui kinerja dari eksperimen ini kita dapat mengandalkan kurva AUC yang dihasilkan. Perbandingan hasil Kurva AUC menggunakan algoritma C4.5 dan SVM dapat kita lihat pada gambar 4.10 dan 4.11 dibawah ini:



Gambar 4.10 Kurva AUC C4.5

Gambar 4.10 menunjukkan kinerja algoritma C.45 dalam mengklasifikasi penerima beasiswa dengan menghasilkan *Area Under Curva* (AUC) 0.945 (*Excellent Classification*).



Gambar 4.11 Kurva AUC SVM

Gambar 4.11 menunjukkan kinerja algoritma SVM dalam mengklasifikasi penerima beasiswa dengan menghasilkan *Area Under Curva* (AUC) 0.905 (*Excellent Classification*).

4.5 Evaluasi

Dalam pembuatan model klasifikasi, dapat digunakan banyak metode. Model *classifier* yang dihasilkan untuk melakukan perhitungan dimana nantinya perhitungan tersebut diambil sebagai suatu tetapan. Ketetapan yang digunakan dalam algoritma svm adalah mencari bobot, *bias*, *support vektor* dan juga *hyperplane* dan juga penentuan kelas yang optimal. Sedangkan untuk *C4.5* tetapan yang digunakan adalah *entropy* dan juga *gain* untuk membangun sebuah pohon keputusan sebagai implementasinya.

Proses klasifikasi yang dimiliki ke dua algoritma ini berbeda beda dalam proses mengklasifikasikan. Terlihat dari hasil akurasi yang ada dan terlihat pada saat melakukan proses perhitungan manual. Pada *SVM* saat mengklasifikasikan terlebih dahulu mencari bobot dari setiap atribut nya. Di dalam pembobotan ini dipengaruhi

dari jumlah data pada satu *record* data nya. Proses pengklasifikasian *SVM* mendapatkan nilai akurasi yang rendah terlihat pada saat memproses *SVM* secara manual. Pada saat proses *SVM* tergolong klasifikasi yang generalisasi, karena pada proses nya mengambil perhitungan berdasarkan *hyperplane*. Sedangkan pada algoritma *C4.5* hasil akurasi menunjukkan cukup lebih baik dari *SVM*. Hal tersebut dapat terjadi saat pengujian. Dapat dikatakan lebih baik dari *svm* karena dalam tahapan klasifikasi tidak dilakukan secara general. Yang dimaksudkan tidak dengan general, di setiap akan memulai eksekusi di setiap atributnya *C4.5* mengelompokkan setiap atributnya dengan sangat baik dan memberi perlabelan atau range nilai pada setiap proses eksekusi per atribut nya. Sehingga klasifikasi nya dapat dikatakan lebih khusus. Karena dilakukan dengan memperlakukan proses per atribut nya. Di buktikan juga pada saat satu kali proses mencari nilai entropy per partisi data, hanya untuk mendapatkan salah satu nilai *gain* terbesar dari setiap atribut. Metode klasifikasi bisa dievaluasi berdasarkan beberapa kriteria seperti tingkat akurasi, presisi, *recall*, dan hasil AUC. Dari hasil evaluasi ternyata Algoritma *C4.5* terbukti paling akurat dibanding Algoritma *SVM* untuk perbandingan Hasil penelitian dengan menggunakan algoritma *Decision Tree C4.5* dan algoritma *Support Vector Machine* ditunjukkan pada tabel 4.4 dibawah ini:

Tabel 4.4 Perbandingan Hasil Penelitian

Metode	<i>accuracy</i>	<i>precision</i>	<i>recall</i>	AUC
C4.5	90.37%	83.03%	98.33%	0.945
SVM	86.42%	84.14%	85.62%	0.905

Pada tabel 4.4 disajikan hasil dari pengujian algoritma *C4.5* dan *SVM* kemudian Hasil dari eksperimen ini menghasilkan tingkat akurasi yang tinggi dengan penambahan seleksi fitur yaitu sebesar 90.46%, *precision* 82.47% dan *recall* 99.79% serta nilai AUC 0.950 termasuk dalam predikat *Excellent Classification* (0.90 – 1.00). Metode ini bisa digunakan untuk penentuan beasiswa KIP dan membantu bagian kemahasiswaan dalam seleksi calon penerima beasiswa yang jumlah kuotanya sekitar 200 dengan pendaftar yang melebihi kuota yang sudah ditentukan oleh pemerintah.