

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN METODE *DESICION TREE* DAN *ARTIFICIAL NEURAL NETWORK*

Eko Prasetyo Rohmawan
Dosen Universitas Dehasen Bengkulu
Jl. Meranti Raya No. 32 Sawah Lebar Bengkulu
Sur-el : prasetiyoko1@gmail.com

Abstract : *Students graduation on time is one of assessment in accreditation process of university. The problem is how to know the students will be able to pass on time or not. In this study focus on how the prediction of graduation students on time with data mining using decision tree algorithm and artificial neural network. This research uses data of students who have passed the data as training and students who have not passed as data testing. This study aims to determine the graduation of students on time or not, the expected results can provide information and input for universities in making policy for improvement in the future.*

Keywords: Graduation, Data Mining, Decision Tree, Artificial Neural Network

Abstrak : *Kelulusan mahasiswa tepat waktu merupakan salah satu penilaian dalam proses akreditasi perguruan tinggi. Permasalahannya adalah bagaimana mengetahui mahasiswa nantinya bisa lulus tepat waktu atau tidak. Dalam penelitian ini fokus pada bagaimana prediksi kelulusan mahasiswa tepat waktu dengan data mining menggunakan algoritma decision tree dan artificial neural network. Penelitian ini menggunakan data mahasiswa yang sudah dinyatakan lulus sebagai data training dan mahasiswa yang belum lulus sebagai data testing. Penelitian ini bertujuan untuk mengetahui kelulusan mahasiswa tepat waktu atau tidak, yang diharapkan hasilnya dapat memberikan informasi dan masukan bagi pihak perguruan tinggi dalam membuat kebijakan demi perbaikan di masa yang akan datang.*

Kata Kunci : Kelulusan, Data Mining, Decision Tree, Artificial Neural Network

1. PENDAHULUAN

Kelulusan merupakan salah satu item penilaian dalam proses akreditasi institusi suatu perguruan tinggi. Elemen penilaian pada lulusan perguruan tinggi salah satunya adalah perguruan tinggi memiliki angka efisien edukasi yang ideal (BANPT, 2007). Sehingga jika mahasiswa lulusnya tepat waktu akan membantu penilaian akreditasi suatu perguruan tinggi. Di sisi lain, mahasiswa akan beruntung jika lulus tepat waktu karena tidak akan membayar biaya kuliah lagi dan bisa cepat bekerja setelah lulus kuliah.

Namun waktu kelulusan mahasiswa tidak selalu dapat dideteksi secara dini, sehingga bisa mengakibatkan keterlambatan lulusan. Untuk mengatasi hal tersebut perlu ada teknik untuk bisa melakukan prediksi terhadap kelulusan mahasiswa. Adapun teknik yang sering digunakan adalah dengan menggunakan data mining. Dan metode yang sering digunakan untuk prediksi kelulusan mahasiswa adalah metode klasifikasi.

Menurut David hand, Heikki Mannila, dan Padhraic Smyth dalam Widodo, Handayanto dan Herlawati (2013) data mining adalah analisa terhadap data (biasanya data yang berukuran

besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut. Data mining sudah ada sejak lama dan teori-teorinya pun sudah banyak dibahas dalam literatur. Teori-teori tersebut antara lain: Naive-Bayes dan Nearest Neighbour, Pohon Keputusan, Aturan Asosiasi, k-Means Clustering, dan Text Mining (Bramer, 2007). Sedangkan perkembangan terkini menghadirkan algoritma-algoritma yang baru dikembangkan antara lain: Jaringan Syaraf Tiruan (JST), Algoritma Genetik, *Fuzzy C-Means*, *Support Vector Machine* (SVM), dan lain-lain (Daniel, 2006).

Metode klasifikasi merupakan pendekatan untuk menjalankan fungsi klasifikasi dalam data mining yaitu menggolongkan data. Teknik klasifikasi ini dapat pula digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma *Decision Tree C.45*, *Artificial Neural Networks* (ANN), *K-Nearest Neighbour* (KNN), Algoritma *Naive Bayes*, Algoritma Genetik, *Rough Set*, Metode Berbasis Aturan, *Memory Based Reasoning*, dan *Support Vector Machine* (Widodo dkk., 2013).

Penelitian yang menggunakan data mining pada data set akademik dan kemahasiswaan telah banyak dilakukan, antara lain adalah penelitian yang dilakukan Alvi Syahrin (2013) yaitu implementasi algoritma k-Means untuk klusterisasi mahasiswa berdasarkan waktu kelulusan. Penelitian ini mengimplementasi algoritma *K-Means* ke dalam studi kasus tersebut. Aplikasi ini terdiri dari empat fungsi,

yakni '*Cluster*', '*Show Centroid*', '*Show the Graphic*', dan '*Evaluate the Cluster*'. '*Cluster*' digunakan untuk membagi data mahasiswa ke dalam kelas-kelas berdasarkan prediksi waktu kelulusannya. '*Show Centroid*' digunakan untuk melihat centroid akhir dari proses iterasi. '*Show the Graphic*' digunakan untuk menampilkan posisi tingkat kelulusan mahasiswa. '*Evaluate the Cluster*' digunakan untuk menghitung nilai optimal dari hasil cluster tersebut.

Penelitian serupa dilakukan Nuqson M. Huda (2010) yaitu aplikasi data mining untuk menampilkan informasi tingkat kelulusan mahasiswa. Dalam penelitian ini kategori tingkat kelulusan diukur dari lama studi dan IPK. Algoritma yang digunakan adalah algoritma *apriori*, informasi yang ditampilkan berupa nilai *support* dan *confidence* dari masing-masing kategori tingkat kelulusan.

Dalam penelitian ini penulis akan membandingkan dua metode data mining yaitu metode *decision tree* dan *artificial neural network*. Kedua metode tersebut sudah banyak digunakan dalam penelitian-penelitian sebelumnya, antara lain penelitian Andri (2013) Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma. Dalam penelitian tersebut hasil prediksi kelulusan memiliki tingkat akurasi di atas 90%. Adapun penelitian yang berkaitan dengan metode *artificial neural network* antara lain penelitian yang dilakukan oleh Kusumawati (2015) Prediksi kelulusan mahasiswa menggunakan metode *neural network* dan *particle swarm optimization*. Dalam penelitian tersebut tingkat akurasi yang dihasilkan sebesar 87,31%.

2. METODOLOGI PENELITIAN

Dalam penelitian ini teknik data mining yang digunakan adalah *Decision Tree* dan *Artificial Neural Network*.

Data yang digunakan dalam penelitian ini sebagai berikut:

a. Data training dan data testing

Data ini akan digunakan sebagai proses mining dan pengujian, berupa sampel data induk dan data akademik mahasiswa angkatan 2006-2010 yang sudah dinyatakan lulus. Data ini memiliki atribut NIM, jenis kelamin, asal sekolah, jalur masuk, nilai ujian nasional, gaji orang tua, IP semester 1-4, IPK semester 1-4 dan keterangan lulus.

b. Data target

Data ini merupakan sampel data induk data akademik mahasiswa angkatan 2011-2012 yang diasumsikan belum lulus. Data ini memiliki atribut NIM, jenis kelamin, asal sekolah, jalur masuk, nilai ujian nasional, gaji orang tua, IP semester 1-4, dan IPK semester 1-4. Data ini akan dijadikan testing berdasarkan data training.

Algoritma *decision tree* yang digunakan adalah C4.5. secara umum untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai node akar
2. Buat cabang untuk tiap-tiap nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai node akar, didasarkan pada nilai *Gain* tertinggi dari atribut-

atribut yang ada. Untuk menghitung *Gain* digunakan rumus seperti tertera dalam persamaan berikut: $Gain(S,A) = Entropy(S) -$

$$\sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S_i| : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai *Gain* adalah mencari nilai *Entropy*. *Entropy* digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan keluaran atribut. Rumus dasar dari *Entropy* tersebut adalah sebagai berikut:

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

Keterangan:

S : himpunan kasus

A : Fitur

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

Jaringan syaraf tiruan merupakan salah satu representasi buatan dari otak manusia yang selalu mencoba untuk mensimulasikan proses pembelajaran otak manusia tersebut (M.F. Andrijasa, 2010). Jaringan syaraf tiruan tercipta sebagai suatu generalisasi model matematis dari pemahaman manusia (human cognition) yang didasarkan atas asumsi sebagai:

1. Pemrosesan informasi terjadi pada elemen sederhana yang disebut neuron.
2. Isyarat mengalir diantara sel syaraf (neuron) melalui suatu sambungan penghubung.
3. Setiap sambungan penghubung memiliki

bobot yang bersesuaian

4. Setiap sel syaraf akan merupakan fungsi aktivasi terhadap isyarat hasil penjumlahan berbobot yang masuk kepadanya untuk menentuka isyarat keluarannya

Adapun pemodelan dengan Jaringan Syaraf Tiruan secara umum terdiri dari empat langkah yaitu:

1. Penyiapan data
2. Pemilihan arsitektur
3. Pembelajaran (learning)
4. Pengujian (testing)

3. HASIL

Dalam pengumpulan data terdapat sumber data, sumber data yang dihimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder. Data yang diperoleh adalah data sekunder karena diperoleh dari *database* mahasiswa yang dimiliki oleh Universitas Dehasen Bengkulu. Data yang dikumpulkan adalah data mahasiswa Universitas Dehasen Bengkulu Program Studi Teknik Informatika tahun akademik 2009, 2010 dan 2011.

Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 600 data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan data awal (*preparation data*). Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan adalah sebagai berikut:

1. *Data validation*, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*),

data yang tidak konsisten dan data yang tidak lengkap (*missing value*).

2. *Data integration and transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorial. Data ditransformasikan ke dalam *software RapidMiner*.
3. *Data size reduction and dicritization*, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif. Dalam penelitian ini atribut yang tak relevan seperti nim, nama, tanggal lahir, periode masuk, status, IP semester 7, IPK semester 7, total SKS semester 7, IP semester 8, IPK semester 8 dan total SKS semester 8 dihapuskan.

Setelah dilakukan tahapan *data preparation*, total data berkurang menjadi 553 *record*. Adapun rekap data berdasarkan jenis kelamin, angkatan dan keterangan lulus tepat atau terlambat sebagai berikut.

Tabel 1. Rekapitulasi data mahasiswa lulus tepat dan terlambat berdasarkan angkatan dan jenis kelamin setelah *data preparation*

Angkatan	Jenis kelamin				Jumlah
	Laki-Laki		Perempuan		
	Tepat	Terlambat	Tepat	Terlambat	
2009	33	91	28	26	178
2010	31	89	34	31	185
2011	23	103	36	28	190
	87	283	98	85	553

Sumber: hasil pengolahan data mahasiswa setelah pengolahan data awal

3.1. Proses Pohon Keputusan (*Decision Tree*)

Proses pengklasifikasian data menggunakan metode *decison tree* ini untuk mengetahui

atribut-atribut yang sangat berpengaruh dalam proses prediksi. Metode ini berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Karena metode ini memadukan antara eksplorasi data dan pemodelan, maka sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

Dari tabel training diketahui bahwa jumlah kasus 553 kasus, mahasiswa yang termasuk kelas lulus berjumlah 185 Record, dan mahasiswa yang termasuk kelas tidak lulus berjumlah 368 record sehingga didapatkan *Entropy* nya sebagai berikut :

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

$$\text{Entropy (S)} = (-185/553.\text{Log}_2(185/553)) + (-368/553.\text{Log}_2(368/553))$$

$$\text{Entropy(S)} = 0,919497$$

Tabel 2. Perhitungan Node

Node	Jumlah data	Lulus tepat	Lulus terlambat	Entropy	Gain
Total	553	185	368	0,919497	
Jenis Kelamin					0,06331
Laki-laki	370	87	283	0,7866	
Perempuan	183	98	85	0,996357	
IP semester 1					0,08364
< 2,5	143	12	131	0,41583	
>= 2,5	410	173	237	0,98235	
				1	

IP semester 2					0,16876
< 2,5	185	8	177	0,25697	
>= 2,5	368	177	191	0,998956	
IP semester 3					0,18068
< 2,5	165	3	162	0,131107	
>= 2,5	388	182	206	0,997238	
IP semester 4					0,1987
< 2,5	182	4	178	0,152407	
>= 2,5	371	181	190	0,999575	
IP semester 5					0,18965
< 2,5	164	2	162	0,095017	
>= 2,5	389	183	206	0,997477	
IP semester 6					0,27653
< 2,5	224	4	220	0,129234	
>= 2,5	329	181	148	0,99273	
IPK Semester 2					0,12676
< 2,5	157	8	149	0,290442	
>= 2,5	396	177	219	0,99287	
IPK semester 3					0,17696
< 2,5	139	0	139	0	
>= 2,5	414	185	229	0,991837	

IPK semester 4					0,1 93 70
< 2,5	179	4	175	0,15 441 9	
>= 2,5	374	181	193	0,99 925 7	
IPK semester 5					0,1 86 41
< 2,5	162	2	160	0,09 597	
>= 2,5	391	183	208	0,99 704 9	
IPK semester 6					0,2 43 26
< 2,5	207	4	203	0,13 762 6	
>= 2,5	346	181	165	0,99 845 7	
SKS total semester 2					0,0 63 68
< 40	130	13	117	0,46 899 6	
>= 40	423	172	251	0,97 469 1	
SKS total semester 3					0,1 56 79
< 60	143	2	141	0,10 618 8	
>= 60	410	183	227	0,99 167 6	
SKS total semester 4					0,1 76 83
< 80	156	2	154	0,09 895 9	
>= 80	397	183	214	0,99 559 7	
SKS total semester 5					0,1 86 98
< 100	155	1	154	0,05 622	
>= 100	398	184	214	0,99	

					100 589 8
SKS total semester 6					0,1 80 00
< 120	158	2	156	0,09 794 1	
>= 120	395	183	212	0,99 610 8	

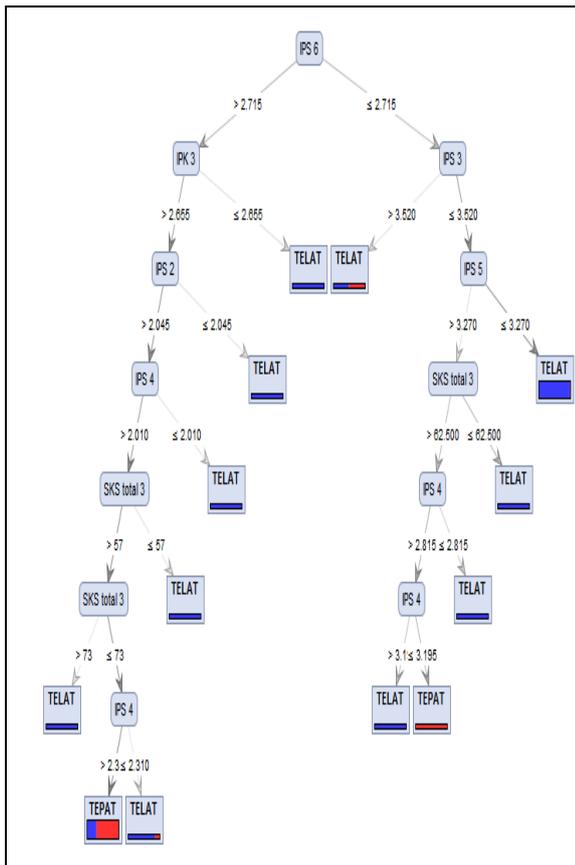
Sumber: hasil pengolahan data decision tree

3.2 Metode Artificial Neural Network

Metode kedua yang digunakan adalah *Artificial Neural Network*. Dan algoritma yang digunakan adalah *Backpropagation*. Atribut masukan yang digunakan sama dengan atribut pada *Decision Tree*.

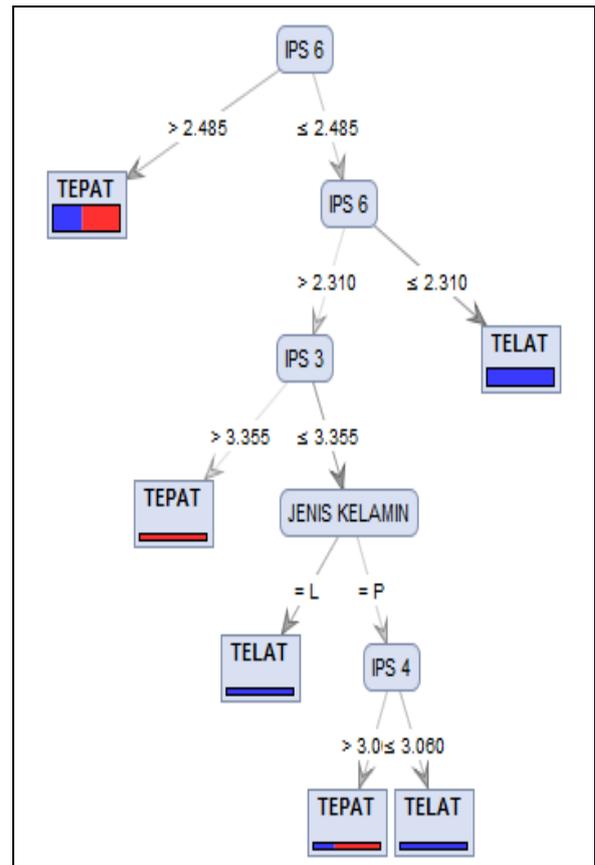
3.3 Implementasi Metode Decision Tree dengan RapidMiner

Implementasi dengan metode decision tree, variabel yang digunakan adalah: jenis kelamin, IP semester 1, IPK semester 1, total sks semester 1, semester 2, IP semester 2, IPK semester 2, total sks semester 2, semester 3, IP semester 3, IPK semester 3, total sks semester 3, semester 4, IP semester 4, IPK semester 4, total sks semester 4, semester 5, IP semester 5, IPK semester 5, total sks semester 5, semester 6, IP semester 6, IPK semester 6, total sks semester 6, keterangan (tepat/telat).



Gambar 1. Pohon keputusan berdasarkan gain ratio

Gain ratio merupakan varian dari *information gain*. Metode ini menghasilkan *information gain* untuk setiap atribut yang memberikan nilai atribut yang seragam. Berdasarkan pohon keputusan di atas atribut yang dijadikan *root* awal adalah Indeks Prestasi semester 6, karena pada tahap penghitungan node awal memiliki *gain* tertinggi yaitu 0,276537056. Atribut di bawahnya adalah IPK semester 3 dan IP semester 3, jika IP semester 6 lebih dari 2,715 maka akan dilanjutkan ke atribut IPK semester 3 dan jika IP semester 6 kurang dari 2,715 maka akan dilanjutkan ke atribut IP semester 3. Kemudian seterusnya sampai ke simpul TEPAT dan TELAT. Dan berikut ini hasil pohon keputusan berdasarkan informatif *gain*.



Gambar 2. Pohon keputusan berdasar information gain

Information gain adalah metode yang semua *entropy* dihitung, kemudian atribut dengan *entropy* minimum yang dipilih untuk dilakukan perpecahan pohon (*split*). Berikut hasil pengujian dengan menggunakan metode *decision tree*.

Tabel 3. hasil pengujian metode *decision tree*

<i>Accuracy</i> : 74,51			
	True TELAT	True TEPAT	class precision
Pred. TELAT	73	7	91,25%
Pred. TEPAT	32	41	56,16%
Class recall	69,52%	85,42%	

Berdasarkan hasil pengujian metode *decision tree*, tingkat akurasi yang dihasilkan adalah 74.5%. dengan prediksi telat sebesar 91,25% dan prediksi tepat sebesar 56,16%.

3.4. Implementasi Metode *Artificial Neural Network* dengan *RapidMiner*

Berikut hasil dari proses metode *artificial neural network* dengan *RapidMiner*:

Tabel 4. Bobot dari unit output

NODE	CLASS	
	TEPAT	TELAT
Node 1	2.300	-2.270
Node 2	2.593	-2.596
Node 3	1.693	-1.719
Node 4	1.745	-1.770
Node 5	3.006	-2.999
Node 6	1.781	-1.780
Node 7	3.239	-3.246
Node 8	2.334	-2.344
Node 9	1.642	-1.658
Node 10	1.789	-1.786
Node 11	1.772	-1.713
Node 12	5.488	-5.479
Threshold	-1.809	1.809

Dan berikut hasil pengujian dengan menggunakan *neural network*

Tabel 5. Hasil pengujian metode *artificial neural network*

Accuracy: 79,74

	True TELAT	True TEPAT	class precision
Pred. TELAT	82	8	91,11%
Pred. TEPAT	23	40	63,49%
Class recall	78,10%	83,33%	

Hasil pengujian dengan menggunakan *neural network* menunjukkan bahwa *accuracy* yang dihasilkan adalah 79,74%, dengan rincian prediksi telat sebesar 91,11% , dan prediksi tepat sebesar 63,49%

3.5. Perbandingan hasil pengujian antara metode *Decision Tree* dan *Artificial Neural Network*

Berdasarkan hasil dua pengujian sebelumnya, berikut rekap perbandingan antara metode *decision tree* dan *artificial neural network*.

Tabel 6. Perbandingan hasil pengujian

Metode	Accuracy
Decision Tree	74,51%
Neural Network	79,74%

Berdasarkan tabel 6, prediksi kelulusan mahasiswa tepat waktu dengan menggunakan metode *Artificial Neural Network* memiliki tingkat *accuracy* yang lebih tinggi jika dibandingkan dengan metode *decision tree* yaitu : 79,74%.

3.6. Analisis Perbandingan hasil pengujian antara metode *Decision Tree* dan *Artificial Neural Network*

Berdasarkan hasil pengujian di atas, dalam penelitian ini metode *artificial neural network* memiliki tingkat akurasi sedikit lebih tinggi jika dibandingkan dengan metode *decision tree*.

Ada beberapa analisis yang menyebabkan metode *artificial neural network* memiliki akurasi lebih tinggi, antara lain sebagai berikut:

- a. Dilihat dari karakteristik masing-masing metode. Metode *decision tree* bersifat eksploratif, metode ini akan mengeksplorasi semua atribut yang terdapat dalam data. Dan hasil yang didapatkan rata-rata lebih stabil.

Sedangkan metode *artificial neural network* bersifat heuristik, metode ini berusaha mendapatkan performa komputasi atau penyederhanaan konseptual, untuk meningkatkan keakuratan atau presisi.

- b. Data yang akan dilakukan mining. data yang digunakan biasanya terdapat dua, yaitu data dengan label (terdapat target) dan data tanpa label (tanpa target) biasanya berupa *time series*.

Kedua hal di atas yang menyebabkan metode *artificial neural network* memiliki tingkat akurasi lebih tinggi jika dibandingkan dengan metode *decision tree*. Dalam penelitian ini data yang digunakan berupa data label, yang mana jika menggunakan metode *artificial neural network*, tingkat akurasi yang dihasilkan lebih tinggi karena dalam *neural network* terdapat *backward* yang mengembalikan jika terdapat eror.

4. SIMPULAN

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa:

- a. Penggunaan data mining dengan metode *Decision Tree* dan *Artificial Neural Network* dapat diterapkan dalam melakukan prediksi kelulusan mahasiswa tepat waktu
- b. Berdasarkan hasil pengujian metode *decision tree* memiliki akurasi sebesar 74,51% dan *artificial neural network* sebesar 79,74%.
- c. Metode *artificial neural network* memiliki akurasi lebih tinggi jika dibandingkan

dengan *decision tree* karena data yang digunakan dalam penelitian ini berupa data label.

DAFTAR RUJUKAN

- Andri, A., Yesi, N. K., & Sri Murniati, M. (2013). *Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang*.
- Andrijasa, M.F., Mistianingsih. 2010. *Penerapan Jaringan Syaraf Tiruan Untuk Memprediksi Jumlah Pengangguran di Provinsi Kalimantan Timur Dengan Menggunakan Algoritma Pembelajaran Back propagation*. Jurnal Informatika Mulawarman Vol 5 No. 1 Februari 2010
- BANPT. 2007. *Akreditasi Institusi Perguruan Tinggi*. Jakarta: Depdiknas.
- Bramer, Max.2007. *Principles of Data Mining*. London:Springer-Verlag.
- Daniel T., Larose. 2006. *Data Mining, Methods and Models*. New Jersey: John Wiley & Son
- Huda, Nuqson M. 2010. *Aplikasi Data Mining untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa*. FMIPA Universitas Diponegoro.
- Kusumawati, D., Winarno, W. W., & Arief, M. R. (2015). *Prediksi Kelulusan Mahasiswa Menggunakan Metode Neural Network dan Particle Swarm Optimization*. SEMNASTEKNOMEDI A ONLINE, 3(1), 3-8.
- Syahrin, Alvi. 2013. *Implementasi Algoritma K-Means untuk Klasterisasi Mahasiswa berdasarkan prediksi waktu kelulusan*. Jawa Timur : UPN "Veteran".
- Widodo, dkk. 2013. *Penerapan Data Mining dengan MATLAB*. Bandung: Rekayasa Sains