

PREDIKSI CALON MAHASISWA BARU MENGGUNAKAN METODE KLASIFIKASI DECISION TREE

Mambang¹Finki Dona Marleny²

¹AKBID Sari Mulia, Banjarmasin

²Jurusan Sistem Informasi, STMIK Indonesia, Banjarmasin

e-mail: mbgcche@yahoo.co.id , finkidona@gmail.com

Abstrak

Sebelum penyelenggaraan pendidikan tenaga kesehatan memulai tahun ajaran baru, maka langkah awal akan dilaksanakan seleksi penerimaan mahasiswa baru yang berasal dari lulusan pendidikan menengah umum maupun kejuruan yang sederajat. Seleksi penerimaan mahasiswa baru ini bertujuan untuk menyaring calon mahasiswa dari berbagai latar belakang yang disesuaikan dengan standar yang telah ditentukan oleh lembaga. Dalam penelitian ini bagaimana akurasi algoritma C4.5 untuk memprediksi kelulusan calon mahasiswa baru. Model decision tree merupakan metode prediksi klasifikasi untuk membuat sebuah tree yang terdiri dari root node, internal node dan terminal node. Berdasarkan hasil eksperimen dan evaluasi yang dilakukan maka dapat disimpulkan bahwa Algoritma C4.5 dengan Uncertainty didapatkan Akurasi 80,39%, Precision 94,44%, Recall 75,00% sedangkan dengan Algoritma C4.5 dengan Information Gain Ratio Akurasi 88,24%, Precision 98,28%, Recall 83,82%.

Kata Kunci : Decision Tree, C4.5, Recall, Precision

Abstract

Prior to the organization of health education begin the new school year, then the first step will be carried out selection of new admissions from general secondary education graduates and vocational sederajat. Pada this study, predicting new students to take multiple data attributes. The model is a decision tree classification prediction method to create a tree consisting of a root node, internal nodes and terminal nodes. While the root node and internal nodes are variables / features, the terminal node is labeled kelas. Berdasarkan experimental results and evaluations are done, it can be concluded bahwa Algoritma C4.5 with 80.39% accuracy obtained Uncertainty, Precision 94.44%, Recall of 75.00 % while the C4.5 algorithm with Information Gain Ratio Akurasi 88.24%, 98.28% Precision, 83.82% Recall.

Keywords: Decision Tree, C4.5, Recall, Precision

1. PENDAHULUAN

Pendidikan merupakan kebutuhan, kenyataan ini disebabkan oleh karena pendidikan merupakan fenomena kehidupan yang tidak dapat dipisahkan dengan gejala sosial lainnya. Penyelenggaraan pembangunan nasional yang berwawasan kesehatan serta pembangunan kesehatan memerlukan berbagai jenis tenaga kesehatan kebidanan yang memiliki kemampuan melaksanakan pelayanan kesehatan. Khususnya dalam tenaga bidan professional yang memberikan pelayanan langsung di bidang kesehatan ibu dan anak.

Melihat realitas yang ada, maka akademi kebidanan sari mulia banjarmasin berusaha untuk mencetak tenaga-tenaga kesehatan, khususnya tenaga bidan yang professional dan berkualitas. Langkah awal sebelum memulai tahun ajaran baru yaitu dilaksanakan seleksi penerimaan mahasiswa baru yang berasal dari lulusan pendidikan menengah umum maupun kejuruan yang sederajat. Dalam proses seleksi penerimaan mahasiswa baru yang bertujuan untuk mendapatkan mahasiswa yang berkompeten diantara seluruh calon dan dapat dipilih menjadi tenaga kesehatan

yang profesional pada umumnya dan tenaga bidan pada khususnya. Seleksi penerimaan mahasiswa baru tidak hanya menilai dari kemampuan akademisnya saja, tetapi kemampuan lain yang dapat mendukung untuk menimbulkan kepekaan sosial. Serta minat yang tepat harus disesuaikan dengan kemampuan dan kepribadian, bukan karena pengaruh orang lain atau trend di lingkungan masyarakat saat itu [1]. Proses seleksi calon mahasiswa baru memiliki beberapa tahapan proses dimana para calon mahasiswa baru harus dapat mengumpulkan beberapa kriteria nilai yang di ujikan serta mengikuti serangkaian tes selain tes akademis, seperti tes psikotes dan kesehatan serta wawancara. Total dari keseluruhan nilai dari serangkaian tes ini yang akan menentukan kelulusan calon mahasiswa baru. Data proses seleksi calon mahasiswa baru dapat dieksplorasi untuk menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target menggunakan pohon keputusan algoritma C4.5 [2]. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan [3].

Malaya, dkk [4], dalam proses penerimaan mahasiswa di perguruan tinggi teknik di India menerapkan pengetahuan teknik keputusan berbasis fungsi heuristik pada algoritma C4.5, keuntungan dari heuristik adalah bahwa informasi perpecahan pernah mendekati nol, maka menghasilkan aturan set yang stabil dalam pohon keputusan. Zhiwu Liu [1], menggunakan aturan klasifikasi dan model prediksi untuk mengetahui faktor apa yang dapat mempengaruhi siswa dan efek pengajaran guru yang dapat diperiksa dari beberapa kebiasaan negatif atau perilaku belajar siswa. Adrian Costea [5], menggunakan metodologi yang didasarkan pada dua teknik, yaitu dengan pengelompokan data analisis cluster dan klasifikasi model prediksi untuk menentukan kemungkinan pada data kinerja ekonomi keuangan kemudian menentukan hasil nilai pada tiap baris data. Sementara itu Sotiris, dkk [6] memprediksi kinerja siswa untuk mengidentifikasi kinerja buruk dan dapat mengaktifkan tutor untuk mengambil langkah-langkah perbaikan pada tahap awal untuk memprediksi siswa dalam sistem pembelajaran. Smith Tsang dkk [7], dalam penelitiannya *Decision Trees For Uncertain Data*. Mengungkapkan bahwa Salah satu model klasifikasi paling populer adalah model pohon keputusan. Pohon keputusan yang populer karena praktis dan mudah dimengerti. Aturan juga bisa diekstrak dari pohon keputusan dengan mudah. Pathom dkk [8] menggunakan algoritma C4.5 untuk membantu menemukan model perencanaan pendaftaran mahasiswa dari data Indeks Prestasi Kumulatif dan nilai dari mahasiswa.

Pada penelitian ini memfokuskan tentang bagaimana menggunakan aturan klasifikasi dan model prediksi untuk menentukan kelulusan calon mahasiswa baru menggunakan metode decision tree C4.5.

2. METODE PENELITIAN

Metode penelitian yang dilakukan adalah metode penelitian eksperimen, dengan tahapan penelitian yang terdiri dari : (1) Metode pengumpulan data dan pengolahan data awal. (2) Model yang diusulkan, (3) Pengujian model (4) Hasil Eksperimen, dan (5) Evaluasi dan Validasi Hasil.

A. Pengumpulan dan pengolahan data awal

Data yang digunakan adalah data seleksi penerimaan calon mahasiswa baru pada AKBID Sari Mulia Banjarmasin yang memiliki beberapa data diantaranya adalah Data tes Kesehatan, MMPI, wawancara dan ujian tertulis.

Berikut contoh data awal seleksi penerimaan calon mahasiswa baru sebanyak 20 data dari 340 data yang akan di uji:

Tabel 1. Data seleksi penerimaan calon mahasiswa baru

Data	Kesehatan	MMPI	Wawancara	Status	MAT	INGG	IPA	IND	Total Nilai	Ujian Tulis	Hasil Akhir
1	direkomendasi	LULUS	direkomendasi	Blm menikah	11	17	11	7	46	4L	diterima
2	direkomendasi	LULUS	direkomendasi	Blm menikah	10	16	7	7	40	3L	diterima
3	direkomendasi	LULUS	direkomendasi	Blm menikah	8	12	8	7	35	3L	diterima
4	direkomendasi	LULUS	direkomendasi	Blm menikah	8	6	5	7	26	1L	tdk diterima
5	tdk direkomendasi	LULUS	direkomendasi	Blm menikah	10	4	9	5	28	2L	tdk diterima
6	direkomendasi	LULUS	direkomendasi	Blm menikah	8	4	5	5	22	1L	tdk diterima
7	direkomendasi	LULUS	direkomendasi	Blm menikah	8	8	4	6	26	2L	tdk diterima
8	direkomendasi	LULUS	direkomendasi	Blm menikah	12	7	7	3	29	1L	tdk diterima
9	tdk direkomendasi	LULUS	direkomendasi	Blm menikah	8	10	10	5	33	2L	tdk diterima
10	direkomendasi	LULUS	direkomendasi	Blm menikah	13	11	11	7	42	4L	diterima
11	direkomendasi	LULUS	direkomendasi	Blm menikah	14	12	5	8	39	3L	diterima
12	direkomendasi	LULUS	direkomendasi	Blm menikah	9	9	10	4	32	3L	diterima
13	direkomendasi	LULUS	direkomendasi	Blm menikah	12	8	9	8	37	4L	diterima
14	direkomendasi	LULUS	direkomendasi	Blm menikah	6	5	8	7	26	2L	tdk diterima
15	tdk direkomendasi	LULUS	direkomendasi	Blm menikah	9	5	10	9	33	3L	diterima
16	direkomendasi	LULUS	direkomendasi	Blm menikah	6	7	6	6	25	1L	tdk diterima
17	tdk direkomendasi	LULUS	direkomendasi	Blm menikah	7	2	8	5	22	1L	tdk diterima
18	direkomendasi	LULUS	direkomendasi	Blm menikah	8	8	13	5	34	2L	tdk diterima
19	direkomendasi	LULUS	direkomendasi	Blm menikah	10	7	9	4	30	2L	tdk diterima
20	direkomendasi	LULUS	direkomendasi	Blm menikah	12	16	5	8	41	3L	diterima
21	Obesitas	LULUS	direkomendasi	Blm menikah	6	6	2	6	20	1L	tdk diterima
22	direkomendasi	LULUS	direkomendasi	Blm menikah	9	7	4	6	26	2L	tdk diterima
23	direkomendasi	LULUS	direkomendasi	Blm menikah	5	3	7	6	21	1L	tdk diterima
24	TB Kurang	LULUS	direkomendasi	Blm menikah	8	4	3	6	21	1L	tdk diterima

B. Model/Metode Yang Diusulkan

Metode yang diusulkan untuk Prediksi mahasiswa baru adalah algoritma Decision Tree C4.5. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut[9]:

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

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada [10]. Untuk menghitung gain digunakan rumus seperti tertera dalam Rumus 1

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

Sedangkan perhitungan nilai entropy dapat dilihat pada rumus 2 berikut:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Perhitungan node 1, sebagai berikut:

Tabel 2. Perhitungan node

No	Cabang		Jumlah Kasus(S)	Diterima (S1)	Tidak diterima (S2)	Entropy	Gain
	Total		340	83	257	0.80182099	
1	KESEHATAN	Direkomendasi	313	79	234	0.750329046	0.02094332
		tdk direkomendasi	17	1	16	0.016185452	
		TB Kurang	5	2	3	0.014363174	
		Obesitas	4	0	4	0	
2	MMPI	Lulus	54	16	38	0.139243175	0.16162721
		Tdk Lulus	0	0	0	0	
		Maskulinitas	129	22	107	0.250011332	
		Feminimitas	97	27	70	0.250939272	
3	WAWANCARA	Direkomendasi	337	83	254	0.798240082	0.00358091
		tdk direkomendasi	3	0	3	0	
4	STATUS	Menikah	0	0	0	0	0.00000000
		Blm menikah	340	83	257	0.801820989	
5	KET Tertulis	4L	25	25	0	0	0.37729306
		3L	58	17	41	0.148871757	
		2L	103	25	78	0.136390721	
		1L	154	39	115	0.139265449	

Baris TOTAL kolom Entropy pada Tabel 2 dihitung dengan rumus 2, sebagai berikut:

$$Entropy(Total) = \left(-\frac{83}{340} * \log_2 \left(\frac{83}{340} \right) \right) + \left(-\frac{257}{340} * \log_2 \left(\frac{257}{340} \right) \right)$$

$$Entropy(Total) = 0.80182099$$

Sementara itu, nilai Gain pada baris Kesehatan dihitung dengan menggunakan persamaan 1 sebagai berikut.

$$Gain(total, kesehatan) = Entropy(total) - \sum_{i=1}^n \frac{|Outlook_i|}{|Total|} * Entropy(kesehatan_i)$$

$$Gain(total, Kesehatan) = 0.80182099 - \left(\left(\frac{313}{340} * 1,7501 \right) + \left(\frac{17}{340} * 0.0161 \right) + \left(\frac{5}{340} * 0,143 \right) + \left(\frac{4}{340} * 0 \right) \right)$$

$$Gain(total, Kesehatan) = 0.02094332$$

$$Gain(total, MMPI) = 0.80182099 - \left(\left(\frac{54}{340} * 0.1392 \right) + \left(\frac{0}{340} * 0 \right) + \left(\frac{129}{340} * 0,2501 \right) + \left(\frac{97}{340} * 0.2509 \right) \right)$$

$$Gain(total, MMPI) = 0.16162721$$

$$Gain(total, Wawancara) = 0.80182099 - \left(\left(\frac{337}{340} * 0,251 \right) + \left(\frac{3}{340} * 0 \right) \right)$$

$$Gain(total, wawancara) = 0.00358091$$

$$Gain(total, Status) = 0.80182099 - \left(\left(\frac{0}{340} * 0 \right) + \left(\frac{340}{340} * 0.7982 \right) \right)$$

$$\text{Gain}(\text{total}, \text{Status}) = 0.0$$

$$\text{Gain}(\text{total}, \text{Ket Tulis}) = 0.80182099 - \left(\left(\frac{25}{340} * 0 \right) + \left(\frac{58}{340} * 0.1488 \right) + \left(\frac{103}{340} * 0.1363 \right) + \left(\frac{154}{340} * 0.1392 \right) \right)$$

$$\text{Gain}(\text{total}, \text{Ket Tulis}) = 0.37729306$$

C. Pengujian Model

Data pada tabel 1 diimplementasikan kedalam algoritma C4.5 melalui suatu model simulasi menggunakan RapidMiner. Sebanyak 90 % data akan digunakan untuk membangun struktur pohon keputusan melalui metode C4.5. Sedangkan 10 % lainnya digunakan sebagai data uji.

3. HASIL DAN PEMBAHASAN

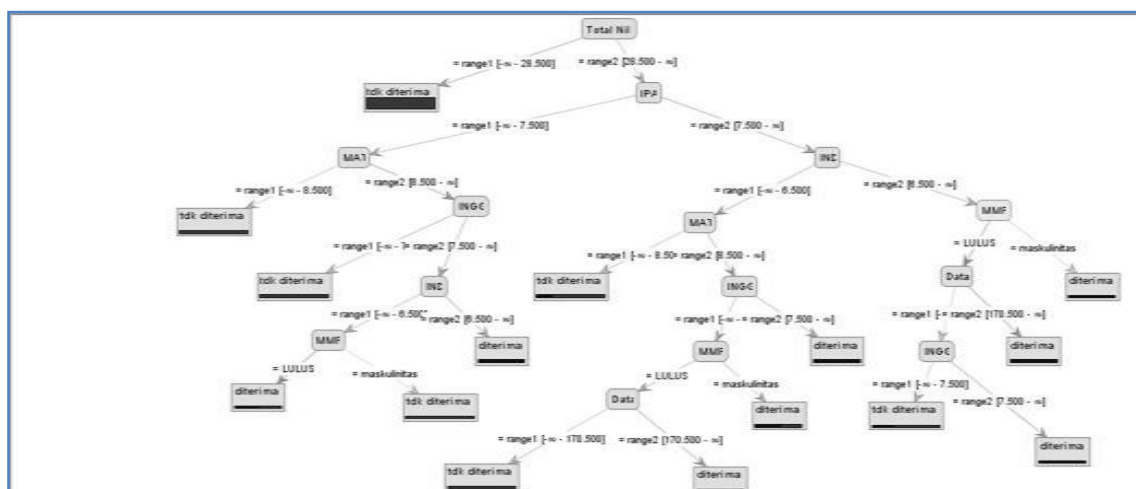
A. Hasil Pengujian Menggunakan Weight by Uncertainty

Dari hasil percobaan dan pengujian menggunakan simulasi rapidminer dengan model decision tree menggunakan weight by uncertainty didapatkan hasil training dan testing pada tabel dibawah ini.

Tabel 3. Hasil menggunakan menggunakan weight by uncertainty

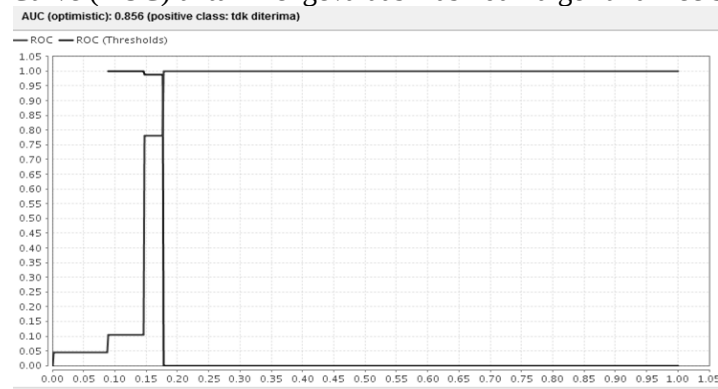
Attribute	weight
Wawancara	0
Status	0
Data	0.053
MMPI	0.057
Kesehatan	0.071
IND	0.346
INGG	0.359
MAT	0.397
IPA	0.436
Total Nilai	1

Hasil yang dibentuk oleh weight by uncertainty pada tiap-tiap atribut menghasilkan pohon keputusan dengan akurasi 80.39% ditampilkan pada gambar 1.



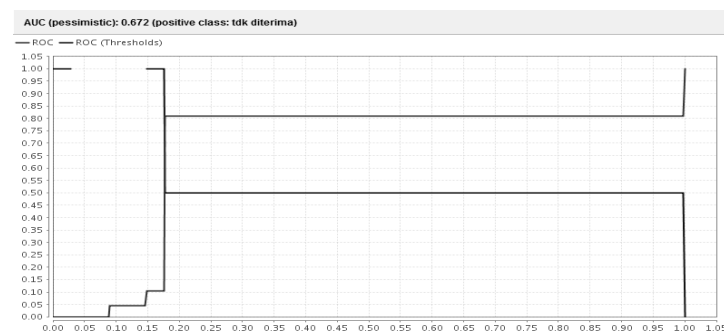
Gambar 1. Pohon keputusan dengan Weight by Uncertainty

Setelah pohon keputusan terbentuk pengukuran data dilakukan dengan confusion matrix dan ROC Curve (AUC) untuk mengevaluasi hasil dari algoritma Decision Tree C4.5.



Gambar 2. Kurva ROC Hasil Percobaan Menggunakan Weight by Uncertainty

disimpulkan bahwa Garis berwarna merah merupakan kurva ROC dengan nilai sebesar 0.856 termasuk nilai klasifikasi baik, sedangkan garis berwarna biru merupakan kurva ambang (thresholds).



Gambar 3. Kurva ROC (pessimistic) Positive class: tdk diterima

Gambar Kurva ROC pessimistic memperlihatkan hasil akurasi dari 0.672 yang berarti dapat termasuk dalam poor classification.

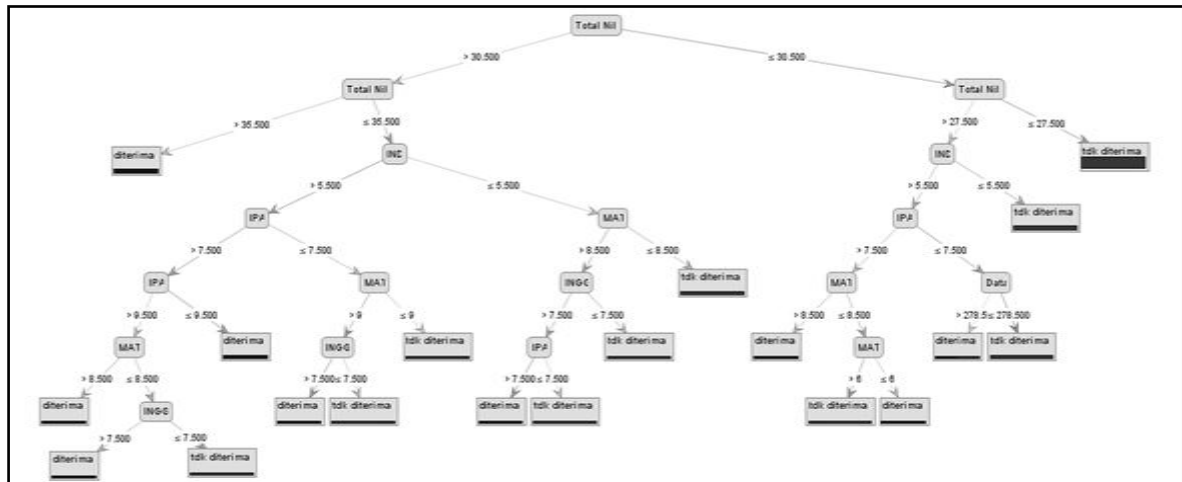
B. Hasil Percobaan Menggunakan Weight by Information Gain Ratio

Dengan data pada tabel 1 pengujian dilakukan menggunakan Weight by Information Gain Ratio didapatkan hasil training dan testing pada tabel 4.

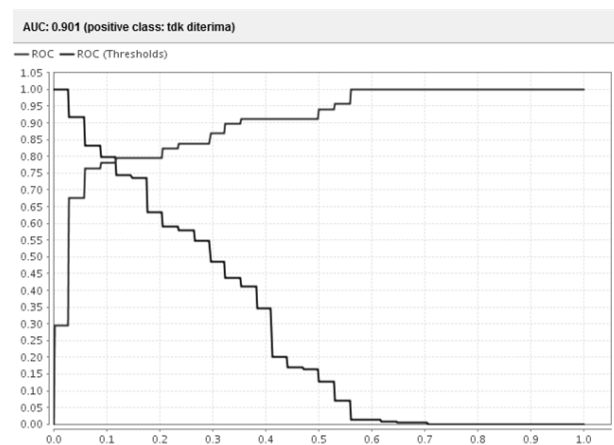
Tabel 4. Pengujian menggunakan Weight by Information Gain Ratio

Attribute	weight
Wawancara	0
Status	0
Data	0.513
MMPI	0.032
Kesehatan	0.053
IND	0.427
INGG	0.513
MAT	0.567
IPA	0.542
Total Nilai	1

Pengujian menggunakan Weight by Information Gain Ratio pada setiap atribut menghasilkan pohon keputusan dengan akurasi 88.24% sebagaimana terlihat pada gambar 4.

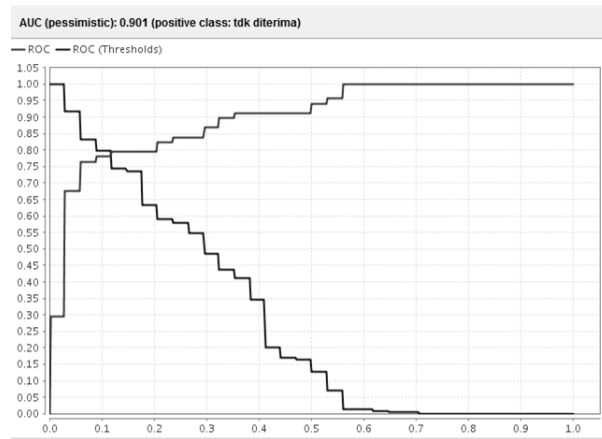


Gambar 4. Pohon keputusan dengan Weight by Information Gain Ratio



Gambar 5. AUC Positive Class

Pada gambar Kurva ROC Positive class memperlihatkan nilai 0.901 yang berarti dapat termasuk dalam excellent classification.



Gambar 6. AUC Pessimistic

Gambar 6 Kurva ROC pessimistic memperlihatkan hasil akurasi dari 0.901 yang berarti dapat termasuk dalam good classification.

Tabel 5 Hasil Perbandingan

	<i>Decision Tree dengan Uncertainty</i>	<i>Decision Tree dengan Information Gain Ratio</i>
<i>Akurasi</i>	80,39%	88,24%
<i>Precision</i>	94,44%	98,28%
<i>Recall</i>	75,00%	83,82%

Secara keseluruhan dalam kasus ini algoritma Decision tree kinerjanya lebih baik dengan menggunakan information Gain Ratio dibandingkan dengan kinerja uncertainty. Decision tree dengan information gain ratio mendapatkan akurasi yang lebih baik yaitu sebesar 88,24%, precision 98,28% dan recall 83,82%, decision tree dengan uncertainty mendapatkan akurasi sedikit lebih rendah dari metode sebelumnya yaitu 80,39%, precision 94,44%, recall 75,00%.

4. KESIMPULAN

Dari hasil penelitian dan pengukuran, dapat disimpulkan:

1. Pohon keputusan menggunakan Weight by Information Gain Ratio dapat digunakan pada data sejenis, karena telah memiliki akurasi 88,24%, demikian juga untuk pohon keputusan menggunakan weight by uncertainty yang sedikit lebih rendah akurasinya yang hanya memiliki 80,39% akurasi.
2. Algoritma decision tree C4.5 dapat digunakan untuk klasifikasi model prediksi calon mahasiswa baru dengan pembobotan.

5. SARAN

Berdasarkan hasil penelitian, penerapan sistem ini dapat membawa efek positif dalam proses prediksi calon mahasiswa baru, namun terdapat beberapa hal yang perlu penulis sarankan bagi pengembangan penelitian ini antara lain:

1. Ada baiknya pengujian data dilakukan dengan menambahkan parameter lainnya agar hasil prediksi lebih akurat.

2. Untuk mengestimasi akurasi sebuah algoritma akan lebih baik jika jumlah data sampel yang digunakan mendekati populasi yang ada.
3. Pada penelitian selanjutnya jumlah data set untuk Training data yang diperlukan sebaiknya meliputi data testing dan data training. Data yang digunakan juga lebih banyak agar pengklasifikasian data jauh lebih akurat

DAFTAR PUSTAKA

- [1] Zhiwu liu and Xiuzhi Zhang, "Prediction and Analysis for Students' Marks Based on Decision Tree Algorithm," in 2010 *Third International Conference on Intelligent Networks and Intelligent Systems*, 2010.
- [2] Daniel T Larose, *Discovering knowledge in data: An Introduction to Data Mining*. Canada: Wiley Interscience, 2005.
- [3] Oded Maimon Lior Rokach, *Data Mining and Knowledge Discovery Handbook*, 2010
- [4] Malaya Dutta Borah, Rajni Jindal, and Daya Gupta, "Application of knowledge based decision technique to Predict student enrollment decision," *2011 International Conference on Recent Trends in Information Systems IEEE*, 2011.
- [5] Adrian Costea and Tomas Eklund, "A Two-Level Approach to Making Class Predictions," *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03) IEEE*, , 2002.
- [6] Sotiris B. Kotsiantis & Panayiotis E. Pintelas, "Predicting Students' Marks in Hellenic Open University," *Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*, vol. 0-7695-2338-2/05, 2005.
- [7] Smith Tsang, Kao, Ben., "Decision Tree for Uncertain Data", IEEE Computer Society, 2009.
- [8] Pathom Pumouang and Anongnart Srivihok, "Comparisons of Classifier Algorithms: Bayesian Network, C4.5, Decision Forest and NBTree for Course Registration Planning Model of Undergraduate Students," *IEEE*, vol. 1-4244-2384-2/08/, 2008.
- [9] Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques Second Edition*. San Fransisco: Morgan Kaufmann, 2007.
- [10] Xindong Wu and Vipin Kumar, *The Top Ten Algorithms In Data Mining*. London: Chapman, 2009.