

KOMPARASI MODEL DECISION TREE UNTUK PREDIKSI KELULUSAN MAHASISWA DI UNIVERSITAS PANDANARAN

Abdul Rohman¹, Anief Ruffyanto²

email: *¹ abdulrohman15@gmail.com, ²aniefruffyanto@gmail.com

Abstrak

Dalam memprediksi kelulusan mahasiswa dalam perguruan tinggi dengan menggunakan metode klasifikasi data mining *Decision Tree*, memberikan manfaat dalam pengambilan keputusan. Mahasiswa di Universitas Pandanaran merupakan objek penelitian ini yang memiliki data mahasiswa kelas reguler dan mahasiswa kelas karyawan dan kebanyakan statusnya sudah bekerja.

Metode *data mining* yang digunakan adalah *Decision Tree C4.5*, *ID3* dan *CHAID* Sehingga menghasilkan nilai akurasi tinggi adalah *ID3* dengan nilai 73,19% sedangkan nilai *AUC* tertinggi adalah *C4.5* dengan nilai 0,874.

Kata Kunci:

Data Mining, *Decision Tree*, *C4.5*, *ID3*, *CHAID*

PENDAHULUAN

Data mahasiswa menjadi sangat penting untuk bahan mengambil suatu keputusan, jika data tersebut diolah dan analisa dengan menggunakan *data mining*, karena setiap kumpulan data dapat memberikan pengetahuan penting dan informasi yang sangat berharga bagi perguruan tinggi. Pada perguruan tinggi, suatu sistem informasi dapat digunakan untuk memperoleh informasi yang menunjang setiap pada pengambilan suatu keputusan

Data mahasiswa yang diolah dengan menggunakan *data mining* akan memberikan informasi yang sangat berharga untuk pengambilan keputusan terutama menggunakan algoritma atau metode *decision tree*, dimana metode ini menghasilkan pohon keputusan dan pola yang mudah untuk dipahami [1],[2].

Penelitian yang kaitannya dengan menerapkan *data mining* untuk memprediksi kelulusan mahasiswa dengan menggunakan metode *Decision Tree*, sudah banyak dilakukan dengan data mahasiswa reguler [3], [4], [5], [6], [7], Sedangkan dalam penelitian ini mahasiswa di Universitas Pandanaran terutama Diploma 3 Fakultas Teknik, memiliki data mahasiswa reguler dan karyawan dan kebanyakan statusnya sudah bekerja.

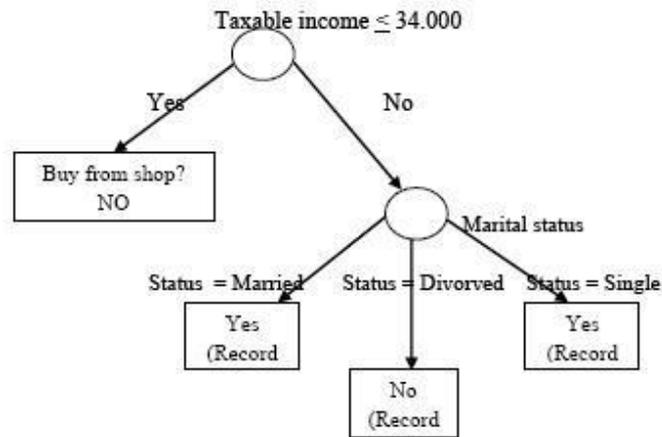
KAJIAN PUSTAKA

a. Klasifikasi *Data Mining*

Data mining merupakan proses mengolah dan menganalisis data untuk memecahkan suatu masalah dengan kegiatan penumpukan, pemakaian data histori sehingga memunculkan sebuah pola yang teratur [8],[9].

b. Metode *Decision Tree*

Metode *decision tree* merupakan metode yang mudah dimengerti, fleksible, dan menarik dalam membangun sebuah pohon keputusan dapat visualkan dalam bentuk gambar.[10].



Gambar 1. Contoh Konsep Pohon Keputusan Sederhana

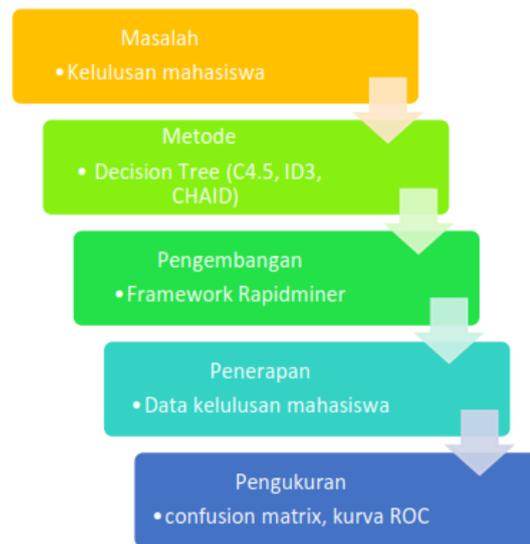
Ada 6 tahapan dalam membuat pohon keputusan dengan menggunakan algoritma atau metode *C4.5* yaitu; (1) mempersiapkan data training, (2) menentukan akar pohon dengan index entropy, (3) menghitung nilai gain, (4) menghitung split information, (5) menghitung rasio dan (6) mengulangi langkah ke-2 sampai semua recordnya terpatisi [10].

Algoritma atau metode *ID3* membentuk pohon keputusan secara ringkas dengan menghitung entropy ruang sampel dan entropy ruang sampel dan atribut [11].

Algoritma atau metode *CHAID* membentuk pohon atau diagram keputusan dengan melalui 3 tahap yaitu (1) penggabungan, (2) pemisahan dan (3) penghentian [12].

c. Kerangka Pemikiran

Kerangka pemikiran dalam penelitian ini, seperti gambar dibawah ini.



Gambar 2. Kerangka Pemikiran

METODE PENELITIAN

Data penelitian ini adalah data kelulusan mahasiswa di Universitas Pandanaran dengan jenjang pendidikan D3 di Fakultas Teknik berjumlah 235 mahasiswa yang terdiri dari 151 mahasiswa yang lulus tepat waktu dan 84 mahasiswa yang terlambat.

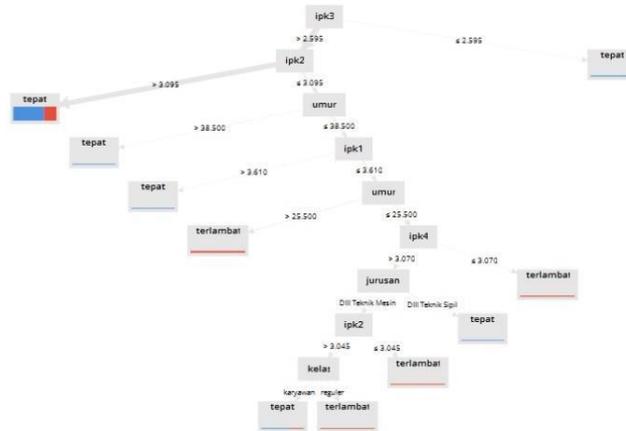
Tabel 1. Dataset Kelulusan Mahasiswa

jurusan	umur	jk	pekerjaan	ipk1	ipk2	ipk3	ipk4	label
DIII Teknik Sipil	23	Laki-laki	bekerja	3.62	3.76	3.67	3.5	tepat
DIII Teknik Sipil	28	Laki-laki	bekerja	3.29	3.19	3.05	3.2	terlambat
DIII Teknik Sipil	29	Laki-laki	bekerja	3.29	3.14	2.76	3.5	terlambat
DIII Teknik Sipil	35	Laki-laki	bekerja	3.19	3.1	2.9	3.1	terlambat
DIII Teknik Sipil	35	Laki-laki	bekerja	3.62	3.52	3.62	3.3	terlambat
DIII Teknik Sipil	27	Laki-laki	bekerja	2.86	2.86	3	3.1	terlambat
DIII Teknik Sipil	23	Laki-laki	bekerja	3.05	3.19	3.29	3.4	terlambat
DIII Teknik Sipil	33	Perempuan	bekerja	3.29	3.05	3.48	3.3	terlambat
DIII Teknik Sipil	21	Laki-laki	bekerja	3.48	3.14	3.38	3.3	terlambat
DIII Teknik Sipil	22	Laki-laki	bekerja	3.23	3.24	3.29	3.1	terlambat
DIII Teknik Sipil	25	Perempuan	bekerja	3.57	3.9	3.17	3.4	tepat
DIII Teknik Sipil	39	Laki-laki	bekerja	3.67	3.57	3.48	3	tepat
DIII Teknik Sipil	27	Laki-laki	bekerja	3.48	3.67	3.48	3	tepat
DIII Teknik Sipil	33	Laki-laki	bekerja	3.48	3.1	3.05	3.5	tepat
DIII Teknik Sipil	34	Laki-laki	bekerja	3.24	2.86	2.57	2.8	tepat
DIII Teknik Sipil	23	Laki-laki	bekerja	2.9	3	3	3	terlambat
DIII Teknik Sipil	30	Laki-laki	bekerja	3.19	3.24	3.57	3.1	tepat
DIII Teknik Sipil	31	Laki-laki	bekerja	3.48	3.14	2.86	2.9	tepat
DIII Teknik Sipil	28	Laki-laki	bekerja	3.43	3.57	2.76	2.8	tepat
DIII Teknik Sipil	24	Laki-laki	bekerja	3.29	3.29	2.57	2.7	tepat
DIII Teknik Sipil	43	Laki-laki	bekerja	3.38	3.24	2.48	2.4	tepat
DIII Teknik Sipil	25	Laki-laki	bekerja	3.38	3.57	2.38	2.4	tepat
DIII Teknik Sipil	30	Laki-laki	bekerja	3.19	3.24	3.57	3.1	tepat
DIII Teknik Sipil	30	Laki-laki	bekerja	3.05	3.43	3.19	3.5	terlambat
DIII Teknik Sipil	26	Laki-laki	bekerja	2.86	2.95	2.62	3.3	terlambat
DIII Teknik Sipil	27	Laki-laki	bekerja	3.05	3.35	3	3.2	terlambat
DIII Teknik Sipil	26	Laki-laki	bekerja	2.9	3.26	2.84	3.2	terlambat
DIII Teknik Sipil	28	Laki-laki	bekerja	3.29	3.26	2.84	3.3	terlambat

HASIL DAN PEMBAHASAN

a. Algoritma C4.5

Pengolahan data kelulusan mahasiswa dengan menggunakan C4.5 menghasilkan model pohon keputusan serta menghasilkan 10 rule/pola sebagai berikut:

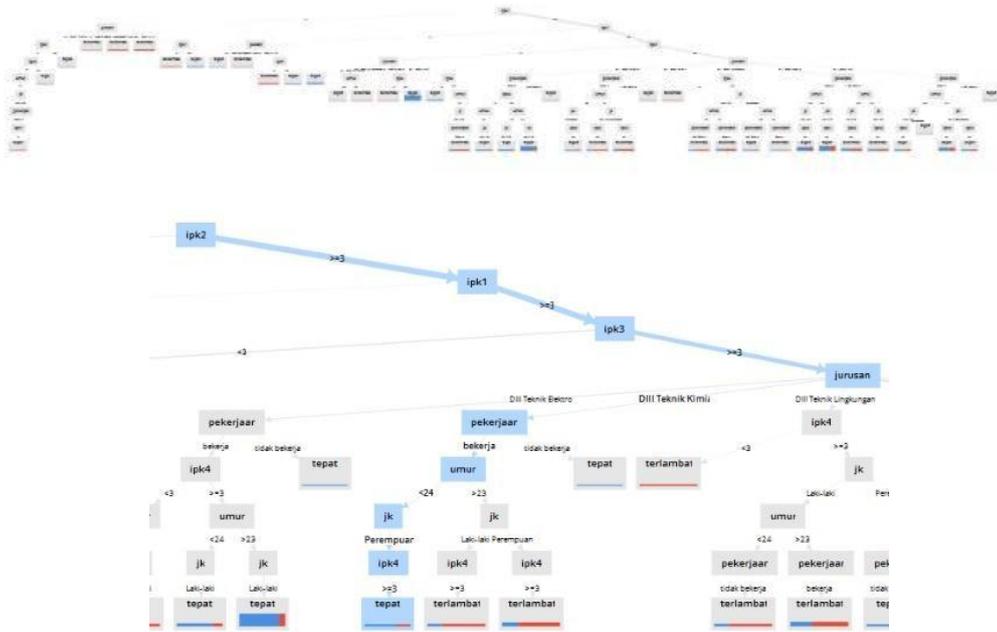


Gambar 3. Model Pohon keputusan C4.5

Setelah menghasilkan diagram dan 10 rule/pola tersebut diatas, selanjutnya mengetahui nilai akurasi yaitu 65,5% dan nilai AUC (Area Under Curve) sebesar 0,874 sehingga dapat dikategorikan klasifikasi Baik.

b. Algoritma ID3

Pengolahan data kelulusan mahasiswa dengan menggunakan ID3 menghasilkan model pohon keputusan serta menghasilkan 38 rule/pola sebagai berikut:

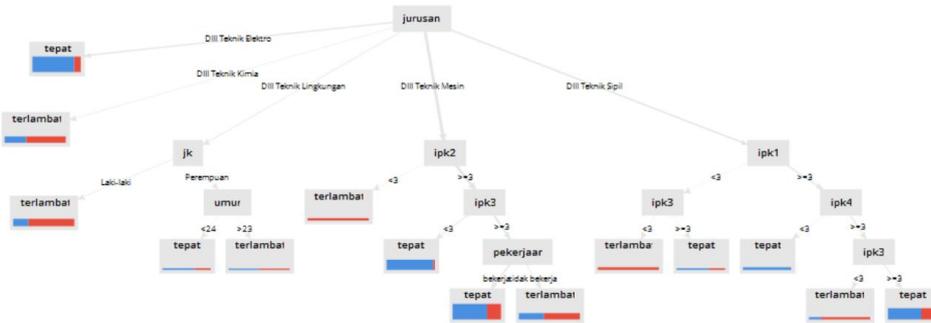


Gambar 4. Model Pohon keputusan ID3

Setelah menghasilkan diagram dan 38 rule/pola tersebut diatas, selanjutnya mengetahui nilai akurasinya yaitu 73,19% dan nilai *AUC* (*Area Under Curve*) sebesar 0,806 sehingga dapat dikategorikan klasifikasi Baik.

c. Algoritma CHAID

Pengolahan data kelulusan mahasiswa dengan menggunakan CHAID menghasilkan model pohon keputusan serta menghasilkan 13 rule/pola sebagai berikut:



Gambar 5. Model Pohon keputusan CHAID

Setelah menghasilkan diagram dan 13 rule/pola tersebut diatas, selanjutnya mengetahui nilai akurasinya yaitu 71,87% dan nilai *AUC* (*Area Under Curve*) sebesar 0,785 sehingga dapat dikategorikan klasifikasi Cukup.

Dari ketiga metode *Decision Tree* diatas memiliki tingkat akurasi dan *AUC* yang berbeda-beda, yaitu dengan perbandingan sebagai berikut:

Tabel 2. *Perfomance Decision Tree*

No	Decisio Tree	Perfomance	
		Akurasi	AUC
1	C4.5	65,98%	0,874
2	ID3	73,19%	0,806
3	CHAID	71,87%.	0.785

Dari hasil tabel diatas, maka dari ketiga metode *Decision Tree* yang memiliki akurasi tinggi adalah *ID3* dengan nilai 73,19% sedangkan nilai *AUC* tertinggi adalah *C4.5* dengan nilai 0,874.

KESIMPULAN

Dalam penelitian ini dilakukan pengolahan data kelulusan mahasiswa Universitas Pandanaran dengan menggunakan metode *Decision Tree C4.5*, *ID3* dan *CHAID*, dan menghasilkan model dan beberapa rule/pola. Hasil yang didapat dari ketiga metode *Decision Tree* yang memiliki akurasi tinggi adalah *ID3* dengan nilai 73,19% sedangkan nilai *AUC* tertinggi dalah *C4.5* dengan nilai 0,874

REFERENSI

- [1] Asriningtias 2014 “Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa,” *J. Inform.*, vol. 8, no.1, pp. 837–848
- [2] Ridwan, 2013, Suyono, Sarosa, “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier,” *Eccis*, vol. 7, no. 1, pp. 59–64.
- [3] Himawan 2011, “Aplikasi Data Mining Menggunakan Algoritma ID3 Untuk Mengklasifikasi Kelulusan Mahasiswa Pada Universitas Dian Nuswantoro Semarang,” *Fak. Ilmu Komputer*.
- [4] Muhammad Sholeh, 2014, “Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014 Yogyakarta, 15 November 2014 ISSN: 1979-911X,” *Snast*, no. November, pp. 211–216.
- [5] Romadhona, Agus; suprapedi; himawan 2017, “Prediksi Kelulusan Tepat Waktu Mahasiswa Stmik- Ymi,” *J. Teknol. Inf.*, vol. 13, no. 1, pp. 69–83.
- [6] Marselina Silvia Suhartinah, 2010, “Graduation Prediction Of Gunadarma University Students Using Algorithm And Naive Bayes C4.5 Algorithm,” *Fac. Ind. Technol. Gunadarma Univ.*
- [7] Siska Haryati, Aji Sudarsono, 2015, “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu),” *J. Media Infotama*, vol. 11, no. 2, pp. 130–138.
- [8] Witten 2007, “Data Mining Data Mining Complications: Overfitting Statistical modeling One attribute does all the work,”.
- [9] Budi Santosa, 2007, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- [10] Gorunescu, 2011, *Data Mining: Concepts, Models and Techniques (Intelligent Systems Reference Library)*.
- [11] Munthe, Sihombing, 2018, “Klasifikasi Algoritma Iterative Dichotomizer (ID3) untuk Tingkat kepuasan pada Sarana Laboratorium Komputer,” *Jutikomp*, vol. 1, no. 2, pp. 27–34.
- [12] Yustisia Wirania 2013, *Pembentukan Pohon Klasifikasi dengan Metode Chaid*. Pontianak: UNTAN.