

ALGORITMA DECISION TREE C.45 DALAM ANALISA KELULUSAN MAHASISWA PROGRAM STUDI MANAJEMEN INFORMATIKA UMPP

Aslam Fatkhudin¹, M. Yusuf Febrianto², Fenilinas Adi Artanto³,
M. Waffa Najib Hadinata⁴, Riza Fahlevi⁵

^{1,3,5} Informatika, Universitas Muhammadiyah Pekajangan Pekalongan, Pekalongan, Indonesia

^{2,4} Manajemen Informatika, Universitas Muhammadiyah Pekajangan Pekalongan, Pekalongan, Indonesia

¹ aslamfatkhudin@umpp.ac.id, ² m.yusuf.febrianto@umpp.ac.id, ³ fenilinasadi@gmail.com,

⁴ mahasiswa.fastikom@umpp.ac.id, ⁵ mahasiswa.fastikom@umpp.ac.id

ABSTRAK

Sebuah Program Studi pada perguruan tinggi dapat dinilai dari bagaimana ketepatan lulusan mahasiswanya. Apakah banyak mahasiswa yang dapat menyelesaikan pendidikan tepat pada waktunya atau malah terdapat mahasiswa yang terlambat bahkan sampai tidak dapat menyelesaikan pendidikannya pada program studi tersebut. Oleh karena hal tersebut maka dilakukanlah penelitian agar dapat diketahui penyebab kelulusan mahasiswa pada Program Studi Diploma Tiga Manajemen Informatika. Dikarenakan dengan meningkatkan kelulusan mahasiswa dengan tepat waktu juga akan berpengaruh terhadap akreditasi Progra Studi tersebut. Pada analisa kelulusan mahasiswa digunakan algoritma *decision tree C4.5* dikarenakan algoritma tersebut cukup populer, selain mudah untuk digunakan algoritma tersebut juga mempunyai tampilan estimasi berupa pohon keputusan yang mudah untuk dipahami. Dari analisa yang dilakukan didapatkan bahwa tahun masuk mahasiswa menjadi faktor variabel utama dalam kelulusan mahasiswa dengan didapatkan akurasi algoritma *decision tree* sebesar 73,48%. Dengan Tahun mahasiswa menjadi point utama dikarenakan kondisi pemenuhan SDMDan fasilitas dalam menangani mahasiswa untuk segera menyelesaikan Tugas Akhirnya. Dengan adanya hasil analisis ini diharapkan Program Studi Diploma Tiga Manajemen Informatika Universitas Muhammadiyah Pekajangan Pekalongan dapat meningkatkan ketepatan waktu untuk kelulusan mahasiswanya.

Kata Kunci— *Decision Tree, Data Mining, Kelulusan Mahasiswa.*

ABSTRACT

A study program at a university can be judged by how accurate the graduates are. Are there many students who can complete their education on time or are there students who are late and even unable to complete their education in the study program. Because of this, research was carried out in order to find out the cause of student graduation in the Manajemen Informatika Diploma Three Study Program. Because increasing student graduation on time will also affect the accreditation of the Study Program. In the analysis of student graduation, the *decision tree C4.5* algorithm is used because the algorithm is quite popular, besides being easy to use, the algorithm also has an estimation display in the form of a decision tree that is easy to understand. From the analysis carried out, it was found that the year of student entry became the main variable factor in student graduation with the accuracy of the *decision tree* algorithm being 73.48%. With the results of this analysis, it is hoped that the Diploma Three Study Program of Manajemen Informatika at the Universitas Muhammadiyah Pekajangan Pekalongan can improve the timeliness of graduation for its students.

Keywords— *Decision Tree, Data Mining, Graduation.*

1. PENDAHULUAN

Sebuah Program Studi untuk menjaga kualitas dan meendapatkan nilai akreditasi yang baik harus dapat menjaga tingkat kelulusan mahasiswanya. Banyak tahapan yang harus dilalui mahasiswa dalam menyelesaikan proses studi sampai akhirnya dapat lulus seperti mengikuti mata kuliah wajib, melaksanakan Praktek Kerja Lapangan (PKL) dan menyelesaikan Tugas Akhir (TA) sesuai dengan proses yang telah ditentukan oleh universitas [1].

Dengan banyaknya mahasiswa yang telah bergabung ke dalam program studi untuk meningkatkan kualitas akreditasi maka diperlukan sebuah analisa tentang kelulusan mahasiswa. Dengan banyaknya data yang ada maka proses analisa perlu menggunakan metode *data mining*. *Data Mining* merupakan metode ataupun proses dalam mencari sebuah informasi yang didapatkan dari menambang sekumpulan data yang berjumlah banyak, sehingga informasi yang didapatkan dapat berguna [2]. Untuk menganalisis data yang berjumlah banyak dengan metode *data mining* digunakanlah metode klasifikasi

dengan algoritma *decesion tree*. *Decesion Tree* merupakan metode klasifikasi *data mining* yang cukup poluler dan banyak digunakan [3]. Dengan *Decesion Tree* dibentuk sebuah pohon keputusan dengan atribut data yang akan menghasilkan atura-aturan dalam mengklasifikasi [4].

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data Mining merupakan sebuah proses dalam menambang data dari sekumpulan data yang banyak untuk mendapatkan informasi yang berguna dari kumpula data yang besar tersebut [5]. *Data Mining* adalah bagian dari *knowledge discover in databases* (KDD). KDD terdiri dari beberapa proses yaitu pembersihan data, integrasi data, pemilihan data, transformasi data, *data mining*, evaluasi pola, dan penyajian pengetahuan dari data [6]. *Data mining* sendiri dibagi dalam beberapa kelompok, yaitu deskripsi, estimasi, prediksi, klasifikasi, clustering, dan asosiasi [7].

2.2. Klasifikasi

Klasifikasi merupakan bagian dari *data mining*. Pada awalnya klasifikasi digunakan pada tanaman untuk mengklasifikasikan tanaman berdasarkan jenis spesiesnya sesuai dengan karakteristik fisik pada tanaman [7]. Klasifikasi memetakan data dalam kelompok ataupun kelas-kelas tertentu. Klasifikasi juga bagian dari metode *supervised learning* yang menghasilkan aturan dalam mengklasifikasikan data yang di uji kedalam kelompok ataupun kelas yang telah ditentukan [8]. Beberapa teknik atau metode dalam klasifikasi adalah *decesion tree support machine*, *neural-network*, *naive bayes* dan *ruled-based clasisifier*.

2.3. Pohon Keputusan (Decision Tree)

Pohon Keputusan yang biasa disebut *Decesion Tree* adalah metode dari klasifikasi pada *data mining*. Metode ini banyak digunakan dikarenakan mudah dalam menginterpretasikan hasil dari estimasi metode tersebut karena berbentuk pohon keputusan [1]. Pohon Keputusan merupakan struktur yang digunakan untuk membagi kelompok dari data dalam jumlah besar menjadi himpunan-himpunan *record* yang menerapkan aturan keputasn [3].

Pohon Keputusan memiliki node yang mendisripsikan atribut pada setiap cabang yang menggambarkan hasil dari atribut yang menggambarkan kelas [4]. Secara umumnya Pohon Keputusan membangun bentuk klasifikasi dengan urutan berikut [7]:

- Pilih atribut sebagai akar
- Buatlah cabang pada setiap nilai
- Membagi kasus pada tiap cabang
- Ulangi setiap proses hingga seluruh kasus mendapatkan kelas

Dalam memilih atribut akar berdasarkan pada nilai *gain* yang dihitung menggunakan rumus Sebagai berikut [9]:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy S_i$$

Dimana:

- S = Himpunan kasus
A = Atribut
N = Jumlah Atribut A
|Si| = Jumlah kasus pada partisi ke-i
|S| = Jumlah kasus dalam S

3. METODE YANG DIUSULKAN

4.1. Data

Data yang digunakan didapatkan dari mahasiswa Manajemen Informatika Universitas Muhammadiyah Pekajangan Pekalongan/

4.2. Variabel

Data yang akan diambil berdasarkan dari variabel berikut ini:

Tabel 1 Variabel dan indikator penilaian

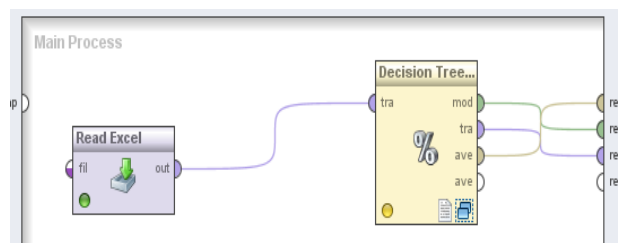
Variabel	Indikator	Skala Pengukuran
X1	Jenis Kelamin	Laki-laki dan Perempuan
X2	IPK	0-4
X3	Tahun Masuk	Tahun
X4	Tahun Lulus	Tahun
X5	Jumlah Semester	Semester
Y	Kelulusan	Tepat waktu, Terlambat dan Sangat Terlambat

4. HASIL PENELITIAN

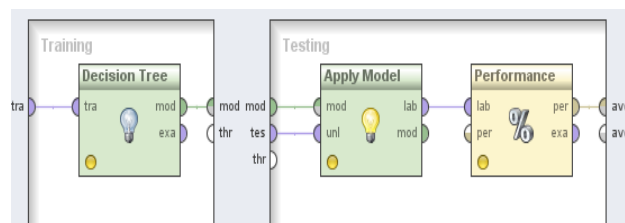
4.1. Analisa

Data yang sudah didapatkan dibangun model dengan menggunakan program Rapidminer dengan tampilan model sebagai berikut:

berikut:



Gambar 1 Main Proses Rapidminer dengan crossvalidation



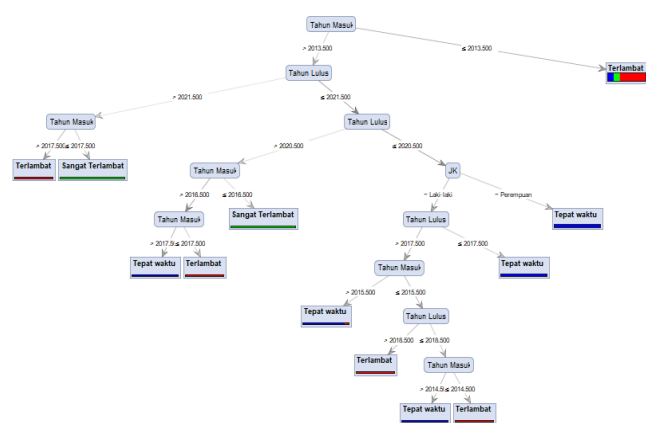
Gambar 2 Training dan Testing Rapidminer

Untuk pengaturan pada rapid miner dapat menggunakan pengaturan berikut ini:

1. Pada data digunakan Y sebagai label
2. Dipilih *Decesion Tree X-Validation*
3. Isi number of value dengan nilai 10
4. Pada data training diberikan nilai *minimaml size 4, minimal leaf size 2, minimal gain 0,1, maximal dept 20, confidence 0,2 dan nilai number of prepruning 3*.

4.2. Klasifikasi *decision tree*

Dari 114 data yang didapatkan dan di analsis dengan menggunakan rapidminer didapatkan hasil pohon keputusan sebagai berikut:



Gambar 3 Pohon Keputusan

Dari gambar pohon keutusan terlihat jalur utamanya adalah variabel Tahun Masuk (X3). Selain itu juga diddapatkan bentuk *text* sebagai berikut:

```

Tahun Masuk > 2013.500
| Tahun Lulus > 2021.500
| | Tahun Masuk > 2017.500:
| | | Terlambat {Tepat waktu=0, Sangat
| | | Terlambat=0, Terlambat=3}
| | | Tahun Masuk <= 2017.500:
| | | Sangat Terlambat {Tepat waktu=0,
| | | Sangat Terlambat=2, Terlambat=0}
| | Tahun Lulus <= 2021.500
| | | Tahun Lulus > 2020.500
| | | | Tahun Masuk > 2016.500
| | | | | Tahun Masuk >
| | | | | 2017.500: Tepat waktu {Tepat
| | | | | waktu=6, Sangat Terlambat=0,
| | | | | Terlambat=0}
| | | | | Tahun Masuk <=
| | | | | 2017.500: Terlambat {Tepat waktu=0,
| | | | | Sangat Terlambat=0, Terlambat=3}
| | | | | Tahun Masuk <= 2016.500:
| | | | | Sangat Terlambat {Tepat waktu=0,
| | | | | Sangat Terlambat=3, Terlambat=0}
| | | Tahun Lulus <= 2020.500
| | | | JK = Laki-laki
| | | | | Tahun Lulus >
| | | | | 2017.500
    
```

```

| | | | | Tahun Masuk >
| | | | | 2015.500: Tepat waktu {Tepat
| | | | | waktu=10, Sangat Terlambat=0,
| | | | | Terlambat=1}
| | | | | Tahun Masuk <=
| | | | | 2015.500
| | | | | Tahun Lulus
| | | | | > 2018.500: Terlambat {Tepat
| | | | | waktu=0, Sangat Terlambat=0,
| | | | | Terlambat=8}
| | | | | Tahun Lulus
| | | | | <= 2018.500
| | | | | Tahun
| | | | | Masuk > 2014.500: Tepat waktu {Tepat
| | | | | waktu=3, Sangat Terlambat=0,
| | | | | Terlambat=0}
| | | | | Tahun
| | | | | Masuk <= 2014.500: Terlambat {Tepat
| | | | | waktu=0, Sangat Terlambat=0,
| | | | | Terlambat=2}
| | | | | Tahun Lulus <=
| | | | | 2017.500: Tepat waktu {Tepat
| | | | | waktu=14, Sangat Terlambat=0,
| | | | | Terlambat=0}
| | | | | JK = Perempuan: Tepat
| | | | | waktu {Tepat waktu=28, Sangat
| | | | | Terlambat=0, Terlambat=0}
| | | | | Tahun Masuk <= 2013.500: Terlambat
| | | | | {Tepat waktu=23, Sangat
| | | | | Terlambat=23, Terlambat=101}
    
```

Estimasi menggunakan algoritma *decesion tree C4.5* mendapatkan hasil akurasi sebesar 73,48% dengan rincian sebagai berikut:

Tabel 2 Akurasi C4.5

Accuracy: 73,48% +/- 7,89% (mikro: 73,48%)				
	true Tepat waktu	true Sangat Terlambat	True Terlambat	class precision
pred. Tepat Waktu	58	2	7	86,57%
pred. Sangat Terlambat	0	1	1	50,00%
pred. Terlambat	26	25	110	68,32%
class recall	69,05%	3,57%	93,22%	

4.3. Pembahasan

Hasil dari estimasi menggunakan *decision tree* melihat bahwa kualitas website menjadi faktor utama dalam pendaftaran siswa baru, untuk lebih jelasnya bisa dilihat dalam penjelasan berikut ini:

Jika Tahun Masuk > 2013.500

Tahun Lulus > 2021.500

Tahun Masuk > 2017.500: maka Terlambat

Tahun Masuk <= 2017.500: maka Sangat Terlambat

Tahun Lulus $\leq$ 2021.500
Tahun Lulus $>$ 2020.500
Tahun Masuk $>$ 2016.500
Tahun Masuk $>$ 2017.500: maka Tepat waktu
Tahun Masuk $\leq$ 2017.500: maka Terlambat
Tahun Masuk $\leq$ 2016.500: maka Sangat Terlambat
Tahun Lulus $\leq$ 2020.500
Jenis Kelamin = Laki-laki
Tahun Lulus $>$ 2017.500
Tahun Masuk $>$ 2015.500: maka Tepat waktu
Tahun Masuk $\leq$ 2015.500
Tahun Lulus $>$ 2018.500: maka Terlambat
Tahun Lulus $\leq$ 2018.500
Tahun Masuk $>$ 2014.500: maka Tepat waktu
Tahun Masuk $\leq$ 2014.500: maka Terlambat
Tahun Lulus $\leq$ 2017.500: maka Tepat waktu
Jenis Kelamin = Perempuan: maka Tepat waktu
Tahun Masuk $\leq$ 2013.500: maka Terlambat

Dengan nilai akurasi estimasi *decision tree* sebesar 73,48 %.

5. KESIMPULAN

Dari data yang telah didapatkan yang berupa Jenis Kelamin, IPK, Tahun Masuk dan Jumlah Semester yang di estimasi didapatkan bahwa tahun masuk menjadi faktor utama dalam menjadi alasan dalam kelulusan mahasiswa. Dengan tingkat akurasi estimasi *decision tree* sebesar 73,48%. Dengan hasil estimasi tersebut faktor utama dalam kelulusan mahasiswa adalah tahun masuk tahun lulus dan jenis kelamin. Tahun masuklah yang mejadi faktor utama dikarenakan dengan fasilitas yang dipunyai oleh Perguruan Tinggi.

Penelitian ini belum maksimal dikarenakan variabel faktor penunjang kelulusan mahasiswa belum banyak dan penggunaan metode klasifikasi hanya sebatas *decesion tree* saja. Kiranya dapat kelak bisa menggunakan metode lainnya yang juga terdapat penambahan jumlah data dan jumlah variabelnya.

Daftar Pustaka

- [1] C. N. Dengen, K. Kusriani, and E. T. Luthfi, "Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *Sisfotenika*, vol. 10, no. 1, p. 1, 2020, doi: 10.30700/jst.v10i1.484.
- [2] A. Fatkhudin, A. Khambali, and F. A. Artanto, "Decision Tree Dalam Mengklasifikasi Mata Kuliah Terhadap Pemahaman Sistem Pemasaran," *J. Ilm. Ilmu Komput. Fak. ...*, vol. 7, no. 2, pp. 52–55, 2021, [Online]. Available: <http://ejournal.fikom-unasman.ac.id/index.php/jikom/article/view/204>.
- [3] F. D. Marleny and Mambang, "Prediksi Calon Mahasiswa Baru Menggunakan Metode Klasifikasi Decision Tree," *CSRID J.*, vol. 7, no. 1, pp. 46–54, 2015.
- [4] A. H. Nasrullah, "Penerapan Metode C4.5 untuk

Klasifikasi Mahasiswa Berpotensi Drop Out," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 244–250, 2018, doi: 10.33096/ilkom.v10i2.300.244-250.

- [5] S. Santosa and F. A. Artanto, "PREDIKSI LOYALITAS PELANGGAN TELEKOMUNIKASI MENGGUNAKAN LOGISTIC REGRESSION DENGAN SELEKSI FITUR PARTICLE SWARM OPTIMIZATIONssion Dengan Seleksi Fitur Particle Swarm Optimization," vol. 11, no. April, pp. 90–99, 2015.
- [6] M. Ayub, "Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer," no. May, pp. 21–30, 2018.
- [7] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [8] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [9] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017, doi: 10.15575/join.v2i1.71.