

Penerapan Metode Decision Tree Dalam Menentukan Kelulusan Mahasiswa

Fitria Rahmadayanti, Inda Anggraini*

Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Pagaralam, Pagar Alam, Indonesia

Email: ¹ria.ria.rr71@gmail.com, ^{2,*}indaanggraini@gmail.com

Email Penulis Korespondensi: indaanggraini@gmail.com

Submitted: 22/12/2021; Accepted: 29/12/2021; Published: 31/12/2021

Abstrak—Tujuan dari penelitian ini adalah menghasilkan sistem prediksi penentuan kelulusan mahasiswa tepat waktu dengan metode Decision Tree pada Sekolah Tinggi Teknologi Pagaralam. Apabila mahasiswa banyak yang lulus tidak tepat waktu atau melebihi batas yang ditentukan akan mengakibatkan menumpuknya mahasiswa dalam jumlah besar dikarenakan tidak seimbangnya jumlah mahasiswa yang masuk dan keluar disetiap periode kelulusan sehingga dapat mengakibatkan proses akademik tidak berjalan dengan maksimal. Decision Tree merupakan algoritma klasifikasi yang dapat melakukan prediksi data dalam jumlah besar. Metode pengembangan yang digunakan yaitu metode Rapid Application Development (RAD) yang terdiri dari Requirement Planning (Perencanaan Persyaratan), Desain Workshop, Implementation (Penerapan). Penelitian ini dapat membantu pihak Sekolah Tinggi Teknologi Pagaralam dalam melihat apakah mahasiswa akan lulus tepat waktu atau tidak.

Kata Kunci: Sistem Prediksi; Kelulusan Mahasiswa; Decision Tree; RAD

Abstract—The purpose of this study is to produce a prediction system for determining the determination of student graduation on time with the Decision Tree method at Pagaralam High School of Technology. If many students graduate not on time or exceed the specified limit will result in the accumulation of students in large numbers due to the imbalance of the number of students entering and exiting each graduation period so that it can cause the academic process does not run optimally. Decision Tree is a classification algorithm that can predict large amounts of data. The development method used is the Rapid Application Development (RAD) method consisting of Requirement Planning (Requirements Planning), Workshop Design, Implementation (Implementation). This research can help the Pagaralam High School of Technology in seeing whether students will graduate on time or not.

Keywords: Prediction System; Student Graduation; Decision Tree; RAD

1. PENDAHULUAN

Perguruan Tinggi merupakan penyelenggara pendidikan akademik bagi mahasiswa salah satunya faktor untuk mendukung peningkatan mutu yaitu ketuntasan mahasiswa dalam proses belajar dengan tepat waktu. Hal ini menjadi tolak ukur untuk keberhasilan mahasiswa dan tingginya tingkat ketidakpastian dalam menyelesaikan studi menjadi suatu masalah[1]. Mahasiswa merupakan salah satu parameter penting dalam sebuah evaluasi program studi[2] Penyebab mahasiswa tidak lulus tepat waktu adalah kurangnya pencapaian atau perolehan angka kredit semester dari tahun ke tahun, kurang aktifnya dalam lingkungan kampus, kurangnya tingkat kedisiplinan dan ketidaktahuan dalam perhitungan sks.

Prediksi merupakan cara atau proses untuk memprediksi atau memperkirakan secara urut dan sistematis mengenai sesuatu yang mungkin dapat terjadi pada masa depan berdasarkan tentang informasi pada masa lalu dan sekarang yang dimiliki, agar tingkat *error* dan kesalahannya dapat diperkecil. Dalam prediksi tidak harus memberikan suatu jawaban secara pasti tentang kejadian yang nanti akan terjadi pada masa yang akan datang, melainkan berusaha untuk mencari jawaban yang akurat mungkin nanti akan terjadi[3].

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar[4]. Menurut Larose, data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat di lakukan, yaitu [4]: (a) Deskripsi Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. (b)Estimasi Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. (c)Prediksi Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. (d) Klasifikasi Dalam klasifikasi, terdapat target variabel kategori. (e) Pengklusteran Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain. Clustering merupakan salah satu metode data mining yang bersifat tanpa arahan (unsupervised). (f) Asosiasi Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja[4][5].

Kelulusan mahasiswa adalah mahasiswa dapat menyelesaikan program Pendidikan sesuai dengan kecepatan belajar masing-masing dan tidak melebihi ketentuan batas waktu yang ditetapkan oleh Perguruan Tinggi[6]. Decision Tree adalah salah satu metode yang ada pada teknik klasifikasi dalam data mining. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target[7].

Algoritma C4.5 adalah salah satu algoritma yang digunakan dalam *decision tree*. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser 3*) yang dikembangkan oleh J.Ross Quinlan. Tahapan algoritma C4.5 dalam membangun *decision tree* adalah (1) Pilih atribut sebagai akar; (2) Buat cabang untuk tiap-tiap nilai; (3) Bagi kasus dalam cabang; (4) Ulangi proses untuk setiap proses cabang sampai semua kasus pada cabang memiliki kelas yang sama[8].

Penelitian dilakukan merujuk pada penelitian terdahulu yakni, menghasilkan sistem prediksi kelulusan mahasiswa tepat waktu yang digunakan untuk membantu fakultas mengetahui hasil prediksi kelulusan mahasiswa dengan mudah dan tepat menggunakan Algoritma *Decision Tree* ID3 sehingga menghasilkan model dan *rule*. Dan dari hasil model kelulusan mahasiswa tersebut dilakukan evaluasi algoritma sehingga menghasilkan nilai akurasi 73,19 % dengan nilai AUC 0,806 % dan termasuk klasifikasi yang baik[9].

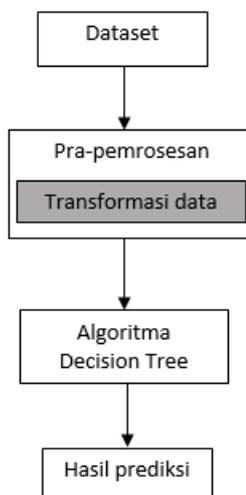
Penelitian sebelumnya kedua Menghasilkan sistem prediksi kelulusan mahasiswa tepat waktu dalam penelitian ini mahasiswa yang masuk di perguruan tinggi STMIK Adiguna di setiap tahunnya sangat banyak sedangkan yang lulus jauh lebih kecil sehingga menyebabkan proses belajar mengajar tidak efektif. Sistem ini diimplementasikan menggunakan metode *Decision Tree* yang memiliki kompleksitas yang lebih besar[10].

Penelitian sebelumnya ketiga menghasilkan sistem prediksi kelulusan mahasiswa tepat waktu untuk mengetahui lebih dini bahwa mahasiswa tersebut akan lulus tepat waktu atau tidak dengan mudah dan tepat. Sistem ini diimplementasikan menggunakan metode Algoritma C4.5 hasil analisis menunjukkan bahwa Algoritma C4.5 dapat memprediksi dengan nilai akurasi sebesar 82 %[11].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahap penelitian ini menggunakan metode pengembangan sistem *Rapid Application Development* (RAD) yang merupakan metode pengembangan perangkat lunak dengan pendekatan berorientasi objek (*object oriented approach*) terhadap pengembangan sistem. *Rapid Application Development* (RAD) sebuah metode pengembangan sistem dengan pendekatan prototyping yang dirancang untuk menghasilkan sistem yang berkualitas tinggi dengan waktu yang relatif cepat serta biaya yang cukup rendah[12][13]. Lalu Peneliti ini menggunakan metode *Decision Tree* dalam menentukan ketepatan waktu lulus mahasiswa. Tahapan dalam penelitian ini dapat dilihat pada gambar berikut.



Gambar 1. Tahapan Penelitian

Adapun prosesnya adalah sebagai berikut : (1) tahapan dimulai dengan membaca dataset berupa data mahasiswa dengan format file *exls*; (2) setelah pembacaan dataset selanjutnya adalah proses transformasi data agar data sesuai dengan format yang akan di baca pada aplikasi dan menghindari data *missing value* atau data tidak lengkap; (3) tahap selanjutnya adalah proses prediksi dengan menggunakan algoritma *decision tree*. Model diuji untuk mendapatkan hasil prediksi berupa jumlah mahasiswa yang akan lulus tepat waktu dan jumlah mahasiswa yang diprediksi tidak lulus tepat waktu.

2.2 Metode yang Digunakan

Metode yang digunakan dalam penelitian ini yaitu metode algoritma *decision tree*. Algoritma C4.5 merupakan salah satu algoritma yang telah secara luas digunakan, khususnya di area machine learning yang memiliki beberapa perbaikan dari algoritma sebelumnya, ID3, yaitu dalam hal metode pemangkasanya (*prunning*). Adapun perbaikannya adalah sebagai berikut: (1) Algoritma C4.5 menghitung *gain ratio* untuk masing-masing atribut, dan atribut yang memiliki nilai yang tertinggi akan dipilih sebagai simpul. Penggunaan *gain ratio* ini memperbaiki

kelemahan dari ID3 yang menggunakan information gain; (2) Pemangkasan dapat dilakukan pada saat pembangunan pohon (tree) ataupun pada saat proses pembangunan pohon selesai; (3) Mampu menangani continues attribute; (4) Mampu menangani missing data; (5) Mampu membangkitkan rule dari sebuah pohon. Untuk menghitung gain digunakan rumus sebagai berikut [8];

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

Keterangan :

S : Himpunan Kasus

A : Atribut

n : Jumlah partisi Atribut

|S_i| : Jumlah Kasus Pada partisi ke-i

|S| : Jumlah Kasus dalam S

Sementara itu, perhitungan nilai entropi dengan rumus dasar *Entropy* adalah sebagai berikut[8]:

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

Keterangan :

S : Himpunan kasus

n : Jumlah partisi S

Pi : Proporsi dari Si terhadap S.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dalam penelitian ini digunakan dataset dengan ekstensi file exls. Jumlah dataset yang digunakan yaitu sebanyak 295 data mahasiswa yang diambil di tahun 2019-2020 dengan 12 atribut seperti pada tabel berikut ini :

Tabel 1. Atribut Dataset

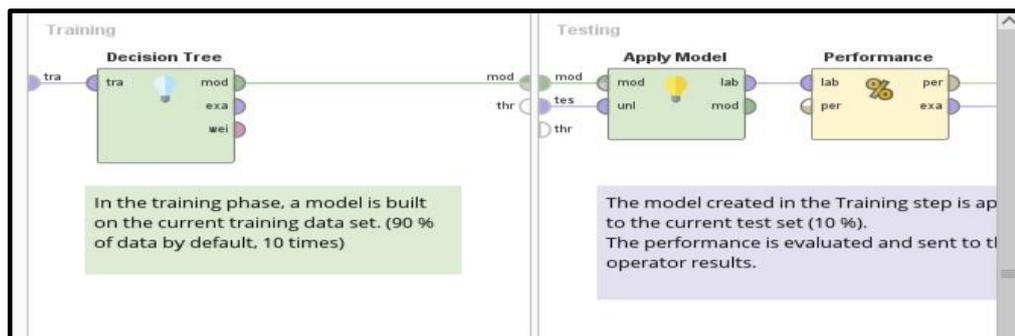
No	Atribut	Type Data
1	NPM	Varchar
2	Nama	Varchar
3	Jenis_Kelamin	Varchar
4	Index_Prestasi1	float
5	Index_Prestasi2	float
6	Index_Prestasi3	float
7	Index_Prestasi4	float
8	Index_Prestasi5	float
9	Index_Prestasi6	float
10	Index_Prestasi7	float
11	Index_Prestasi8	float
12	Index_Prestasi_Kumulatif	float

Tabel 2. Dataset Mahasiswa

NPM	Nama	Jenis_Kelamin	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPS8	IPK
15.420.001	Febri Pratama	Laki-laki	3.61	3.16	3.30	3.80	3.40	2.25	3.27	3.53	
15.420.002	Romadon Alvike	Laki-laki	3.20	3.30	3.14	3.14	2.84	3.13	3.25	2.25	
15.420.003	Bahari Saputra	Laki-laki	3.70	3.15	3.60	3.64	3.40	3.57	2.50	2.70	
15.420.004	Mertanti Rualiza	Perempuan	3.27	3.47	3.55	3.89	3.73	3.46	3.46	3.00	
15.420.005	Jurgi	Laki-laki	3.23	3.22	3.16	3.20	2.97	2.80	3.33	3.33	
15.420.006	Erwin Sepriadi	Laki-laki	3.28	3.28	2.75	2.75	2.75	3.00	3.00	3.00	
15.420.007	Mufel	Laki-laki	3.77	3.77	3.60	3.73	3.86	3.92	3.89	3.50	
15.420.008	Kasih	Perempuan	3.58	3.50	3.37	3.60	3.88	3.82	3.67	3.90	
15.420.009	Neni Agustin	Perempuan	3.35	3.35	3.22	3.53	3.34	3.28	3.73	3.52	
15.420.010	Agus Salim	Laki-laki	3.92	3.92	3.78	3.78	3.90	3.81	3.94	2.70	
15.420.011	Okta	Perempuan	3.56	2.50	2.50	2.50	2.75	2.75	3.00	3.30	
15.420.012	Riko Anggriansyah	Laki-laki	3.33	3.33	3.30	3.25	3.78	3.45	3.35	3.30	
15.420.013	Tedi Juliansyah	Laki-laki	3.30	3.27	3.47	3.55	3.89	3.73	3.46	3.46	
15.420.014	Zandi Ronaldo	Laki-laki	3.70	3.23	3.22	3.16	3.20	2.97	2.80	3.33	

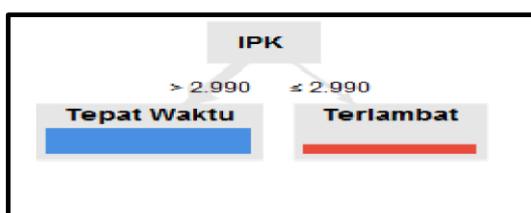
3.2 Hasil Implementasi Algoritma Decision Tree

Pada proses ini data akan diuji dengan menggunakan aplikasi rapid miner guna mendapatkan pola untuk membuat sebuah pohon keputusan.



Gambar 2. Dataset Rapid Miner

Pola ini menghasilkan 2 kemungkinan yaitu jika IPK lebih dari >2.90 maka akan lulus tepat waktu dan jika IPK kurang dari <2.90 maka akan terlambat.



Gambar 3. Pohon keputusan akhir

3.3 Hasil Confusion Matrik Prediksi

Hasil dari confusion matrix prediksi kelulusan mahasiswa dengan dataset yang berjumlah 295 data memiliki tingkat akurasi sebesar 99,67%. Dengan jumlah mahasiswa yang diprediksi akan lulus tepat waktu berjumlah 210 orang sedangkan mahasiswa yang diprediksi akan terlambat berjumlah 83 orang mahasiswa. Hasil dari confusion matrix juga dapat dilihat pada table berikut.

Table 3. Hasil Confusion Matrix Untuk Prediksi

Accuracy: 99,67 %			
	True tepat waktu	True terlambat	Class precision
Pred Tepat Waktu	210	0	100,00%
Pred Terlambat	1	83	98,81%
Class Recall	99.53%	100,00 %	

4. KESIMPULAN

Simpulan dalam Penelitian ini menghasilkan sebuah sistem Penentuan Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Decision Tree* agar dapat membantu pihak Kampus Sekolah Tinggi Teknologi Pagar Alam dalam memprediksi kelulusan mahasiswa tepat waktu. Hasil dari proses prediksi penentuan kelulusan mahasiswa tepat waktu melalui aplikasi *rapidminer* sama dengan hasil yang di terapkan di sistem yang telah di bangun yang menghasilkan pohon/pola yaitu jika IPK lebih dari > 2,990 maka tepat dan jika IPK kurang dari < 2.990 maka terlambat.

REFERENCES

- [1] R. Rosnelly *et al.*, "ANALISIS METODE DECISION TREE DALAM," pp. 131-140.
- [2] M. Ridwan, "Sistem Rekomendasi Proses Kelulusan Mahasiswa berbasis Algoritma Klasifikasi C4.5," *J. Ilm. Inform.*, vol. 2, no. 1, pp. 105-111, 2017, doi: 10.35316/jimi.v2i1.460.
- [3] R. H. Kusumodestoni and S. Sarwido, "Komparasi Model Support Vector Machines (Svm) Dan Neural Network Untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Harga Saham," *J. Inform. Upgris*, vol. 3, no. 1, 2017, doi: 10.26877/jiu.v3i1.1536.
- [4] N. Y. Septian, "Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro," *J. Semant.* 2013, pp. 1-11, 2009.
- [5] J. Fadlil and W. F. Mahmudy, "Pembuatan Sistem Rekomendasi Menggunakan Decision Tree dan Clustering," *Kursor*, vol. 3, No.1, no. 1, pp. 1-10, 2007.
- [6] Endang Supriyadi, "Metode SVM Berbasis PSO untuk Meningkatkan Prediksi Ketepatan Waktu Kelulusan Mahasiswa," *J. Sist. Inf.*, vol. 6, no. 2, p. 335171, 2017, doi: 10.31227/osf.io/4pyrt.
- [7] F. Dwi Meliani Achmad, Budanis, Slamet, "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree," *J. IPTEK*, vol. 16, no. 1, pp. 18-23, 2012, [Online]. Available: <http://jurnal.itats.ac.id/wp->



content/uploads/2013/06/3.-BUDANIS-FINAL-hal-17-23.pdf.

- [8] R. T. Wulandari, “Pengertian Data Mining,” *Data Min.*, vol. 7, no. 3, pp. 3–9, 2010.
- [9] K. Pustaka, “Penerapan Algoritma Decision Tree Id3 Untuk Prediksi Kelulusan Mahasiswa Jenjang Pendidikan D3 Di Fakultas Teknik Universitas Pandanaran,” *Neo Tek.*, vol. 5, no. 2, pp. 2–6, 2019, doi: 10.37760/neoteknika.v5i2.1391.
- [10] A. Romadhona, Suprapedi, and H. Himawan, “Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree,” *J. Teknol. Inf.*, vol. 13, pp. 69–83, 2017.
- [11] I. P. Astuti, “Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5,” *Fountain Informatics J.*, vol. 2, no. 2, p. 5, 2017, doi: 10.21111/fij.v2i2.1067.
- [12] A. Rohman and A. Rufiyanto, “Komparasi Model Decision Tree Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandanaran,” *Neo Tek.*, vol. 6, no. 1, pp. 8–12, 2020, doi: 10.37760/neoteknika.v6i1.1427.
- [13] M. A. R. Sikumbang, R. Habibi, and S. F. Pane, “Sistem Informasi Absensi Pegawai Menggunakan Metode RAD dan Metode LBS Pada Koordinat Absensi,” *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 59, 2020, doi: 10.30865/mib.v4i1.1445.