

Pemanfaatan Algoritma ID3 sebagai Prediktor Kelulusan Mahasiswa AMIK Harapan Bangsa Surakarta

Ismail setiawan¹, Eko Purbiyanto²

¹Dosen Tetap, AMIK Harapan Bangsa Surakarta

Alamat (Telp. (0271) 630029 Fax. (0271) 630029; e-mail: ismailsetiawan@amikhb.ac.id)

Abstract— Student graduation is one of the assessments in the accreditation of study programs and institutions. The high rate of student graduation on time is directly proportional to the value of accreditation. Therefore each college tries its best to make its students graduate on time. Some strategies were finally carried out by universities such as making meetings with each Academic Advisor for each generation so that there was communication that could ultimately provide a solution if there was a potential for students who did not graduate on time. This study tries with another approach to predict students who have the potential to not pass on time. Computation calculations by paying attention to existing data are interesting topics to be discussed. The ID3 algorithm has the ability to make decision trees about a problem. The algorithm is used in this study and found that the first factor that needs to be considered in graduating students at AMIK Harapan Bangsa Surakarta is gender. Gender with male types has a greater percentage of graduates than women with a percentage of 60% versus 40%. This can be a concern for AMIK Harapan Bangsa Surakarta towards final year students to detect as early as possible the potential of students who do not graduate on time.

Keywords— Students, graduate on time, ID3, AMIK Harapan Bangsa Surakarta, Gender.

Intisari— Kelulusan mahasiswa menjadi salah satu penilaian dalam akreditasi program studi maupun institusi. Tingginya angka kelulusan mahasiswa tepat waktu berbanding lurus dengan nilai akreditasi. Oleh karena itu masing-masing perguruan tinggi berusaha sebaik mungkin untuk membuat mahasiswanya lulus tepat waktu. Beberapa strategi akhirnya dilakukan oleh perguruan tinggi seperti membuat pertemuan dengan Pembimbing Akademik masing-masing angkatan sehingga terjadi komunikasi yang akhirnya dapat memberikan solusi jika terdeteksi adanya potensi mahasiswa yang tidak lulus tepat waktu. Penelitian ini mencoba dengan pendekatan lain untuk memprediksi mahasiswa yang memiliki potensi tidak lulus tepat waktu. Perhitungan komputasi dengan memperhatikan data yang telah ada sebelumnya menjadi topik yang menarik untuk dibahas. Algoritma ID3 memiliki kemampuan untuk membuat pohon keputusan mengenai sebuah masalah. Algoritma tersebut digunakan dalam penelitian ini dan didapatkan hasil bahwa faktor pertama yang perlu di perhatikan dalam kelulusan mahasiswa di AMIK Harapan Bangsa Surakarta adalah jenis kelamin. Kelamin dengan jenis pria memiliki prosentase lulus lebih besar dari wanita dengan prosentase 60% berbanding 40%. Hal ini dapat menjadi perhatian bagi AMIK Harapan Bangsa Surakarta terhadap mahasiswa tingkat akhir untuk mendeteksi sedini mungkin potensi mahasiswa yang tidak lulus tepat waktu.

Kata kunci— Mahasiswaa, lulus tepat waktu, ID3, AMIK Harapan Bangsa Surakarta, Jenis Kelamin.

I. PENDAHULUAN

Berkembangnya ilmu pengetahuan dewasa ini menghasilkan bidang ilmu baru. Salah satu ilmu yang sedang berkembang pesat adalah teknologi informasi dan komunikasi (TIK). Perkembangan TIK menghasilkan banyak cabang ilmu baru diantaranya perancangan sistem informasi. Pada ilmu tersebut dibahas bagaimana membangun sistem informasi yang di desain khusus untuk mempermudah pekerjaan manusia dan menghurangi tingkat human error. Penggunaan aplikasi yang dihasilkan dari perancangan sistem informasi dewasa ini menjadi tren dan mampu meningkatkan nilai bisnis sebuah instansi (Setiawan et al., 2009).

Perguruan tinggi sebagai instansi Formal yang bergerak dibidang pendidikan memiliki mahasiswa sebagai komponen utama untuk berjalannya proses pendidikan. Perguruan tinggi dan swasta memiliki aturan yang berbeda pada perguruan tinggi negeri. Pada perguruan tinggi swasta Mahasiswa perlu dikelola secara baik sehingga mereka merasa nyaman hingga akhir masa kuliahnya (Andriani, 2012). Hal ini disebabkan perguruan tinggi swasta memerlukan sumbangan dari mahasiswa untuk melakukan kegiatan perkuliahan. Semakin banyak jumlah mahasiswanya maka semakin banyak fasilitas yang dapat dibangun untuk pelayanan.

Tidak dipungkiri terkadang mahasiswa tidak melanjutkan perkuliahan hingga selesai masa studinya. Banyak faktor yang menjadi alasan mahasiswa tidak melanjutkan kuliah hingga selesai (Nugroho, 2014). Hal ini menjadi persoalan tersendiri yang perlu di berikan perhatian khusus oleh kemahasiswaan. Selain berpengaruh terhadap rasio jumlah mahasiswa terhadap dosen bagian keuangan juga akan merasakan dampaknya (Fakhrurriqfi & Wardoyo, 2013). Kegiatan

bisa tidak terlaksana dikarenakan tidak adanya anggaran yang masih bersumber dari sumbangan mahasiswa.

AMIK Harapan Bangsa Surakarta merupakan salah satu perguruan tinggi swasta di kota Surakarta. Perguruan tinggi tersebut memiliki satu program studi manajemen informatika dan jumlah mahasiswa sekitar 500 orang. Jumlah tersebut harus tetap dijaga sejak awal masuk hingga lulus untuk menjaga kestabilan keuangan instansi. Perlu perlakuan khusus kepada mahasiswa yang memiliki indikasi keluar di tengah jalan perkuliahan.

Peran utama data mining adalah melakukan estimasi, prediction, clasification, clustering dan association (Wu & Kumar, 2009). Untuk mendeteksi mahasiswa yang memiliki kemungkinan untuk berhenti ditengah masa perkuliahan maka tugas prediksi cocok untuk mengolah datanya. Algoritma ID3 cocok untuk melakukan tugas prediksi (Slocum, 2012). Penelitian ini menggunakan algoritma ID3 untuk melakukan tugas prediksi mahasiswa yang memiliki kecenderungan untuk berhenti di tengah masa perkuliahan

II. TINJAUAN PUSTAKA

1. Data Selection

Pada tahap ini akan dilakukan pengumpulan data yang berhubungan dengan mahasiswa. Data diambil pada bian BAAK dan kemahasiswaan. Data tersebut akan di lakukan penggabungan atau sinkronisasi untuk kemudian dilakukan pemilihan atribut data yang sesuai dengan jalannya penelitian. Data yang telah mengalami proses seleksi akan disimpan dalam database tersendiri untuk proses selanjutnya.

2. Pre-processing

Data yang telah mengalami proses seleksi, data akan dibersihkan dari atribut yang tidak mendukung dalam proses prediksi (Elmande & Widodo, 2012). Atribut yang dibersihkan adalah data yang memiliki duplikasi. Selain itu proses pre-processing akan melakukan perbaikan data yang salah ketik (tipografi). Berikutnya pada tahap ini akan

dilakukan pemerayaan data yang sudah ada dengan data atau informasi lain yang relevan untuk proses KDD.

3. Transformation

Tahap selanjutnya adalah pencarian fitur yang membantu dalam mencapai keputusan. Data yang terpilih pada proses ini akan dilakukan proses transofrmasi sehingga mudah terbaca oleh model yang dibangun nanti. Data pada tugas prediksi merupakan data nominal, oleh karena itu data dirubah menjadi nilai yang berbentuk nominal (Lavanya & Rani, 2011).

4. Data Mining

Pada tahap ini data akan di pisah menjadi data learning dan data testing. Pada data learning algoritma akan membentuk model. Model tersebut diujikan menggunakan data testing untuk mengetahui tingkat akurasi dari model yang terbentuk (Ugulino, Cardador, & Vega, 2012). Apabila nilai akurasi dari model yang terbentuk dirasa belum memuaskan maka perubahan indikator dan nilai penghitungan akan dilakukan pula pada tahap ini. Setelah model yang dibangun memiliki akurasi yang dirasa cukup memadai maka tahap ini dianggap selesai.

5. Interpretation

Tahap ini merupakan implementasi dari model yang telah berhasil dibangun. Model dikembangkan menggunakan aplikasi berbasis web, kemudian siap digunakan untuk melakukan tugas prediksi pada data baru yang diinputkan. Hasil dari prediksi dapat digunakan manajemen untuk mendukung keputusan

III. HASIL DAN PEMBAHASAN

A. Data selection

Proses awal dari pembahasan penelitian ini adalah membuat data selection atau pengumpulan data. Pengumpulan data dilakukan dengan mengambil dari bagian akademik. Untuk tahapannya dilakukan dua kali agar data yang digunakan semakin banyak dan proses komputasi menjadi semakin baik. Data awal disajikan pada tabel 1 sebagai berikut.

Tabel 1. Data Akademik Siswa Angkatan 2012-2013

no	jenis kelamin	jenis seleksi	pendapatan ayah	pendidikan ibu	ip s1	ip s2	ip s3	ip s4	sks s1	sks s2	sks s3	sks s4	status
1	Wanita	Undangan beasiswa	Rp 2.000.000	S1	2,77	3,62	3,48	3,39	22	22	22	22	Tidak tepat waktu
2	Wanita	Mandiri	Rp 1.500.000	SMA	2,82	3,62	3,19	3,36	22	22	22	22	Tepat waktu
3	PRIA	Reguler	Rp 2.000.000	S1	2,88	3,43	3,24	3,20	22	22	22	22	Tidak tepat waktu
.
.
14	9	Wanita	Reguler	Rp 2.500.000	S1	3,23	3,07	3,55	3,51	22	22	22	Tepat waktu
15	0	Wanita	Mandiri	Rp 2.500.000	Tidak Sekolah	3,03	3,56	3,78	2,90	22	22	22	Tidak tepat waktu

Tabel 3. Data Akademik Siswa Angkatan 2016-2017

no	jenis kelamin	jenis seleksi	pendapatan ayah	pendidikan ibu	ip s1	ip s2	ip s3	ip s4	sks s1	sks s2	sks s3	sks s4	status
151	Wanita	Mandiri	Rp 1.000.000	Tidak Sekolah	3,48	3,46	2,87	3,75	22	22	22	22	Tidak tepat waktu
152	Wanita	Mandiri	Rp 2.500.000	Tidak Sekolah	3,01	3,26	2,86	3,63	22	22	22	22	Tidak tepat waktu
153	Wanita	Mandiri	Rp 2.500.000	S1	3,15	3,49	2,70	3,14	22	22	22	22	Tidak tepat waktu
.
.
299	Wanita	Mandiri	Rp 2.500.000	S1	3,75	3,27	3,52	3,29	22	22	22	22	Tepat waktu
300	Wanita	Mandiri	Rp .500.000	Tidak Sekolah	3,45	2,84	3,63	3,57	22	22	22	22	Tepat waktu

B. Pre-processing

Pada tahap ini dilakukan pembersihan data dan pelengkapan data yang tidak ada dalam tabel. Pada kasus ini data yang dikumpulkan telah terisi semua sehingga dapat dilakukan proses selanjutnya.

C. Transformation

Tahap ini seharusnya merubah data yang sifatnya belum nominal menjadi nominal sehingga mudah dalam mengambil keputusan. Kasus yang dibahas ini data sudah dalam bentuk nominal sehingga tidak perlu dilakukan transformasi data. Selanjutnya adalah proses data mining.

D. Data mining

Untuk pengolahan data dengan algoritma ID3 digunakan rumus dengan persamaan sebagai berikut.

$$\text{Entropi (S)} = \sum_{j=1}^k - p_j \log_2 p_j \quad (1)$$

Dimana:

- S adalah himpunan (dataset) kasus
- k adalah banyaknya partisi S
- p_j adalah probabilitas yang di dapat dari $\text{Sum}(Y_a)$ dibagi Total Kasus.

Setelah mendapat nilai entropi, pemilihan atribut dilakukan dengan nilai information gain terbesar.

$$\text{Gain (A)} = \text{entropi (S)} - \sum_{j=1}^k \frac{|S_j|}{|S|} \text{Entropi (S}_j) \quad (2)$$

Dimana:

S = ruang (data) sample yang digunakan untuk training.

A = atribut.

|Si| = jumlah sample untuk nilai V.

|S| = jumlah seluruh sample data.

Entropi(Si) = entropy untuk sample-sample yang memiliki nilai *i*

Pada perhitungan pertama dalam pencarian gain didapati gain terbesar dimiliki oleh jenis kelamin, sehingga jenis kelamin menjadi akar pertama dalam pohon keputusan.

TABEL 3
PROSES PENGHITUNGAN GAIN

node	jumlah kasus	Ya	tidak	entropi	gain
1 total	301	145	156	0,999	
jenis kelamin					0,995293484
Wanita	142	63	79	0,9908	
PRIA	159	82	77	0,9993	
jenis seleksi					0,003601698
Seleksi disekolah	54	22	32	0,9751	
Mandiri	77	38	39	0,9999	
Reguler	123	62	61	1	
Undangan beasiswa	47	23	24	0,9997	
pendapatan ayah					0,007080234
1000000	73	39	34	0,9966	
1500000	67	32	35	0,9986	
2000000	43	16	27	0,9523	
2500000	118	58	60	0,9998	
pendidikan ibu					0,017232093
tidak sekolah	60	32	28	0,9968	
sd	51	29	22	0,9864	
smp	43	16	27	0,9523	
sma	35	12	23	0,9275	
s1	112	56	56	1	
ip s1					4,00648E-06
Dibawah Tiga	91	44	47	0,9992	
Diatas Tiga	210	101	109	0,999	
ip s2					0,001439256
Dibawah Tiga	81	42	39	0,999	
Diatas Tiga	220	103	117	0,9971	
ip s3					0,000808849
Dibawah Tiga	92	42	50	0,9945	
Diatas Tiga	209	106	103	0,9999	
ip s4					0,00033728
Dibawah Tiga	78	39	39	1	
Diatas Tiga	223	106	117	0,9982	
sks 1					0,001507288
20	90	45	45	1	
21	113	56	57	0,9999	
22	98	44	54	0,9925	

sks 2					0,007524501
	20	102	56	46	0,9931
	21	94	40	54	0,9839
	22	105	49	56	0,9968
sks 3					0,018952103
	20	97	36	61	0,9515
	21	109	55	54	0,9999
	22	95	54	41	0,9864
sks 4					0,013789212
	20	108	58	50	0,996
	21	99	38	61	0,9607
	22	94	49	45	0,9987

Setelah dilakukan beberapa iterasi hingga semua kasus sudah memenuhi nilai lulus dan tidak lulus tepat waktu didapatkan hasil pohon keputusan sebagai berikut.

a. Wanita

|sks 4 - 20

|--|sks 3 20

|----|Ip S1 diatas tiga – Lulus tepat waktu / tidak lulus tepat waktu

|----|Ip S1 Dibawah tiga – Lulus tepat waktu / tidak lulus tepat waktu

|--|Sks 3 21

|----|Ip S1 diatas tiga – Lulus tepat waktu / tidak lulus tepat waktu

|----|Ip S1 dibawah tiga – Lulus tepat waktu / tidak lulus tepat waktu

|--|Sks 3 22

|----|Ip S4 diatas tiga – Lulus tepat waktu

|----|Ip S4 dibawah tiga – Lulus tepat waktu

|SKS 4 - 21

|--|Pendapatan ayah 1000000

|----|pendidikan ibu Tidak Sekolah –tidak lulus tepat waktu

|----| pendidikan ibu SD – Lulus tepat waktu / tidak lulus tepat waktu

|----| pendidikan ibu SMP – Lulus tepat waktu / tidak lulus tepat waktu

|----| pendidikan ibu SMA – Lulus tepat waktu

|----| pendidikan ibu S1 – tidak lulus tepat waktu

|--|Pendapatan ayah 1500000

|----|pendidikan ibu Tidak Sekolah – Lulus tepat waktu / tidak lulus tepat waktu

|----|pendidikan ibu SD – tidak lulus tepat waktu

|----| pendidikan ibu SMP – tidak lulus tepat waktu

|----| pendidikan ibu SMA – tidak lulus tepat waktu

|----| pendidikan ibu S1 – tidak lulus tepat waktu

|--|Pendapatan ayah 2000000

|----|SKS 3 - 20 - tidak lulus tepat waktu

|----|SKS 3 – 21 - Lulus tepat waktu

|----|SKS 3 – 22 – Lulus tepat waktu

|--|Pendapatan ayah 2500000

|----|pendidikan ibu Tidak Sekolah –tidak lulus tepat waktu

|----| pendidikan ibu SD – Lulus tepat waktu

|----| pendidikan ibu SMP – Lulus tepat waktu

|----| pendidikan ibu SMA – tidak lulus tepat waktu

|----| pendidikan ibu S1 – Lulus tepat waktu

|sks 4 – 22

|--| IP S4 – dibawah tiga - tidak lulus tepat waktu

|--| IP S4 – diatas 3 - Lulus tepat waktu / tidak lulus tepat waktu

b. Pria

| SKS 4 – 20

- |--| IP S1 – dibawah 3 - tidak lulus tepat waktu
- |--| IP S1 – Diatas 3 - Lulus tepat waktu / tidak lulus tepat waktu
- |Sks 4 – 21
- |--|Seleksi mandiri – Lulus tepat waktu
- |--| Seleksi sekolah – Lulus tepat waktu
- |--| Seleksi reguler – Lulus tepat waktu
- |--|Seleksi undangan – Lulus tepat waktu
- |Sks 4 -22
- |--|Seleksi mandiri – Lulus tepat waktu
- |--|Seleksi reguler – tidak lulus tepat waktu
- |--|Seleksi undangan – Lulus tepat waktu / tidak lulus tepat waktu
- |--|Seleksi disekolah - Lulus tepat waktu

Sedangkan untuk prosentase kelulusan mahasiswa berdasarkan jenis kelamin dapat dilihat pada tabel berikut

Tabel 4. Proses kelulusan berdasarkan jenis kelamin

NO	JENIS KELAMIN	LULUS TEPAT WAKTU	TIDAK LULUS TEPAT WAKTU
1	WANITA	40%	60%
2	PRIA	60%	40%

IV. KESIMPULAN

Berdasarkan hasil perhiungan menggunakan algoritma ID3 pada data akademik mahasiswa AMIK Harapan Bangsa Surakarta tahun 2012-2013 dan 2016-2017 didapatkan bahwa faktor utama penentu kelulusan mahasiswa tepat waktu adalah jenis kelamin. Sedangkan untuk faktor kedua dari jenis kelamin wanita adalah sks yang diambil untuk semester 4. Pada jenis kelamin pria faktor penentu selanjutnya adalah sks yang diambil pada semester 4. Artinya faktor kedua pada masing- masing jenis kelamin adalah sama. Hal ini dapat menjadi perhatian bagi bagian akademik untuk berkordinasi dengan masing-masing pendamping akademik setiap angkatan yang mahasiswanya akan memasuki masa tugas akhir agar menyiapkan strategi khusus untuk mahasiswa tersebut. Prosentase yang paling besar untuk lulus tidak tepat waktu adalah jenis kelamin wanita. Sehingga mahasiswa dengan jenis kelamin wanita

perlu mendapatkan penanganan khusus jika sudah menunjukkan potensi tidak lulus tepat waktu.

DAFTAR PUSTAKA

- [1] Andriani, A. (2012). Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout. Seminar Nasional Matematika, 139–147. Retrieved from [http://demo.pohonkeputusan.com/files/PENERAPAN ALGORITMA C4.5 PADA PROGRAM KLASIFIKASI MAHASISWA DROPOUT.pdf?i=1](http://demo.pohonkeputusan.com/files/PENERAPAN_ALGORITMA_C4.5_PADA_PROGRAM_KLASIFIKASI_MAHASISWA_DROPOUT.pdf?i=1)
- [2] Elmande, Y., & Widodo, P. (2012). Pemilihan Criteria Splitting dalam Algoritma Iterative Dichotomiser 3 (ID3) untuk Penentuan Kualitas Beras: Studi Kasus Pada Perum Bulog Divre Lampung. *Jurnal TELEMATIKA MKOM*, 4(1).
- [3] Fakhrrurriqfi, M., & Wardoyo, R. (2013). Perbandingan Algoritma Nearest Neighbour, C4. 5 dan LVQ untuk Klasifikasi Kemampuan Mahasiswa. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 7(2), 145–154.
- [4] Lavanya, D., & Rani, K. U. (2011). Performance Evaluation of Decision Tree Classifiers on Medical Datasets. *International Journal of Computer Applications*, 26(4), 1–4. <https://doi.org/10.5120/3095-4247>
- [5] Mueller, M., dos Santos, V. G., & Seuring, S. (2009). The contribution of environmental and social standards towards ensuring legitimacy in supply chain governance. *Journal of Business Ethics*, 89(4), 509–523. <https://doi.org/10.1007/s10551-008-0013-9>
- [6] Nugroho, Y. S. (2014). Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi Dan Informatika Universitas Muhammadiyah Surakarta. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014*, (November), 1–6. <https://doi.org/10.13140/RG.2.1.2734.8247>
- [7] Setiawan, B., Widjaja, R. S., Learning, D. T., Quinlan, R., Kapal, K., & Belakang, L. (2009). Untuk Menentukan Kelaiklautan Kapal. *Prosiding Seminar Nasional Manajemen Teknologi X*, 1–7. <https://doi.org/10.1186/1471-2164-15-579>
- [8] Slocum, M. (2012). Decision Making Using Id3 Algorithm. *InSight: RIVIER ACADEMIC JOURNAL*, 8(2), 1–12.
- [9] Ugulino, W., Cardador, D., & Vega, K. (2012). *Wearable Computing : Accelerometers ' Data*, 52–53.
- [10] Wu, X., & Kumar, V. (2009). *the top ten algorithms in data mining.* (V. Kumar, Ed.). london: crc press.