

Analisis Evaluasi Kinerja Akademik Mahasiswa Program Studi Teknik Informatika Universitas Mulawarman Menggunakan Algoritma CART

Alfan Nurfaizin^{*1}, Nataniel Dengen², Herman Santoso Pakpahan³

^{1,2,3}Program Studi Informatika, Universitas Mulawarman, Samarinda

e-mail: ^{*1}alfannurfaizin02@gmail.com, ²ndengen@gmail.com,

³pakpahan.herman891@gmail.com

Abstrak

Selama mahasiswa berada di perguruan tinggi tentunya telah menghasilkan informasi yang sangat banyak berupa profil maupun data nilai mahasiswa. Informasi ini dapat dioptimalkan menggunakan algoritma dengan teknik data mining untuk dianalisis guna menemukan pola-pola yang berhubungan dengan kelulusan mahasiswa tersebut. Pada penelitian ini, data mining diterapkan pada model proses Cross Industry Standart Process for Data Mining (CRISP-DM) yang menyediakan proses standar penggunaan data mining pada berbagai bidang. Algoritma yang digunakan adalah Classification and Regression Trees (CART) untuk evaluasi kinerja akademik mahasiswa. Algoritma ini mengolah dataset profil beserta nilai mahasiswa yang berjumlah 158 data. Dataset telah memiliki label predikat kelulusan, yaitu tepat waktu dan terlambat. Adapun variabel yang digunakan yaitu jenis kelamin, umur mahasiswa saat mendaftar, asal daerah, status sekolah asal, jenis sekolah asal, serta indeks prestasi semester 1 hingga 4. Berdasarkan evaluasi yang telah dilakukan menggunakan metode confusion matrix, didapatkan akurasi algoritma sebesar 71,79%. Serta variabel yang paling berpengaruh terhadap kelulusan adalah Indeks Prestasi Semester 2.

Kata kunci— Evaluasi Akademik Mahasiswa, Decision Tree, Algoritma CART

1. PENDAHULUAN

Perguruan tinggi merupakan bagian terpenting dalam menentukan kualitas suatu negara. Hal tersebut dapat dilihat dari mutu dosen dan tenaga kependidikan, mutu prasarana (ruang kelas, laboratorium, dan sebagainya) yang diberikan kepada mahasiswa dalam menempuh pendidikan. Indikasi mutu mahasiswa dapat dilihat dari etika, keaktifan dalam proses belajar mengajar, serta prestasi akademik [1].

Universitas Mulawarman adalah sebuah perguruan tinggi yang berada di kota Samarinda, Kalimantan Timur. Terdiri dari 14 Fakultas, Universitas Mulawarman merupakan perguruan tinggi dan memiliki 40% dari total mahasiswa di perguruan tinggi di Kalimantan Timur dan Utara. Salah satu dari fakultas yang berada di Universitas Mulawarman adalah Fakultas Ilmu Komputer dan Teknologi Informasi. Jumlah mahasiswa Fakultas Ilmu Komputer dan Teknologi Informasi per Desember 2017 tercatat sejumlah 1313 [2].

Berhubungan dengan pengelolaan lama studi mahasiswa, program studi berkewajiban untuk memantau perkembangan studi dari mahasiswanya. Untuk itu perlu adanya pemantauan maupun evaluasi terhadap kecenderungan mahasiswa lulus tepat waktu maupun terlambat. Dalam mengelola masa studi mahasiswa, sebuah program studi akan menghasilkan informasi berupa profil mahasiswa beserta rekaman akademik mahasiswa selama menempuh masa studi. Informasi ini yang nantinya akan menjadi data yang dapat dioptimalkan untuk dianalisis guna

menemukan pola-pola yang terjadi selama mahasiswa menjalankan masa studi untuk memperoleh informasi baru.

Dalam penelitian ini akan digunakan teknik *data mining* untuk menganalisis evaluasi kinerja akademik mahasiswa. *Data mining* sebagai sebuah teknik analisis berbasis basis data dan statistik dipandang sangat tepat digunakan untuk memperoleh informasi yang terkandung dalam sebuah basis data. Beberapa contoh algoritma *data mining* diantaranya C4.5, *Naïve Bayes*, *K-Nearest Neighbors*. Terdapat alternatif lain dalam pengklasifikasian *data mining*, yaitu *Artificial Neural Network* [3]. Berdasarkan pemaparan latar belakang, penulis memutuskan untuk menganalisis evaluasi kinerja akademik mahasiswa agar diketahui seberapa besar persentase model algoritma *classification and regression trees* (CART) dalam mengevaluasi kinerja akademik mahasiswa.

2. METODE PENELITIAN

2.1 Data Mining

Data mining adalah ekstraksi pola yang menarik dari kumpulan data dalam jumlah besar. Suatu pola dikatakan menarik apabila pola tersebut tidak sepele, implisit, tidak diketahui sebelumnya, dan berguna. Pola yang disajikan haruslah mudah dipahami, berlaku untuk data yang akan diprediksi dengan derajat kepastian tertentu, berguna, dan baru [4].

2.2 Algoritma CART

Algoritma CART merupakan metodologi statistik *non-parametrik* yang dikembangkan untuk topik analisis klasifikasi, baik untuk peubah respon kategorik maupun kontinu. Algoritma CART menghasilkan suatu pohon klasifikasi jika peubah responnya kategorik, dan menghasilkan pohon regresi jika peubah responnya kontinu. Tujuan utama Algoritma CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Ciri khas algoritma CART ini adalah noktah keputusan yang selalu bercabang dua atau bercabang biner [5] [6].

2.2.1 First Phase

Menyusun calon cabang (candidate split). Penyusunan ini dilakukan terhadap seluruh variabel prediktor yang masing-masing dipilah menjadi 2, yaitu calon cabang kiri dan calon cabang kanan.

2.2.2 Second Phase

Menilai kinerja keseluruhan calon cabang yang ada pada daftar calon cabang kiri. Kinerja dari setiap calon cabang akan diukur melalui ukuran yang disebut kesesuaian (goodness). Kesesuaian dari calon cabang s pada noktah keputusan t dilambangkan dan didefinisikan dalam persamaan (1).

$$\Phi(s | t) = 2P_L P_R \sum_{j=1}^{\text{jumlah kategori}} |P(j|t_L) - P(j|t_R)|$$

1

2.2.3 Third Phase

Menentukan calon cabang manakah yang benar-benar dijadikan cabang. Hal ini dilakukan dengan memilih calon cabang yang memiliki nilai kesesuaian $\Phi(s | t)$ terbesar. Setelah itu menggambar percabangan sesuai hasil dengan algoritma. Jika tidak ada lagi noktah

keputusan, pelaksanaan algoritma CART akan dihentikan. Namun jika masih terdapat noktah keputusan, pelaksanaan algoritma dilanjutkan dengan kembali ke langkah kedua, dengan terlebih dahulu membuang calon cabang yang telah berhasil menjadi cabang.

2.3 Confusion Matrix

Confusion Matrix adalah pengukuran performa untuk masalah klasifikasi machine learning dimana keluaran dapat berupa dua kelas atau lebih confusion matrix adalah metode yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai aktual dan prediksi pada klasifikasi. Confusion matrix 2 kelas dapat dilihat pada tabel 1.

Tabel 1. Confusion Matrix

Aktual/Prediksi		Prediksi	
		Negatif	Positif
Aktual	Negatif	a	b
	Positif	c	d

Pada penelitian ini akan dilakukan perhitungan tingkat akurasi model algoritma CART menggunakan persamaan (2).

$$AC = \frac{a + d}{a + b + c + d}$$

2

2.4 CRISP-DM

Penelitian ini mengacu pada model Cross-Standard Industry for Data Mining (CRISP-DM). CRISP-DM terdiri dari 6 fase, yaitu [8] bussiness understanding, data understanding, data Preparation dan modeling. Pada tahap bussiness understanding penelitian akan berfokus pada persentase keakuratan model algoritma CART dalam evaluasi kinerja akademik mahasiswa. Pada tahap data understanding dimana data telah diperoleh data mahasiswa yang akan menjadi dataset. Pada tahap data preparation, dari data yang didapat akan dilakukan pemilihan atribut serta dilakukan data cleaning dan data transformation. Pada algoritma CART data terlebih dahulu akan disusun menjadi calon cabang (candidate split) terhadap seluruh variable secara lengkap, sehingga terbentuk daftar calon cabang seperti pada tabel 2.

Tabel 2. Calon Cabang

No	Calon Cabang Kiri	Calon Cabang Kanan
1	Jenis Kelamin = Laki-laki	Jenis Kelamin = Perempuan
2	Jenis Kelamin = Perempuan	Jenis Kelamin = Laki-laki
3	Umur < 17	Umur > 17
4	Umur = 17	Umur ∈ {< 17, > 17}
5	Umur = 18	Umur ∈ {< 18, > 18}
6	Umur = 19	Umur ∈ {< 19, > 19}
7	Umur > 19	Umur < 19
8	Asal Daerah = Samarinda	Asal Daerah = Luar Samarinda
9	Asal Daerah = Luar Samarinda	Asal Daerah = Samarinda
10	Status Sekolah Asal = Negeri	Status Sekolah Asal = Swasta
11	Status Sekolah Asal = Swasta	Status Sekolah Asal = Negeri
12	Jenis Sekolah Asal = SMA	Jenis Sekolah Asal ∈ {SMK, MA}
13	Jenis Sekolah Asal = SMK	Jenis Sekolah Asal ∈ {SMA, MA}
14	Jenis Sekolah Asal = MA	Jenis Sekolah Asal ∈ {SMA, SMK}

15	IP Semester 1 bobot 1	IP Semester 1 $\in \{< 1, > 1\}$
16	IP Semester 1 bobot 2	IP Semester 1 $\in \{< 2, > 2\}$
17	IP Semester 1 bobot 3	IP Semester 1 $\in \{< 3, > 3\}$
18	IP Semester 1 bobot 4	IP Semester 1 $\in \{< 4, > 4\}$
19	IP Semester 2 bobot 1	IP Semester 2 $\in \{< 1, > 1\}$
20	IP Semester 2 bobot 2	IP Semester 2 $\in \{< 2, > 2\}$
21	IP Semester 2 bobot 3	IP Semester 2 $\in \{< 3, > 3\}$
22	IP Semester 2 bobot 4	IP Semester 2 $\in \{< 4, > 4\}$
23	IP Semester 3 bobot 1	IP Semester 3 $\in \{< 1, > 1\}$
24	IP Semester 3 bobot 2	IP Semester 3 $\in \{< 2, > 2\}$
25	IP Semester 3 bobot 3	IP Semester 3 $\in \{< 3, > 3\}$
26	IP Semester 3 bobot 4	IP Semester 3 $\in \{< 4, > 4\}$
27	IP Semester 4 bobot 1	IP Semester 4 $\in \{< 1, > 1\}$
28	IP Semester 4 bobot 2	IP Semester 4 $\in \{< 2, > 2\}$
29	IP Semester 4 bobot 3	IP Semester 4 $\in \{< 3, > 3\}$
30	IP Semester 4 bobot 4	IP Semester 4 $\in \{< 4, > 4\}$

Selanjutnya akan dicari rasio setiap cabang terhadap jumlah *dataset* dengan menggunakan persamaan (1). Penentuan *data training* serta *data testing* menggunakan *linear sampling*, yang mana data akan dibagi berdasarkan urutan yang sebenarnya tanpa harus mengacak *dataset*. *Data training* akan berada pada urutan teratas, sementara sisanya merupakan *data testing*. Setelah semua rasio calon cabang ditemukan maka akan dapat menghitung nilai kesesuaian (*goodness*). Yang mana nilai kesesuaian tertinggi akan menjadi cabang dari pohon keputusan. Pada tahap *evaluation*, akan dilakukan pengukuran akurasi algoritma CART dengan metode *confusion matrix*. Model yang telah dihasilkan dari algoritma CART telah dapat diterapkan untuk evaluasi kinerja akademik mahasiswa.

3. HASIL DAN PEMBAHASAN

Data pada penelitian ini adalah data kelulusan mahasiswa FKTI angkatan 2012 hingga 2014, yang berjumlah 158 data. Akan dilakukan 9 percobaan untuk mencari model algoritma yang memiliki tingkat akurasi tertinggi dalam mengevaluasi kinerja akademik mahasiswa. Pada rasio 10% data training, algoritma CART akan memodelkan sejumlah 16 data. Setelah dilakukan perhitungan menggunakan persamaan (1), maka diketahui node awal (z) adalah IPS 3, karena memiliki nilai $\Phi(s | t)$ tertinggi seperti pada tabel 3.

Tabel 3. Node Awal 10% Data Training

No.	Variabel	Nilai $\Phi(s t)$
1	Jenis Kelamin	0,2656
2	Umur Daftar	0,3750
3	Asal Daerah	0,3125
4	Status Sekolah Asal	
5	Jenis Sekolah Asal	0,2188
6	Indeks Prestasi Semester 1	0,0156
7	Indeks Prestasi Semester 2	0,1719

8	Indeks Prestasi Semester 3	0,4219
9	Indeks Prestasi Semester 4	0,4219

Pada rasio 20% *data training*, algoritma CART akan memodelkan sejumlah 32 data. Setelah dilakukan perhitungan menggunakan persamaan (1), maka diketahui node awal (*root*) adalah Umur Daftar, karena memiliki nilai $\Phi(s | t)$ tertinggi seperti pada tabel 4.

Tabel 4. Node Awal 20% *Data Training*

No.	Variabel	Nilai $\Phi(s t)$
1	Jenis Kelamin	0,1313
2	Umur Daftar	0,1563
3	Asal Daerah	0,0938
4	Status Sekolah Asal	0,0781
5	Jenis Sekolah Asal	0,0781
6	Indeks Prestasi Semester 1	0,1094
7	Indeks Prestasi Semester 2	0,01533
8	Indeks Prestasi Semester 3	0,0469
9	Indeks Prestasi Semester 4	0,0859

Pada rasio 30% *data training*, algoritma CART akan memodelkan sejumlah 48 data. Setelah dilakukan perhitungan menggunakan persamaan (1), maka diketahui node awal (*root*) adalah IPS 4, karena memiliki nilai $\Phi(s | t)$ tertinggi seperti pada tabel 5.

Tabel 5. Node Awal 30% *Data Training*

No.	Variabel	Nilai $\Phi(s t)$
1	Jenis Kelamin	0,0660
2	Umur Daftar	0,1667
3	Asal Daerah	0,1771
4	Status Sekolah Asal	0,0903
5	Jenis Sekolah Asal	0,0625
6	Indeks Prestasi Semester 1	0,0417
7	Indeks Prestasi Semester 2	0,0114
8	Indeks Prestasi Semester 3	0,2257
9	Indeks Prestasi Semester 4	0,2465

Pada rasio 40% *data training*, algoritma CART akan memodelkan sejumlah 64 data. Setelah dilakukan perhitungan menggunakan persamaan (1), maka diketahui node awal (*root*) adalah IPS 2, karena memiliki nilai $\Phi(s | t)$ tertinggi seperti pada tabel 6.

Tabel 6. Node Awal 40% *Data Training*

No.	Variabel	Nilai $\Phi(s t)$
-----	----------	---------------------

1	Jenis Kelamin	0,1563
2	Umur Daftar	0,0625
3	Asal Daerah	0,2188
4	Status Sekolah Asal	0,0313
5	Jenis Sekolah Asal	0,0313
6	Indeks Prestasi Semester 1	0,1250
7	Indeks Prestasi Semester 2	0,3438
8	Indeks Prestasi Semester 3	0,2500
9	Indeks Prestasi Semester 4	0,2188

Pada rasio 50% *data training*, algoritma CART akan memodelkan sejumlah 80 data. Setelah dilakukan perhitungan menggunakan persamaan (1), maka diketahui node awal (*root*) adalah IPS 2, karena memiliki nilai $\Phi(s | t)$ tertinggi seperti pada tabel 7.

Tabel 7. Node Awal 50% *Data Training*

No.	Variabel	Nilai $\Phi(s t)$
1	Jenis Kelamin	0,2294
2	Umur Daftar	0,0638
3	Asal Daerah	0,2088
4	Status Sekolah Asal	0,0656
5	Jenis Sekolah Asal	0,1169
6	Indeks Prestasi Semester 1	0,1950
7	Indeks Prestasi Semester 2	0,4556
8	Indeks Prestasi Semester 3	0,2906
9	Indeks Prestasi Semester 4	0,2550

Pada rasio 60% *data training*, algoritma CART akan memodelkan sejumlah 96 data. Setelah dilakukan perhitungan menggunakan persamaan (1), maka diketahui node awal (*root*) adalah IPS 2, karena memiliki nilai $\Phi(s | t)$ tertinggi seperti pada tabel 8.

Tabel 8. Node Awal 60% *Data Training*

No.	Variabel	Nilai $\Phi(s t)$
1	Jenis Kelamin	0,3229
2	Umur Daftar	0,0729
3	Asal Daerah	0,1589
4	Status Sekolah Asal	0,0299
5	Jenis Sekolah Asal	0,1471
6	Indeks Prestasi Semester 1	0,3229

7	Indeks Prestasi Semester 2	0,4883
8	Indeks Prestasi Semester 3	0,3542
9	Indeks Prestasi Semester 4	0,3255

Pada rasio 70% *data training*, algoritma CART akan memodelkan sejumlah 112 data. Setelah dilakukan perhitungan menggunakan persamaan (1), maka diketahui node awal (*root*) adalah IPS 2, karena memiliki nilai $\Phi(s | t)$ tertinggi seperti pada tabel 9.

Tabel 9. Node Awal 70% *Data Training*

No.	Variabel	Nilai $\Phi(s t)$
1	Jenis Kelamin	0,3227
2	Umur Daftar	0,1378
3	Asal Daerah	0,1224
4	Status Sekolah Asal	0,0290
5	Jenis Sekolah Asal	0,0475
6	Indeks Prestasi Semester 1	0,3036
7	Indeks Prestasi Semester 2	0,4372
8	Indeks Prestasi Semester 3	0,3198
9	Indeks Prestasi Semester 4	0,2752

Pada rasio 80% *data training*, algoritma CART akan memodelkan sejumlah 128 data. Setelah dilakukan perhitungan menggunakan persamaan (1), maka diketahui node awal (*root*) adalah IPS 2, karena memiliki nilai $\Phi(s | t)$ tertinggi seperti pada tabel 10.

Tabel 10. Node Awal 80% *Data Training*

No.	Variabel	Nilai $\Phi(s t)$
1	Jenis Kelamin	0,3159
2	Umur Daftar	0,1567
3	Asal Daerah	0,0869
4	Status Sekolah Asal	0,0176
5	Jenis Sekolah Asal	0,0254
6	Indeks Prestasi Semester 1	0,3379
7	Indeks Prestasi Semester 2	0,4536
8	Indeks Prestasi Semester 3	0,3438
9	Indeks Prestasi Semester 4	0,2686

Pada rasio 90% *data training*, algoritma CART akan memodelkan sejumlah 144 data. Setelah dilakukan perhitungan menggunakan persamaan (1), maka diketahui node awal (*root*) adalah IPS 2, karena memiliki nilai $\Phi(s | t)$ tertinggi seperti pada tabel 11.

Tabel 11. 90% Data Training

No.	Variabel	Nilai $\Phi(s t)$
1	Jenis Kelamin	0,2986
2	Umur Daftar	0,1551
3	Asal Daerah	0,0556
4	Status Sekolah Asal	0,0093
5	Jenis Sekolah Asal	0,0532
6	Indeks Prestasi Semester 1	0,3542
7	Indeks Prestasi Semester 2	0,4282
8	Indeks Prestasi Semester 3	0,2940
9	Indeks Prestasi Semester 4	0,2778

Selanjutnya dilakukan perhitungan akurasi dari model algoritma yang telah didapat. Dari 10% data training akan diuji menggunakan 142 data testing menghasilkan akurasi sebesar 46,48%. Dari 20% data training akan diuji menggunakan 126 data testing menghasilkan akurasi sebesar 58,73%. Dari 30% data training akan diuji menggunakan 110 data testing menghasilkan akurasi sebesar 61,82%. Dari 40% data training akan diuji menggunakan 94 data testing menghasilkan akurasi sebesar 58,38%. Dari 50% data training akan diuji menggunakan 78 data testing menghasilkan akurasi sebesar 71,79%. Dari 60% data training akan diuji menggunakan 62 data testing menghasilkan akurasi sebesar 67,74%. Dari 70% data training akan diuji menggunakan 46 data testing menghasilkan akurasi sebesar 71,74%. Dari 80% data training akan diuji menggunakan 30 data testing menghasilkan akurasi sebesar 60,00%. Dari 90% data training akan diuji menggunakan 14 data testing menghasilkan akurasi sebesar 64,29%. Perbandingan akurasi dapat dilihat pada gambar 1.

Grafik Perbandingan Akurasi



Gambar 1. Grafik Perbandingan Akurasi

4. KESIMPULAN

Setelah melakukan analisis evaluasi akademik mahasiswa, maka diperoleh kesimpulan seperti evaluasi kinerja akademik mahasiswa dapat dilakukan dengan variabel jenis kelamin,

umur saat mendaftar perguruan tinggi, asal daerah, status sekolah asal, jenis sekolah asal, dan indeks prestasi semester 1 hingga semester 4. Variabel yang paling berpengaruh dalam penelitian ini adalah indeks prestasi semester 2. Akurasi terbaik diperoleh pada rasio 50% data training dengan nilai akurasi sebesar 71,79%.

5. SARAN

Dalam analisis evaluasi kinerja akademik mahasiswa ini masih terdapat kekurangan, oleh karena itu penulis ingin memberikan saran seperti menambah variabel yang belum digunakan maupun dataset agar menambah keakuratan model algoritma. Kemudian membandingkan dengan algoritma lain.

DAFTAR PUSTAKA

- [1] H. Suciati, "Aplikasi Berbasis Sistem Pakar Untuk Memprediksi Ketepatan Masa Studi Dengan Pendekatan Case Based Reasoning dan C45," pp. 1-11, 2016.
 - [2] Universitas Mulawarman, "Informasi Publik," 2017. [Online]. Available: <http://unmul.ac.id/page/unmul-dalam-angka-1487836584.html>.
 - [3] H. S. Pakpahan, M. Taruk, H. J. Setyadi, P. P. Widagdo, Havaluddin and L. Hendrajaya, "Prediksi Penyebaran Hidrokarbon Menggunakan Artificial Neural Network (ANN) di Formasi Gumai, Jambi," Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi, pp. 200-203, 2017.
 - [4] E. Prasetyo, Data Mining dan Aplikasi Menggunakan Matlab, Yogyakarta, 2012.
 - [5] W. B. Komalasari, "Metode Pohon Keputusan Untuk Eksploratori Data dengan Peubah yang Banyak dan Kompleks," Informatika Pertanian Vol. 16 No. 1, pp. 967-980, 2007.
 - [6] S. Susanto and D. Suryadi, Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data, Yogyakarta, 2010.
 - [7] D. Darsono, "Penggunaan Algoritma Clasification And Regression Tree (Cart) Untuk Mendukung Keputusan Pada Penilaian Sertifikasi Guru," pp. 66-72, 2010.
 - [8] D. T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, 2005.
 - [9] W. Indrawan, "Komparasi Algoritma Naive Bayes Classifier Dan Tree C4.5 Untuk Evaluasi Kinerja Akademik Mahasiswa," 2017.
 - [10] M. F. Amin, "Penerapan Algoritma Cart Untuk Memprediksi Status Kelulusan Mahasiswa," JUTISI Vol. 5 No. 3, pp. 1215-1222, 2016.
 - [11] P. Priati, "Kajian Perbandingan Teknik Klasifikasi Algoritma C4.5, Naive Bayes Dan Cart Untuk Prediksi Kelulusan Mahasiswa (Studi Kasus : Stmik Rosma Karawang)," Media Informatika Vol. 15 No. 2, pp. 1-17, 2016.
-