

Penerapan C4.5 pada Keaktifan Mahasiswa dalam Pengumpulan Berkas di Biro Akademik

¹Eka Irawan, ²Indra Gunawan
STIKOM Tunas Bangsa

Pematangsiantar, Sumatera Utara - Indonesia

eka.irawan@amiktunasbangsa.ac.id, indragunawan@amiktunasbangsa.ac.id

Abstract— Keberhasilan dan kegagalan mahasiswa dalam menyelesaikan perkuliahan merupakan gambaran dari suatu perguruan tinggi. Mahasiswa aktif adalah mahasiswa yang melakukan pendaftaran ulang pada awal semester dan memberikan hasil registrasi berbentuk foto copy di pendidikan sebagai syarat keaktifan mahasiswa di lihat dari berkas yang telah di kumpulkan. Selain mahasiswa aktif ada juga mahasiswa non-aktif ini sangat berpengaruh terhadap keberlangsungan perguruan tinggi dan mahasiswa nonaktif dapat menyebabkan mahasiswa *drop out*. Hal ini dapat menjadi lebih menarik jika di lakukan penelitian dengan teknik *data mining* metode C.4.5 untuk mencari karakteristik mahasiswa aktif berdasarkan berkas yang di kumpulkan ke Biro Akademik. Data penelitian menggunakan data mahasiswa Semester V (lima) S1 STIKOM TUNAS Bangsa Pematangsiantar. Hasil menunjukkan nilai *accuracy* yang mencapai 97,28 % dan *recall* 99,27%.

Kata Kunci— Mahasiswa non-aktif, C4.5, data mining

I. PENDAHULUAN

Jumlah mahasiswa yang waktu studinya tepat waktu pada suatu perguruan tinggi merupakan salah satu penentu kualitas perguruan tinggi. Berdasarkan matriks penilaian instrument akreditasi program studi Badan Akreditasi Nasional Perguruan Tinggi [1]. Menentukan keaktifan mahasiswa berdasarkan pengumpulan berkas di Tata usaha merupakan salah satu kewajiban mahasiswa mulai dari data pendaftaran sampai dengan registrasi mahasiswa persemesternya agar di serahkan ke bagian tata usaha untuk di data. [2] Pendataan ini di lakukan sebagai syarat untuk naik semester, dalam kondisi tertentu terdapat beberapa mahasiswa yang tidak/belum membuat KRS selama 6 semester, mahasiswa selesai/lulus tidak tepat waktu tidak sama dengan mahasiswa drop out. Tapi mahasiswa yg lulus tidak sesuai dengan waktu yang di tentukan maka berpotensi menjadi DO. Serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data disebut dengan data mining [3]. Salah satu metode data mining adalah pengklasifikasian data. Klasifikasi merupakan kegiatan dalam mengekstrak dan kemudian memprediksi label kategori untuk masing-masing data. Dengan kata lain, klasifikasi adalah penyusunan bersistem dalam kelompok atau

golongan menurut kaidah atau standar yang ditetapkan proses. Salah satu algoritma klasifikasi yang bisa digunakan adalah C4.5 [4]. Penelitian yang dilakukan oleh Gerben W. Dekker [5] menyebutkan bahwa monitoring dan dukungan terhadap mahasiswa di tahun pertama sangat penting dilakukan. Mahasiswa jurusan teknik elektro Universitas Eindhoven yang berhenti studi pada tahun pertama mencapai hingga 40%. Kurikulum yang sulit dianggap sebagai salah satu penyebab tingginya jumlah mahasiswa drop out. Permasalahan ini dapat di lakukan untuk mengatasinya dengan membuat suatu aplikasi data mining yang akan digunakan untuk menentukan mahasiswa yang aktif ataupun keaktifan mahasiswa dengan menggunakan atribut yaitu Fotocopy Kartu Keluarga, Fotocopy KTP, Fotocopy KTM, KRS Semester I – Semester V, IPS Semester I – Semester V maka penulis menggunakan model pohon keputusan algoritma C4.5. Model ini dipilih karena proses learning dan klasifikasi pada algoritma C4.5 sederhana dan cepat [4]. Secara umum, model algoritma C4.5 mempunyai tingkat akurasi yang tinggi.

Menggunakan algoritma C4.5 untuk melakukan klasifikasi dalam menentukan keaktifn Mahasiswa adalah sebuah metode yang baik untuk

melakukannya, maka penulis mengambil judul: “Penerapan C4.5 pada Keaktifan Mahasiswa dalam Pengumpulan Berkas di Biro Akademik”.

II. METODE PENELITIAN

Metode data mining c4.5 merupakan metode klasifikasi yaitu membangun suatu model yang bisa mengklasifikasikan suatu objek berdasar atribut-atributnya. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 [4], yaitu :

- 1) Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
- 2) Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy. Untuk menghitung nilai entropy digunakan rumus:

$$Entropy(A) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

Keterangan dari S: himpunan kasus, k: jumlah partisi S dan pi: probabilitas yang didapat dari jumlah (ya/tidak) dibagi total kasus

- 3) Kemudian hitung nilai gain menggunakan rumus:

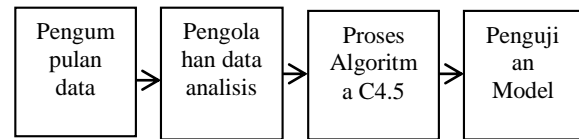
$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan dari S: himpunan kasus, A: fitur, n: jumlah partisi atribut A, |Si|: proporsi Si terhadap S dan |S|: jumlah kasus dalam S

- 4) Ulangi langkah ke-2 hingga semua record terpartisi.
- 5) Proses partisi pohon keputusan akan berhenti saat :
 - a. Semua record dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam record yang dipartisi lagi. Tidak ada record di dalam cabang yang kosong

Untuk melakukan sebuah penilaian dari data yang telah di peroleh maka penulis akan melakukan proses perhitungan dengan menggunakan algoritma klasifikasi C4.5 untuk mendapatkan faktor terbaik dalam menentukan keaktifan mahasiswa yang dapat dari Biro Akademik. Dari faktor tersebut maka akan di tentukan mahasiswa itu aktif atau tidak

berdasarkan berkas yang ada atau yang terkumpul. Secara detail metodologi penelitian ini dirancang seperti diagram blok yang terlihat dalam gambar 2.1.



Gambar 1. Diagram Blok Penelitian

Untuk proses penjelasan dari diagram blok penelitian diatas adalah sebagai berikut:

a. Pengumpulan Data

Jumlah data yang akan di uji adalah sebanyak 40 Mahasiswa yang terdiri dari satu kelas, yang mana Sampel data demografi mahasiswa yaitu Fotocopy Kartu Keluarga, Fotocopy KTP, Fotocopy KTM, KRS Semester I – Semester V, IPS Semester I – Semester V dalam penelitian ini data yang digunakan adalah mahasiswa semester 6.

b. Pengolahan data Analisis

Data mahasiswa yang terdiri atas data akademik dan data demografi yang diperoleh diintegrasikan sebagai satu kesatuan data, kemudian Seleksi berdasarkan faktor pendukung digunakan sebagai input untuk proses klasifikasi yang dilakukan dengan mengambil sebagian variabel pada seluruh atribut yang ada pada data untuk dijadikan atribut penentu dalam melakukan pemberian keputusan. Dan Data cleansing Pada tahap ini, dilakukan penghapusan data yang tidak lengkap

c. Proses Algoritma C4.5

Melakukan penghitungan atau penilaian terhadap data analisis yang telah di dapat dan kemudian akan di lakukan proses perhitungan untuk menentukan keaktifan mahasiswa

d. Pengujian Model

Hasil dari perhitungan maka akan di peroleh hasil terbaik untuk menentukan keaktifan mahasiswa yang mungkin nantinya di kembangkan dari hasil ini akan dilakukan untuk melakukan pengujian data. Presisi adalah persentase data yang diklasifikasikan sebagai model baik yang sebenarnya juga baik. Recall adalah pengukuran tingkat pengenalan positif sebenarnya[6]

Jenis penelitian ini adalah Penelitian *Kuantitatif*, karena menggunakan pengukuran data dan statistik objektif melalui perhitungan ilmiah berasal dari sampel masyarakat dengan memberikan jawaban atas sejumlah pertanyaan yang dituangkan ke dalam kuesioner menggunakan metode pendekatan *Algoritma C4.5*

III. HASIL DAN PEMBAHASAN

Pada bagian ini akan dijelaskan mengenai tahapan penelitian untuk memperoleh hasil klasifikasi mahasiswa yang aktif dan tidak aktif menggunakan metode C4.5. adapun tahapan tersebut adalah tahap persiapan data set, penentuan akar, dan penentuan rule. dan klasifikasi. Data set yang digunakan adalah data mahasiswa STIKOM Tunas Bangsa Prodi Sistem Informasi dari BAAK STIKOM Tunas Bangsa tahun akademik 2016/2017. Data set yang digunakan dalam penelitian ini dikomposisikan sebagai berikut:

Tabel 1. Kelompok Dataset

Kelas Target	Data yang digunakan		Jumlah
	Data Training 50 %	Data Testing 50%	
Aktif	74	70	174
Tidak Aktif	74	70	174

A. Analisis Data

Selanjutnya data disiapkan dengan melakukan seleksi dan transformasi data set (lampiran a).Seleksi data bertujuan untuk mengidentifikasi variable-variabel relevan yang digunakan dalam penelitian ini. Dan transformasi data digunakan untuk mengubah dataset sehingga konten informasi terbaik diambil dan dimasukkan pada tool mining yang digunakan dalam format yang tepat. Tahapan berikutnya yang dilakukan pada penelitian ini yaitu mengikuti cara kerja metode C4.5 mulai dari training data, menghitung entropi hingga menghasilkan rule yang dijadikan sebagai pola untuk Menentukan keaktifan mahasiswa. Atribut adalah bagian data, yang mewakili karakteristik atau fitur dari objek data. Atribut, dimensi, fitur dan variabel sering digunakan secara bergantian dalam literatur.

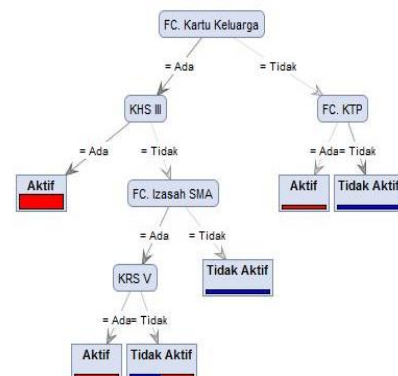
B. Proses algoritma C4.5

Proses algoritma C4.5 dilakukan dengan mengubah beberapa nilai atribut yang awalnya bernilai nominal menjadi nilai-nilai atribut yang sesuai dengan data normalisasi data. Entropy adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari impurity dan homogeneity dari kumpulan data. Dari nilai Entropy tersebut kemudian dihitung nilai information gain masing-masing atribut[7]. Penghitungan nilai Entropy digunakan rumus seperti dalam Persamaan (1).

Tabel 2. Entropy dan Information Gain dari semua atribut sebagai acuan pembuatan pohon keputusan dengan pemilihan atribut/variabel dengan information gain terbesar sebagai simpul akar.

Node		Jlh Kasus	Aktif (S1)	Tidak Aktif (S2)	Entropy	Gain
I	FC. Kartu Keluarga	147	137	10	0,358518	0,124606
	Ada	138	134	4	0,189278	
	Tidak	9	3	6	0,918296	
	FC. KTP					0,052613
	Ada	131	126	5	0,233828	
	Tidak	16	11	5	0,896038	
	FC. KTM					0,017987
	Ada	133	126	7	0,297472	
	Tidak	14	11	3	0,749595	
	FC. Izasah SMA					0,046534
	Ada	129	124	5	0,236576	
	Tidak	18	13	5	0,852405	
	KRS I					0,004523
	Ada	132	124	8	0,329846	
	Tidak	15	13	2	0,56651	
	KRS II					0,079874
	Ada	137	132	5	0,225991	
	Tidak	10	5	5	1	
	KRS III					0,183516
	Ada	129	125	4	0,199421	
	Tidak	18	12	6	0,918296	
	KRS IV					0,105416
	Ada	130	127	3	0,158384	
	Tidak	17	10	7	0,977418	
	KRS V					0,160532
	Ada	126	125	1	0,06678	
	Tidak	21	12	9	0,985228	
	KHS I					0,096194
	Ada	128	125	3	0,160329	
	Tidak	19	12	7	0,949452	
	KHS II					0,135901
	Ada	129	127	2	0,11539	
	Tidak	18	10	8	0,991076	
	KHS III					0,155089
	Ada	132	130	2	0,113274	
	Tidak	15	7	8	0,996792	
	KHS IV					
	Ada	128	128	0	0,118296	0,160532
	Tidak	19	9	10	0,998001	
	KHS V					0,096194
	Ada	132	132	0	0,013274	
	Tidak	15	5	10	0,918296	

Dari hasil Perhitungan entropy dan information gain yang didapat kemudian diolah kedalam Decision Tree. Berikut hasil Pohon Keputusan (Decision Tree)



Gambar 2. Pohon Keputusan

Dari pohon keputusan (*decision tree*) yang terbentuk pada Gambar 4.1 diatas, didapat aturan—aturan/*rule model* dalam penentuan rekomendasi penerimaan mitra penjualan. Ada 6 aturan yang terbentuk, dapat dilihat sebagai berikut:

if KHS V = Ada then Aktif (0 / 132)
if KHS IV = Tidak and FC. Kartu Keluarga = Tidak then Tidak Aktif (6 / 0)
if KHS IV = Ada then Aktif (0 / 4)
if FC. Izasah SMA = Tidak then Tidak Aktif (2 / 0)
if FC. KTP = Tidak then Tidak Aktif (1 / 0)
if FC. Kartu Keluarga = Ada then Tidak Aktif (1 / 1)

C. Validasi dan Analisa Pengujian

Pengujian dilakukan dengan validasi silang. Salah satu jenis validasi silang adalah *ten-fold cross validation*. Berikut hasil dari *ten-fold cross validation* :

Performance Vector:

Accuracy : 97.28%

ConfusionMatrix:

True	:	Tidak Aktif	Aktif
Tidak Aktif	:	7	1
Aktif	:	3	136

Precision : 97.84% (positive class: Aktif)

ConfusionMatrix:

True	:	Tidak Aktif	Aktif
Tidak Aktif	:	7	1
Aktif	:	3	136

Recall : 99.27% (positive class: Aktif)

ConfusionMatrix:

True	:	Tidak Aktif	Aktif
Tidak Aktif	:	7	1
Aktif	:	3	136

AUC (optimistic): 0.998 (positive class: Aktif)

AUC : 0.700 (positive class: Aktif)

AUC (pessimistic): 0.699 (positive class: Aktif)

Berdasarkan pengujian dengan menggunakan metode *Ten-Fold Cross Validation* menghasilkan nilai akurasi yang hampir sempurna sebesar 97,28 %, presisi sebesar 97,84 % dan recall sebesar 99,27%. Hal ini menunjukkan bahwa, dari proses klasifikasi yang dilakukan akan dapat diterapkan pada menentukan keaktifan mahasiswa diambil dari mahasiswa yang mengumpulkan berkas di BAIK.

IV. KESIMPULAN

Kesimpulan dari penelitian yang dilakukan adalah sebagai berikut:

- 1) Penerapan algoritma klasifikasi C4.5 akan dapat diimplementasikan pada menentukan keaktifan mahasiswa, dilihat dari tingkat accuracy yang mencapai 97,28 % dan recall 99,27%., yang menyatakan bahwa perhitungan yang dilakukan akan mampu memprediksi dan menentukan keaktifan mahasiswa
- 2) Terbentuk 6 aturan-aturan/*rule model* dapat digunakan sebagai acuan dalam perancangan dan pembuatan aplikasi GUI.
- 3) Dengan algoritma ini bisa digunakan untuk menentukan keaktifan mahasiswa berdasarkan kumpulan berkas di biro akademik

V. REFERENCES

- [1] Buku VI Matriks Penilaian Instrumen Akreditasi Program Studi Badan Akreditasi Nasional Perguruan Tinggi,
- [2] Asmaul Husnah Nasrullah” Penerapan Metode C4.5 Untuk Klasifikasi Mahasiswa
- [3] Berpotensi Drop Out” ILKOM Jurnal Ilmiah Volume 10 Nomor 2 Agustus 2018
- [4] Ian H. Witten, Frank Eibe, and Mark A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed., Asma Stephan and Burlington, Eds. United States of America: Morgan Kaufmann,
- [5] Kusriani, “Konsep dan Aplikasi”, Yogyakarta, Andi, 2009.
- [6] Gerben W. Dekker, "Predicting Students Drop Out: A Case Study," In International Conference on Educational Data Mining, Cordoba, Spain, 41-50,
- [7] Y. Altujjar, W. Altamimi, I. Al-turaiki, and M. Al-razgan, “Predicting Critical Courses Affecting Students Performance : A Case Study,” *Procedia - Procedia Comput. Sci.*, vol. 82, no. March, pp. 65–71, 2016
- [8] K. Adhatrao, A. Gaykar, D. Amiraj, R. Jha, and Honroa Vipul, “Predicting Students’ Performance Using Id3 And C4.5 Classification Algorithms,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 3, no. 5, pp. 39–52, 2013