

PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI PREDIKAT KEBERHASILAN MAHASISWA DI AMIK TUNAS BANGSA

Yuni Sara Luvia¹, Dedy Hartama², Agus Perdana Windarto³, Solikhun⁴

¹ Mahasiswa S-1 Sistem Informasi, STIKOM Tunas Bangsa

^{2,3,4} Dosen S-1 Sistem Informasi, STIKOM Tunas Bangsa

Jalan Sudirman Blok A No. 1, 2, 3 Kota Pematangsiantar

E-mail : yuninatani@gmail.com dedyhartama@amiktunasbangsa.ac.id zhantura.gusti@gmail.com

solikhun@amiktunasbangsa.ac.id

Abstrak

Predikat keberhasilan Mahasiswa di perguruan tinggi merupakan hal utama yang menjadi acuan bagi seorang mahasiswa selama menjalani study. Data mahasiswa yang terus meningkat dari tahun ke tahun adalah bukti untuk mengembangkan Data Mining. Banyaknya jumlah Mahasiswa akan diproses Untuk mengetahui berdasarkan kriteria apa saja Mahasiswa layak mendapatkan predikat keberhasilannya dengan beberapa atribut seperti, jenis kelamin, kehadiran, sesi perkuliahan, rerata NEM dan Asal sekolah. Dalam hal ini, peneliti melakukan analisis di Amik Tunas Bangsa Jl. Sudirman no. 1,2,3 Pematangsiantar. Analisis yang digunakan adalah menggunakan Data Mining dengan metode C4.5 dan diproses menggunakan software Rapidminer untuk membuat pohon keputusan. Metode ini diharapkan dapat mempermudah pihak instansi mengklasifikasi dan memprediksi predikat keberhasilan Mahasiswa sehingga data-data yang menumpuk bisa bermanfaat untuk keperluan data mining dalam mengambil keputusan yang baik. Analisis ini diharapkan sebagai salah satu motivasi terhadap mahasiswa untuk meningkatkan IPK.

Kata kunci : Keberhasilan mahasiswa; metode C4.5; data mining; rapidminer; decision tree

Pendahuluan

Predikat keberhasilan adalah kemampuan aktual seorang mahasiswa yang berwujud penguasaan ilmu pengetahuan, sikap, dan keterampilan untuk mencapai tujuan akhir dari proses belajar yaitu Indeks Prestasi. Indeks Prestasi dihitung, baik pada setiap akhir semester dengan hasil yang disebut IP semester, maupun pada akhir program pendidikan lengkap satu jenjang, dengan hasil yang disebut IP kumulatif. Indeks Prestasi Mahasiswa merupakan data yang sangat penting di Amik Tunas Bangsa Pematangsiantar sehubungan dengan terus meningkatnya jumlah Mahasiswa dari tahun ke tahun maka perlu dilakukan Data Mining dengan teknik klasifikasi bertujuan membantu suatu lembaga yang memiliki

data melimpah untuk menyajikan informasi kedalam bentuk pengambilan keputusan.

Banyak teknik klasifikasi di Data Mining salah satunya adalah Algoritma C4.5 yang menyajikan klasifikasi data kedalam bentuk pohon keputusan. Kelebihan utama Algoritma C4.5 dapat membuat pohon keputusan (*decision tree*) yang efisien menangani atribut tipe diskrit dan tipe diskrit- numerik, mudah untuk diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima (Han 2001). Kelemahan Algoritma C4.5 salah satunya terdapat di skalabilitas yaitu *data training* hanya dapat digunakan dan disimpan secara keseluruhan pada waktu yang bersamaan di memori (Veronica, 2007)

Amik Tunas Bangsa Pematangsiantar adalah lembaga pendidikan yang

memiliki banyak mahasiswa. Dengan Data Mining memudahkan lembaga pendidikan mengidentifikasi berdasarkan faktor-faktor apa sajakah yang mempengaruhi mahasiswa mendapatkan Indeks Prestasi Kumulatif baik di kelas. Peneliti menggunakan Algoritma C4.5 dalam pengklasifikasian predikat keberhasilan mahasiswa berdasarkan Atribut jenis kelamin, NEM, sesi perkuliahan, asal sekolah, kehadiran, dan IPK. Data Training yang diambil peneliti adalah data alumni mahasiswa program studi Manajemen Informatika dan Komputerisasi Akuntansi 2010-2012. Dengan demikian proses pengolahan data-data akademik menggunakan teknik klasifikasi akan membantu lembaga pendidikan dalam upaya memotivasi mahasiswa untuk meningkatkan IPK, serta bermanfaat meningkatkan akreditasi program studi di Amik Tunas Bangsa.

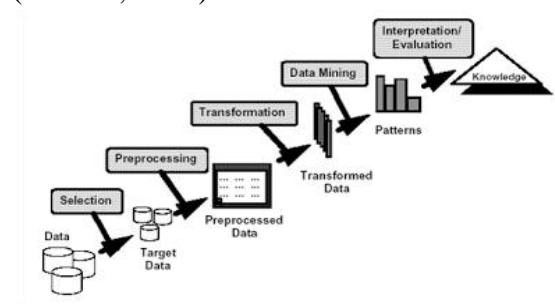
Berdasarkan latar belakang tersebut diterapkan Algoritma C4.5 yang berguna untuk mengklasifikasi dan memprediksi Predikat Keberhasilan Mahasiswa di Amik Tunas Bangsa Pematangsiantar.

Kajian Teori

Data mining merupakan proses iteratif dan interaktif untuk menemukan pola atau model baru yang dapat digeneralisasi untuk masa yang akan datang, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (*massive databases*). Data mining berisi pencarian *trend* atau pola yang diinginkan dalam *database* besar untuk membantu pengambilan keputusan di waktu yang akan datang. (Hermawati, 2013).

Istilah data mining dan knowledge discovery in databases (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD

adalah *data mining*. Proses KDD (*Knowledge discovery in databases*) secara garis besar dapat dijelaskan sebagai berikut (Kusrini, 2009)



Gambar 1. Tahapan *knowledge discovery in databases*

Pengolahan Data Mining terdiri dari beberapa metode pengolahan, yaitu (Larose, 2006):

1. *Association* (Asosiasi) merupakan teknik dalam data mining yang mempelajari hubungan antar data. Contoh penggunaannya seperti untuk menganalisis perilaku mahasiswa yang datang terlambat. Contohnya jika mahasiswa memiliki jadwal dengan dosen A dan B, maka mahasiswa akan datang terlambat.
2. *Predictive modelling* yang merupakan pengolahan data mining dengan melakukan prediksi/peramalan. Tujuan metode ini untuk membangun model prediksi suatu nilai yang mempunyai ciri-ciri tertentu.
3. *Classification* merupakan teknik mengklasifikasikan data. Perbedaannya dengan metode *clustering* terletak pada data, dimana pada *clustering* variabel dependen tidak ada, sedangkan pada *classification* diharuskan ada variabel dependen.
4. *Clustering* merupakan pengelompokan terhadap *record*, yang berfungsi memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan

aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami, juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

- a. Pilih atribut sebagai akar
- b. Buat cabang untuk masing-masing nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera berikut (Craw, 2005):

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

Keterangan :

- S : Himpunan Kasus
- A : Atribut
- N : Jumlah partisi atribut A
- |Si| : Jumlah kasus pada partisi ke i
- |S| : Jumlah kasus dalam S

Sedangkan perhitungan nilai *Entropy* dapat dilihat pada rumus 2 berikut (Craw, 2005):

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

Keterangan:

- S : Himpunan Kasus
- A : Fitur
- N : Jumlah partisi S
- Pi : Proporsi dari Si terhadap S

Metode Penelitian

Metode penelitian ini dilakukan secara sistematis agar mendapatkan alur kerja yang baik, dan dijabarkan menjadi beberapa langkah, yaitu:

1. Studi literatur dilakukan untuk membantu peneliti menelusuri teori-teori yang sedang berkembang mengenai data mining dan Algoritma C4.5 untuk memperoleh metode yang signifikan dengan permasalahan yang ada.
2. Pemilihan objek penelitian dilakukan untuk memudahkan peneliti dalam

pengklasifikasian predikat keberhasilan mahasiswa di Amik Tunas Bangsa, mengingat terus meningkatnya jumlah mahasiswa tiap tahun agar data tersebut menjadi bermanfaat.

3. Menentukan variabel yang akan diproses, untuk mengerjakan pola pengelompokan peneliti memilih predikat keberhasilan mahasiswa sebagai variabel (A) dimana nilainya tergantung (dependen) dengan nilai lainnya, selanjutnya variabel (B) yang tidak memiliki ketergantungan terhadap nilai-nilai variabel lainnya (independen). Variabel (B) terdiri dari:
 - a. B1 : Jenis Kelamin, yang dipertimbangkan sebagai faktor kecerdasan dan kerajinan seseorang sehingga mempengaruhi mahasiswa mendapatkan IPK yang bagus.
 - b. B2 : NEM, mahasiswa yang memiliki NEM bagus sewaktu menyelesaikan studi di SMA Sederajat memiliki pengaruh mendapatkan IPK yang bagus.
 - c. B3: Sesi Perkuliahan, di Amik Tunas Bangsa memiliki 3 sesi perkuliahan yaitu, sesi pagi, siang dan malam. Variabel ini dipilih menjadi salah satu faktor predikat keberhasilan mahasiswa dikarenakan peneliti ingin melihat apakah mahasiswa yang memilih belajar pada sesi pagi termasuk yang dapat IPK bagus, ataupun yang belajar di sesi siang, ataupun yang memilih sesi malam. Karena mayoritas mahasiswa yang memilih sesi malam masing-masing sudah mempunyai pekerjaan dan pasti semangat atau kedisiplinannya berkurang karena terbatasnya waktu yang dimiliki untuk belajar.
 - d. B4 : Asal Sekolah, atribut ini dipilih karena akan menjadi pertimbangan penentuan daerah-daerah strategis yang cocok untuk kegiatan promosi mencari bibit terbaik sebagai calon mahasiswa di Amik Tunas Bangsa.

- e. B5 : Kehadiran, Mahasiswa yang memiliki kehadiran 100% di semua mata kuliah akan menjadi pertimbangan bahwa mahasiswa tersebut memiliki pengaruh mendapatkan predikat keberhasilan yang bagus.
4. Menentukan nilai *class* pada masing-masing variabel (A) dan (B). Nilai *class* variabel (A) disesuaikan pada Amik Tunas Bangsa yang dibedakan menjadi 5 nilai dengan tipe *label*, Jika $A1 \geq 3.50$ adalah *cumlaude*, jika $2.99 \leq A1 < 3.50$ adalah sangat baik, jika $2.5 \leq A1 < 2.99$ adalah baik, jika $1.9 \leq A1 < 2.5$ adalah cukup, sedangkan jika $0 \leq A1 < 1.9$ adalah kurang. Untuk nilai *class* pada variabel (B) dapat dilihat dalam tabel dibawah ini :

Tabel 1. Jenis Kelas Data

Varia-bel	Nama Field	Jenis Kelas Data	Kelas Data yang Digunakan
A	Predikat Keberhasilan	Nominal	Cumlaude, Sangat Baik, Baik, Cukup, Kurang
B1	Jenis Kelamin	Nominal	Pria, Wanita
B2	Rerata NEM	Nominal	$NEM \leq 20$, $NEM > 20$
B3	Sesi Perkuliahan	Nominal	Pagi, Siang, Malam
B4	Asal Sekolah	Nominal	Pematangsiantar, Luar
B5	Kehadiran	Nominal	Kehadiran < 50, Kehadiran > 50

5. Pengumpulan data dengan metode *Slovin* untuk menentukan jumlah sampel pada pengujian proses data mining, adapun rumus *slovin* nya sebagai berikut:

$$n = \frac{N}{1 + Ne^2}$$

Keterangan:

n: jumlah sampel

N: jumlah populasi

e: batas toleransi kesalahan 5%

6. Menganalisis data untuk mengklasifikasi predikat keberhasilan mahasiswa di Amik Tunas Bangsa sesuai dengan entropi berdasarkan masing-masing atribut menggunakan metode C.45.

Hasil dan Pembahasan

Peneliti memperoleh potongan data mahasiswa Amik Tunas Bangsa yang tercatat dalam format excel yang akan diolah untuk mendapatkan informasi. Adapun tabel data mahasiswa, adalah sebagai berikut:

NIM	JENKEL	SESI	KEHADIRAN	NEM	ASAL SEKOLAH	IPK
201001030050	Wanita	Pagi	100	48,5	SMA Taman Siswa Pematangsiantar	3,5
201001030052	Wanita	Pagi	100	48,3	SMA Taman Siswa Pematangsiantar	3,1
201001030358	Pria	Malam	70	49,95	SMK UISU Pematangsiantar	3,09
201001030360	Wanita	Malam	86	46,45	SMK Harapan P.Siantar	3
201101030077	Wanita	Pagi	100	47,25	SMA AL-Wasliyah Perdagangan	3,89
201001030071	Wanita	Pagi	100	35,36	SMA AL-Wasliyah Perdagangan	3,49
201001030056	Wanita	Pagi	40	50,35	SMA YP. Teladan Pematangsiantar	2,88
201101030220	Wanita	Pagi	45	18,56	SMA N I Dolok Batunanggar	2,92
201001030235	Pria	Siang	90	49,95	SMA Pelita Pematangsiantar	3,09
201101030211	Pria	Pagi	100	31,92	SMK GKPS P.Siantar	3,66
201001030227	Wanita	Siang	100	19,94	SMA Al-Kautsar Karang Anom	3,45

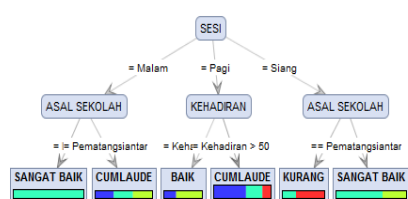
Gambar 2. Data mahasiswa

Adapun hasil *preprocessing* data mahasiswa yang di dapat adalah sebagai berikut:

JENKEL	SESI	KEHADIRAN	NEM	ASAL SEKOLAH	IPK
Wanita	Pagi	Kehadiran > 50	NEM > 20	Pematangsiantar	CUMLAUDE
Wanita	Pagi	Kehadiran > 50	NEM > 20	Pematangsiantar	SANGAT BAIK
Pria	Malam	Kehadiran > 50	NEM > 20	Pematangsiantar	SANGAT BAIK
Wanita	Malam	Kehadiran > 50	NEM > 20	Luar	SANGAT BAIK
Wanita	Pagi	Kehadiran > 50	NEM > 20	Luar	CUMLAUDE
Wanita	Pagi	Kehadiran > 50	NEM > 20	Pematangsiantar	SANGAT BAIK
Wanita	Pagi	Kehadiran < 50	NEM > 20	Pematangsiantar	BAIK
Wanita	Pagi	Kehadiran < 50	NEM < 20	Luar	BAIK
Pria	Siang	Kehadiran > 50	NEM > 20	Pematangsiantar	SANGAT BAIK
Pria	Pagi	Kehadiran > 50	NEM > 20	PematangSiantar	CUMLAUDE
Wanita	Siang	Kehadiran > 50	NEM < 20	Luar	SANGAT BAIK

Gambar 3. Hasil *preprocessing* data mahasiswa

Hasil *preprocessing* yang didapat kemudian diolah kedalam *Decision Tree* menggunakan *software Rapidminer* untuk mengetahui klasifikasi predikat keberhasilan Mahasiswa yang telah menyelesaikan *study*, adapun gambar pohon keputusan sebagai berikut :



Gambar 4. Pohon keputusan klasifikasi predikat keberhasilan

Jika dilihat berdasarkan hasil pohon keputusan klasifikasi predikat keberhasilan mahasiswa, bahwa atribut yang mempunyai pengaruh utama untuk mendapatkan predikat keberhasilan adalah variabel B3 (Sesi) yang menempati sebagai simpul akar. Agar lebih jelas, peneliti membuat model aturan berupa teks, seperti keterangan dibawah ini:

Sesi = Malam

| Asal Sekolah = Luar: Sangat Baik
{Cumlaude=0, Sangat Baik=4, Baik=0, Kurang=0}

| Asal Sekolah = Pematangsiantar:
Cumlaude {Cumlaude=1, Sangat Baik=1, Baik=1, Kurang=0}

Sesi = Pagi

| Kehadiran = kehadiran < 50: Baik
{Cumlaude=1, Sangat Baik=0, Baik=2, Kurang=0}

| Kehadiran = Kehadiran > 50: Cumlaude
{Cumlaude=4, Sangat Baik=2, Baik=0, Kurang=1}

Sesi = Siang

| Asal Sekolah = Luar: Kurang
{Cumlaude=0, Sangat Baik=1, Baik=0, Kurang=2}

| Asal Sekolah = Pematangsiantar: Sangat baik
{Cumlaude=0, Sangat Baik=2, Baik=1, Kurang=0}

Dapat dilihat dari hasil pohon keputusan dan model aturan berbentuk teks bahwa, jika seorang mahasiswa sesi perkuliahan pagi dan memiliki kehadiran > 50 maka mahasiswa tersebut diklasifikasi mendapat predikat keberhasilan *cumlaude*, tanpa melihat kriteria asal sekolah, jenis kelamin dan Nem.

Kesimpulan

Hasil penelitian yang diperoleh disimpulkan bahwa telah di dapat klasifikasi predikat keberhasilan mahasiswa di Amik Tunas Bangsa Pematangsiantar. Variabel yang memiliki prioritas utama terhadap predikat keberhasilan mahasiswa adalah mahasiswa yang memilih sesi perkuliahan pada Pagi hari dan di dukung dengan nilai kehadiran > 50 maka mahasiswa tersebut mendapatkan predikat *cumlaude* dibandingkan dengan mahasiswa yang berada di sesi perkuliahan siang dan malam. Pengaruh ini dapat dilihat dari besarnya semangat belajar mahasiswa pagi yang memiliki banyak waktu untuk diskusi diluar jam belajar sehingga kepedulian dan kedisiplinan mahasiswa tersebut berhak mendapatkan predikat keberhasilan *cumlaude*.

Daftar Pustaka

- Veronica Sri Moertini, 2007, "Pengembangan Skalabilitas Algoritma Klasifikasi C4.5 Dengan Pendekatan Konsep Operator Relasi, studi kasus: pra-pengolahan dan klasifikasi citra batik", Bandung.
- Han, J., Kamber, M., 2001, "Data Mining Concepts and Techniques", Morgan Kaufman Pub., USA.
- Kusrini dan Luthfi. E. Taufiq.(2009).Algoritma *Data Mining*. Yogyakarta: Andi.
- Kamagi, David Hartanto dan Seng Hansun. 2014. Implementasi Data mining dengan Algoritma C 4.5 untuk memprediksi tingkat kelulusan Mahasiswa. Jurnal Seminar Nasional Teknologi Informasi dan Multimedia 2014. ISSN: 2085-4552.
- Hartama, Dedy. 2011. Model aturan keterhubungan data mahasiswa menggunakan Algoritma C4.5 untuk meningkatkan indeks prestasi di Amik Tunas Bangsa. Tesis Program Studi Magister (S2) Teknik Informatika Universitas Sumatera Utara.