

MENENTUKAN MAHASISWA BERPRESTASI DENGAN MENGGUNAKAN ALGORITMA C4.5 : STUDI KASUS UNIVERSITAS MUHAMMADIYAH TANGERANG

R.Tommy Gumelar^{1 (*)}

¹ITB Ahmad Dahlan, Jakarta

Abstract

This research will discuss about the process of determining student achievement of case study Muhammadiyah Tangerang University. The number of students will make the length of time in use in the search process of outstanding students and not objektifnya in determining student achievement, if the process of determining students berprestai if still using the previous way that is using unstructured way. In this research will be made a decision support system in determining student achievement which where the process is more terstruktur by using: c4.5 algorithm to assist in the search process student achievement. As for the result of this research, the accuracy level is better and the time used is shorter than the previous way.

Kata Kunci: *Process, Achievement, Structured, C4.5algorithm, Accuracy*

Juli – Desember 2020, Vol 1 (1) : hlm 17-25
©2020 Institut Teknologi dan Bisnis Ahmad Dahlan.
All rights reserved.

(*) Korespondensi : tommy_gumelar@yahoo.com (R. Tommy Gumelar)

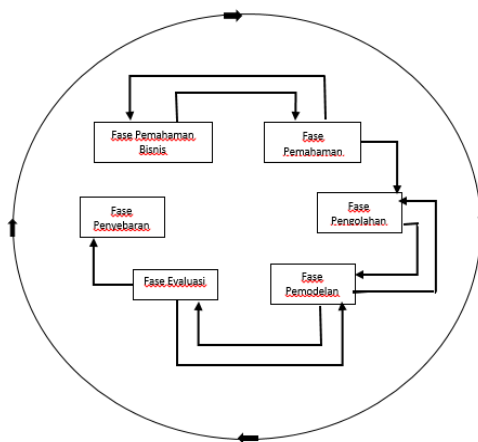
PENDAHULUAN

Universitas Muhammadiyah Tangerang adalah salah satu amal usaha milik Perserikatan Muhammadiyah. Seiring dengan jumlah mahasiswa yang banyak membuat pihak universitas Muhammadiyah Tangerang mengalami kesulitan dalam mencari mahasiswa berprestasi, apabila dilakukan secara manual dan tidak terstruktur. Analisa yang dipergunakan untuk melakukan pencarian mahasiswa berprestasi terjadi kesalahan seperti mahasiswa yang terpilih bukanlah mahasiswa nilai IPK tertinggi persemester dianggap berprestasi padahal masih ada mahasiswa yang memiliki IPK tertinggi tetapi tidak terkatagorikan sebagai mahasiswa berprestasi, di karnakan metode yang di terapkan dalam proses pencarian mahasiswa berprestasi belum terstruktur. Perlu adanya perubahan dalam analisis yang di pergunakan dalam pencarian mahasiswa berprestasi dari yang awalnya hanya dilakukan secara manual dan tidak terstruktur menjadi lebih terstruktur. Sangat diharapkan apabila metode yang diterapkan nantinya akan lebih mempercepat waktu dalam proses pencarian mahasiswa berprestasi pada universitas muhammadiyah tangerang,

2. Tinjauan Pustaka

2.1. Data Mining

Istilah Data mining biasa digunakan suatu istilah yang sering dipakaidalam kegiatan penelitian pengetahuan dalam proses pencarian sebuah database. Pengertian Data mining yaitu peroses pencarian informasi, indentifikasi database besar dengan menggunakan teknik-teknik tertentu seperti kecerdsaan buatan, matematika, statistik, maupun machine learning. Pada tahun 1996 dilakukan analisis mengenai sebuah proses standar Data Mining dari beberapa industry sebagai sample seperti Daimler Chrysler, dan NCR. CRISP DM menyediakan standar proses data mining menjadi strategi untuk pemecahan secara umum dalam sebuah bisnis atau unit penelitian. Menurut CRISP-DM data mining itu sendiri terdiri atas enam fase.



Gambar 1 Proses Data Mining Menurut CRIS-DMP

Enam fase tersebut menurut CRISP-DM yaitu:

1. Tahap Pemahaman Bisnis
 - a. Menentukan focus proyek serta kebutuhan detail dari sebuah bidang bisnis atau keinginan keseluruhan unit penelitian.
 - b. Menentukan tujuan serta batasan dari permasalahan bisnis dalam data mining.
 - c. merancang tahapan-tahapan awal agar dapat mencapai tujuan bisnis tersebut.
2. Tahap Pemahaman Data
 - a. Mengumpulkan data yang diperlukan.
 - b. Mengkaji, menyelidiki dan menganalisis data untuk memahami lebih baik data tersebut dan memahami tahapan awal.
 - c. Mengukur keakurasian data.
 - d. Pengelompokan-pengelompokan data yang dapat dikategorikan apakah dapat dikatakan menjadi sebuah permasalahan.
3. Tahap Pengolahan Data (Data Preparation Phase)
 - a. Mempersiapkan data pada tahapan awal, mengelompokkan data yang diperlukan guna untuk tahapan-tahapan berikutnya. Pada tahap awal ini menjadi tahapan berat agar pada tahapan selanjutnya dapat berjalan dengan baik.
 - b. Menentukan permasalahan dan unsur-unsur apa saja yang akan di analisa dan dapat bermanfaat pada penelitian yang sedang dilakukan.
 - c. Merombak beberapa tahapan apabila kurang sesuai dengan tujuan penelitian.
 - d. Mempersiapkan tahapan awal yang akan di perlukan dalam merancang tujuan akhir bisnis tersebut.
4. Tahap Pemodelan (Modeling Phase)
 - a. Pemilihan aplikasi yang sesuai dengan model yang akan dipergunakan.
 - b. Penentuan model yang akan dipergunakan untuk mendapatkan hasil yang maksimal.
 - c. Melihat teknik-teknik pemodelan dengan permasalahan pada Data Mining yang sama atau mendekati masalahnya.
 - d. Apabila dalam melaksanakan suatu tahapan apabila mengalami sebuah masalah diperbolehkan untuk kembali pada tahapan sebelumnya.
Untuk mendapatkan hasil yang sesuai dengan tujuan.
5. Tahap Evaluasi
 - a. Melakukan evaluasi satu ataupun beberapa model yang nantinya akan digunakan pada tahap pemodelan agar dapat menghasilkan kualitas dan efektivitas sebelum sampai kepada tahapan penyebaran untuk kemudian dipergunakan.
 - b. Memilah-milah apakah model yang akan dipergunakan sesuai kepada tujuan pada tahap awal.
 - c. Menentukan permasalahan penting yang ada dalam bisnis maupun penelitian yang tidak dapat tertangani dengan hasil memuaskan.
 - d. Menentukan hal-hal yang diperlukan dari hasil Data Mining.
6. Tahapan Penyebaran/ Deployment
 - a. Implementasi model-model yang memberikan hasil.

- b. contoh untuk tahap penyebaran/pendistribusian misalnya dalam perancangan laporan.
- c. Contoh spesifik pada tahap penyebaran yaitu Penerapan tahapan-tahapan data mining secara berurutan pada sub-bidang lain.

2.2. Kelompok Data Mining

Sebuah data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dilakukan, yaitu (Larose, 2005) :

1. Deskripsi

Melakukan pencarian untuk dapat menggambarkan model dan kemiripan dalam data. Pengertian dari model dari tingkat kecenderungan sering memberikan kemungkinan pengertian untuk memberikan data.

2. Perkiraan Hasil

Memperkirakan sama pada saat klasifikasi, berbedanya pada bagian memperkirakan tujuan lebih kepada numerik dibandingkan kepada kategori, Perancangan dibuat melalui record data-data lengkap dimiliki yang mempunyai nilai didalamnya dan unsur-unsur tujuan sebagai nilai perkiraan nantinya. Kemudian pada peninjauan berikutnya perkiraan nilai dari unsur-unsur dalam memperkirakan.

3. Prediksi

Prediksi tidak terlalu berbeda dengan klasifikasi dan juga perkiraan tujuan, hanya saja pada saat prediksi dalam nilai hasilnya dalam tahapan berikutnya. Dapat dimisalkan prediksi pada saat bisnis atau pada saat penelitian.

4. Pengklasifikasian

Klasifikasi ialah tahapan dalam pemilihan model maupun fungsi yang menjelaskan atau tahapan dalam membedakan konsep maupun data kelas, dalam tujuan untuk mendapat memprediksi kelas dalam suatu objek pada labelnya yang belum diketahui.

Tujuannya adalah unsur dari kategori, misalkan unsur yang menghasilkan, yang dapat dikategorikan kepada tiga kelas maupun unsur yang lebih tinggi, medium dan juga rendah. Tahapan dalam model Data mining pada saat membaca sebagian besar record dari kelompok record yang berisi informasi dalam sebuah tujuan akhir.

2.3. Algoritma C4.5

Metode Algoritma C4.5 ialah hasil pada pengembangan sebuah metode sebelumnya Algoritma ID3 yang dihasilkan oleh J. Ross Quinlan pada tahun 1986. Algoritma C4.5 itu sendiri merupakan algoritma yang dapat digunakan untuk membentuk pohon keputusan. Penggunaan pohon keputusan yaitu metode klasifikasi dan prediksi yang level akurasi dapat dikatakan sangat kuat dan sudah banyak yang mengakui. Model pada pohon keputusan dalam merubah fakta sangat besar dapat dirubah kedalam pohon keputusan dan menjelaskan aturan. Aturan tersebut yang dapat dipahami dengan lebih mudah serap dan diterima.

2.4. Metode yang diusulkan

Metode yang diusulkan untuk menentukan klasifikasi mahasiswa yang berprestasi adalah metode *Decision Tree* Algoritma C4.5 dengan menggunakan *software* Weka untuk pembuatan analisa dan pengujian model.

Tahapan Algoritma Decision Tree C4.5:

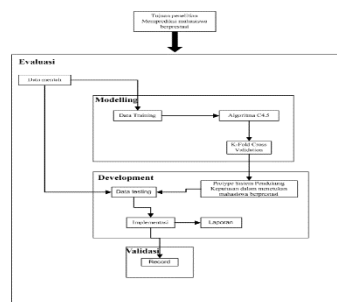
- 1) Menyiapkan data training
- 2) Menentukan akar dari pohon.
- 3) Hitung nilai Gain.

3. Rancangan Sistem Dan Aplikasi

Permasalahan dalam penelitian ini adalah pencarian mahasiswa berprestasi menggunakan cara tidak terstruktur yang berdampak lamanya waktu pengerjaan dan kurang akuratnya hasil yang di capai. Tujuan penelitian ini untuk membuat prediksi pencarian mahasiswa berprestasi menggunakan cara yang terstruktur

3.1. Kerangka Konsep

Analisa data dilakukan pada awal penelitian ini dan untuk memberikan kemudahan maka dibuat suatu kerangka pemikiran yang dapat dilihat pada gambar II dibawah ini:



Gambar II Alur penelitian

Berdasarkan data mahasiswa aktif sebagai *sample*, yaitu semakin besar jumlah *data* semakin menggambarkan jumlah mahasiswa aktif. Dalam sebuah data yang dipakai pada penelitian ini adalah data yang sebelumnya sudah dipakai, diambil 3 tahun terakhir data mahasiswa aktif tahun ajaran 2014, 2015-2016, dan 2016-2017. Dengan menggunakan rumus slovin untuk menentukan jumlah sample berdasarkan tingkat kesalahan 10%. Maka didapat perhitungan sebagai berikut:

Tabel 1 Pemilihan sample

Tahun	Laki-laki	Perempuan
2014-2015	100	211
2015-2016	124	169
2016-2017	160	170
Total	384	550
	934	

Perhitungan untuk menghasilkan nilai entropy dan gain tersebut sebagai berikut :

		jumlah	Lulus	tidak	Entropy	Gain
Total		100	96	4	0.2422	
P. Bisnis dan Finance						0
	A	90	90	0	0	
	B	10	10	0	0	
Tasawuf						0
	A	71	71	0	0	
	B	18	18	0	0	
	E	1	0	1	0	

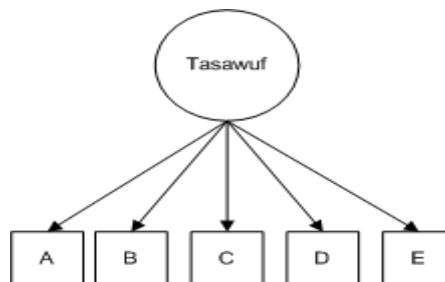
$$Entropy (s) = \sum_{i=1}^n -p_i * \log_2 p_i$$

$$Entropy (T) \left(-\frac{96}{100} * \log_2 \left(\frac{96}{100} \right) \right) + \left(-\frac{4}{100} * \log_2 \left(\frac{4}{100} \right) \right)$$

$$Entropy \text{ Total} = 0.2422$$

Tabel II Perhitungan menentukan Entropy

Dari table diatas dapat dilihat hasil pengclassifikasian memperlihatkan nilai entropy dan nilai Gain yang di peroleh. Setelah di dapat nilai tersebut dapat digambarkan pohon keputusan, seperti berikut:

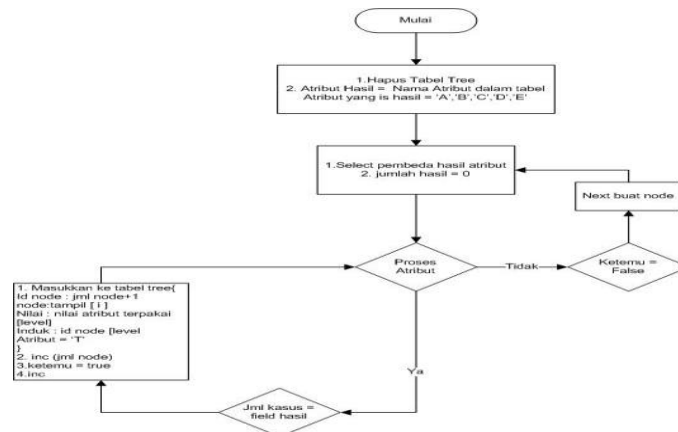


Gambar III Pohon keputusan Hasil penghitungan

Dapat di lihat Gambar III menggambarkan pohom keputusan dari nilai matakuliah, yang didapat dari hasil pengklassifikasian nilai yang di dapat dari tiap-tiap matakuliah.

3.2 Pembentukan Algoritma

Tahapan pembuatan algoritma dilakukan secara berulang-ulang dengan cara memanggil hasil yang sudah ada dan kemudian proses kembali. pada Iniliasiasi tahap awal pembentukan tree yang digambarkan sebagai berikut:



Gambar IV Algoritma C4.5

Gambar diatas memperlihatkan alur algoritma yang diperoleh dari proses pencarian mahasiswa berprestasi Universitas Muhammadiyah Tangerang

3.3 Weka

Berikut adalah hasil uji dengan menggunakan tools WEKA terhadap mahasiswa berprestasi dengan menggunakan algoritma C4.5. Dari 100 record data training yang digunakan untuk diuji dengan 20 atribut didapatkan hasil dengan tingkat akurasi 95%.

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      95          95 %
Incorrectly Classified Instances    5           5 %
Kappa statistic                    0.734
Mean absolute error                 0.0586
Root mean squared error             0.211
Relative absolute error             31.364 %
Root relative squared error         70.3147 %
Total Number of Instances          100
Ignored Class Unknown Instances     1

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.967	0.200	0.978	0.967	0.972	0.735	0.967	0.994	A
	0.800	0.033	0.727	0.800	0.762	0.735	0.964	0.809	B
Weighted Avg.	0.950	0.183	0.953	0.950	0.951	0.735	0.966	0.975	

```

=== Confusion Matrix ===
 a b  <-- classified as
 87 3 | a = A
 2 8 | b = B

```

Gambar IV Hasil Algoritma C 4.5 Weka

Dari hasil yang didapatkan menunjukkan tingkat akurasi yang sangat baik, atau dapat di katakana sangat mencukupi yaitu mencapai 95% dalam penerapan algoritma C4.5 dengan menggunakan bantuan software weka.

4. Pembuktian

Pada tahap implementasi ini Algoritma C4.5 yang dipeterapkan pada proses pencarian mahasiswa berprestasi Universitas Muhammadiyah Tangerang berjalan dengan baik, dan sesuai dengan harapan. Dengan tingkat akurasi sebesar 95% sangat

membuktikan penerapan algoritma C4.5 ini baik di terapkan pada proses pencarian mahasiswa berprestasi Universitas Muhammadiyah Tangerang. Dapat di simpulkan hasil yang di dapat dari penerapan Algoritma C4.5 ini:

1. Analisa yang tepat dan terstruktur dalam proses pencarian mahasiswa berprestasi Universitas Muhammadiyah Tangerang.
2. Waktu pengerjaan dalam proses pencarian mahasiswa berprestasi lebih singkat.

4.1 Implementasi Software

Pada tahap penerapan algoritma C4.5 penelitian ini menggunakan bantuan software weka yang menyediakan algoritma c4.5. Dengan memanfaatkan tools weka ini sangat membantu dari sudut pandang waktu dan kemudahan dalam segi penggunaannya.

5. Hasil dan Pembahasan

Hasil yang didapatkan dari uji coba dalam pencarian mahasiswa berprestasi pada Universitas Muhammadiyah Tangerang menggunakan Algoritma C4.5 menggunakan weka didapatkan hasil yang sangat baik yaitu tingkat keakurasiannya sebesar 95%.

6. Kesimpulan dan Saran

6.1 Kesimpulan

Berdasarkan dari hasil pengujian, prediksi menentukan mahasiswa berprestasi pada Universitas Muhammadiyah Tangerang di dapat hasil sebagai berikut:

1. Hasil yang di dapat dalam proses pencarian mahasiswa berprestasi pada Universitas Muhammadiyah Tangerang berjalan dengan baik dengan menggunakan Algoritma C4.5 tingkat akurasi mencapai 95% dengan error rate 5% dengan hasil itu dapat dikatakan mencapai harapan
2. Dari penerapan Algoritma C4.5 dalam peroses pencarian mahasiswa berprestasi Universitas Muhammadiyah Tangerang waktu yang dipergunakan lebih cepat di banding cara sebelumnya.

6.2 Saran

1. Perlu diadakan pelatihan yang dapat meningkatkan pengetahuan dalam penggunaan algoritma c4.5 pada software weka sehingga dapat mempermudah dan dapat mempersingkat waktu pengerjaan.
2. Pada sisi hardware perlu dilakukan upgrade lagi untuk memperoleh hasil yang lebih cepat dari implementasi prediksi mahasiswa berprestasi pada Universitas Muhammadiyah Tangerang .
3. Untuk pengembangan lebih lanjut dapat menambahkan jumlah sampel dan pada pengujian untuk lebih dilakukan berulang agar lebih mendapatkan hasil yang lebih akurat.

DAFTAR PUSTAKA

Williams, G. (2011) '*Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*',Book. doi: 10.1007/978-1-4419-9890-3.

- Kusrini.dan Emha Taufik Luthfi (2009) 'ALGORITMA DATA MINING', book Penerbit Andi, Yogyakarta, Indonesia
- Kamagi, D. H. and Hansun, S. (2014) 'Implementasi Data Mining dengan Algoritma C4 . 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa', *ULTIMATICS*, Vol. VI, No. 1 | Juni 2014, VI(1), pp. 15–20.
- Hijriana, N. (2008) 'Penerapan metode decision tree algoritma c4.5 untuk klasifikasi mahasiswa berprestasi', pp. 39–43.
- Oktavia, C. A., Budi, P. and Rahmadwati (2015) 'Analisis Kinerja Algoritma C4.5 Pada Sistem Pendukung Keputusan Penentuan Jenis Pelatihan', *Jurnal EECCIS*, 9(2), pp. 144–149.
- Belakang, L. and Keputusan, P. (2009) 'DECISION TREE (POHON KEPUTUSAN)', pp. 1–8.
- Rajeev Bedi Sunil Kumar Gupta, D. (2015) 'Review of Decision Tree Data Mining Algorithms: Id3 and C4.5', pp. 5–8.
- Diansano. 2013,' Metode-metode dalam Data Mining - Seri Data Mining for Business Intelligence', <https://beritati.blogspot.co.id/2013/08/metode-metode-dalam-data-mining-seri.html>
- Perbedaan algoritma ID3C4.5,C4.8,<https://codemath.wordpress.com/2011/06/20/perbedaan-algoritma-id3-c4-5-dan-j48>