

Implementasi Data Mining Untuk Menentukan Masa Studi Mahasiswa Universitas Nasional Dengan Menggunakan Algoritma C.45

Hadziroh Bittauhidil Fitri

Sistem informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional

Jalan Sawo Manila, Pasar Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12520

Ziroh.hf@gmail.com

Abstract - The rapid growth of data accumulation has created conditions for the rich data but still difficult to analyze the solution problem for decision making. One of the algorithms that can be applied in prediction modeling is C4.5. Using algorithm C4.5, the Study Program can predict the study period of the students based on existing case data. In addition, the information displayed in the form of support and confidence values of each category of graduation rate. This final report will describe the attributes used in classifying the student learning period and the model of the decision tree produced. Based on the evaluation of built software, the C4.5 algorithm is able to classify students into predefined classes with an error rate of 27.75% and an accuracy of 72.25%..

Keyword : Study Period, Data Mining, Algorithm C4.5

Abstrak - Pertumbuhan Pesat akumulasi data telah menciptakan kondisi untuk data yang kaya tetapi masih sulit untuk menganalisis masalah solusi untuk pengambilan keputusan. Salah satu algoritme yang dapat diterapkan dalam prediction modelling adalah C4.5. Dengan menggunakan algoritma C4.5, Program Studi dapat memprediksi masa studi mahasiswa berdasarkan data kasus yang sudah ada. Selain itu, informasi yang ditampilkan dalam bentuk dukungan dan nilai keyakinan masing-masing kategori tingkat kelulusan. Laporan akhir ini akan menggambarkan atribut yang digunakan dalam mengklasifikasikan periode belajar siswa dan model pohon keputusan yang dihasilkan. Berdasarkan hasil evaluasi terhadap perangkat lunak yang dibangun, algoritme C4.5 mampu mengklasifikasikan

mahasiswa ke dalam kelas-kelas yang telah ditentukan sebelumnya dengan nilai error 27.75% dan nilai akurasi 72,25%.

Kata Kunci : Masa Studi, Data Mining, Algoritma C45

I. PENDAHULUAN

Tersedianya data yang melimpah pada institusi pendidikan harus dimanfaatkan dengan baik. Namun sulitnya memahami dan menemukan hubungan atribut-atribut data yang mempengaruhi hasil masa studi mahasiswa yang dapat digolongkan sebagai tepat waktu dan tidak tepat waktu, menjadi kajian dalam penelitian ini. Menganalisa kinerja mahasiswa (student performance), mengidentifikasi keunikan keunikan yang ada pada mahasiswa dan membangun suatu strategi pengembangan lebih lanjut serta tindakan-tindakan yang dapat dilakukan untuk masa mendatang, merupakan tantangan utama bagi universitas modern saat ini (Kabakchieva, 2013). Data mining dapat diusulkan sebagai salah satu pendekatan yang dapat dilakukan untuk memprediksi kinerja siswa (Osmanbegovic & Suljic, 2012). Kinerja siswa dalam hal ini adalah capaian kelulusan studi mahasiswa dimana mereka dapat lulus tepat waktu atau tidak tepat waktu.

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban dkk. 2005). Terdapat

beberapa istilah lain yang memiliki makna sama dengan data mining, yaitu Knowledge discovery in databases (KDD), ekstraksi pengetahuan (knowledge extraction), Analisa data/pola (data/pattern analysis), kecerdasan bisnis (business intelligence) dan data archaeology dan data dredging (Larose, 2005). Penerapan metode data mining dalam menganalisis data yang tersedia di lembaga pendidikan didefinisikan sebagai Educational Data Mining (EDM) (Romero & Ventura, 2007).

EDM merupakan suatu aliran yang relatif baru dalam penelitian data mining. EDM menggunakan beberapa teknik seperti Decision Trees, Neural Networks, Naïve Bayes, KNearest Neighbor dan lainnya (Yadav & Pal, 2012). EDM berkaitan dengan pengembangan metode untuk mengeksplorasi jenis yang unik dari data-data pada pengelolaan pendidikan dan menggunakannya untuk lebih memahami siswa dan pengelolaannya (Baker, 2010). Hal tersebut merupakan tujuan yang ingin dicapai dalam pemanfaatan data mining di bidang pendidikan.

Penelitian Yadav dan Pal (Yadav & Pal, 2012) melakukan prediksi pada data pendidikan untuk mengidentifikasi siswa yang lemah dan membantu mereka untuk mencetak nilai yang lebih baik. Algoritma C4.5, ID3 dan CART diterapkan dan dibandingkan akurasi, hasil menunjukkan bahwa teknik C4.5 memiliki akurasi paling tinggi yaitu 67,78% dibandingkan dengan teknik lainnya. Berdasarkan penelitian-penelitian yang telah dilakukan sebelumnya, algoritma C4.5 akan digunakan pada penelitian ini karena memiliki tingkat akurasi yang lebih baik dan dapat memberikan gambaran klasifikasi mahasiswa yang tepat waktu atau tidak tepat waktu berupa pohon keputusan (Decision Tree) yang bermanfaat bagi pengelola akademik Decision Tree menyerupai sebuah struktur pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas (Han & Kamber, 2007).

Pohon keputusan bekerja mulai dari akar paling atas (root node), jika diberikan sejumlah data uji, misalnya X dimana kelas dari data X belum diketahui, maka pohon keputusan akan menelusuri mulai dari akar sampai node dan setiap nilai dari atribut sesuai data X diuji apakah sesuai dengan aturan Decision Tree, kemudian pohon keputusan akan memprediksi kelas dari tupel X.

II. STUDI LITERATUR

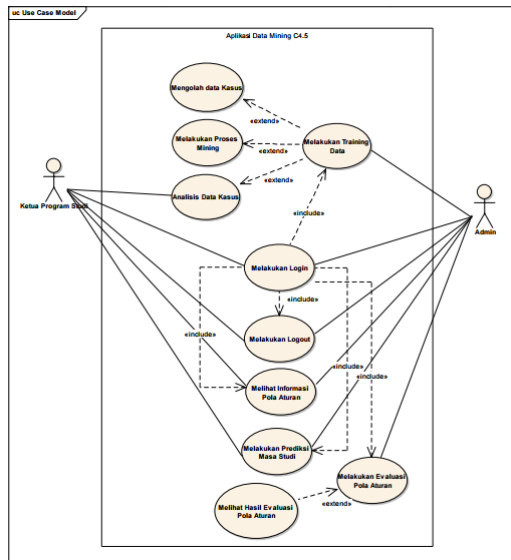
Penelitian ini dilakukan dengan menelusuri literatur serta menelaahnya untuk menggali teori-teori yang sedang berkembang, mencari metode penelitian yang digunakan terdahulu dan untuk memperoleh orientasi yang ada dalam permasalahan. Dalam survei literatur, kami telah melihat bahwa banyak peneliti melakukan penelitian untuk memprediksi masa studi mahasiswa siswa. Terdapat banyak algoritma yang dihasilkan untuk membangun model klasifikasi untuk prediksi mahasiswa, seperti algoritma pohon keputusan C4.5, J48 untuk memprediksi Masa Studi Mahasiswa.

Desain sistem yang digunakan pada pembuatan aplikasi ini dengan berbasis Web dengan bantuan software Sublime dan Xampp, yang mana rancangan aplikasi ini memiliki menu menu atau tools yang memudahkan user dalam penggunaannya atau sering disebut dengan user friendly. Aplikasi ini menerapkan data mining dengan menggunakan algoritma C4.5 untuk mengklasifikasikan status mahasiswa yang berpotensi mangkir. Hasil dari penelitian ini diharapkan mampu membantu Prodi SI untuk melakukan tindakan antisipatif sejak dini pada mahasiswa yang berpotensi mangkir.

Jika masa studi mahasiswa dapat diketahui lebih dini maka pihak jurusan Ilmu Komputer dapat melakukan tindakan-tindakan yang diperlukan supaya mahasiswa dapat lulus dalam waktu kurang dari lima tahun atau tepat lima tahun berdasarkan Peraturan Menteri Pendidikan.

III. METODOLOGI PENELITIAN

a. Use case diagram Aplikasi



Gambar 1 Use Case Diagram Aplikasi Data Mining C4.5

Perancangan sistem dibuat menggunakan UML (Unified Modeling Language).

merupakan use case diagram dari aplikasi data minig. Diagram memiliki dua actor yang berhubungan dengan fungsi sistem, yaitu Admin dan Ketua Program Studi.

C.45

Algoritma yang digunakan membantu pembuatan sistem tersebut ialah Algoritma C45. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (Decision Tree). Pohon keputusan merupakan metode klasifikasi dan prediksi yang terkenal. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain : ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3, Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule.

Tahap tahap algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

- Pilih atribut sebagai akar.

- Buat cabang untuk tiap-tiap nilai.
- Bagi kasus dalam cabang.
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [4]:

1. Pemilihan variabel akar Untuk memilih variabel sebagai akar, didasarkan pada nilai gain tertinggi dari variabel-variabel yang ada. Berikut adalah cara untuk menghitung nilai gain:

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum \text{Entropy}(S_i) \quad (1)$$

dengan:

S : Himpunan kasus

S_i : Himpunan kasus pada partisi ke i

A : Variabel

n : Jumlah partisi atribut A

$|S_i|$: Jumlah kasus pada partisi ke i

$|S|$: Jumlah kasus dalam S Sebelum mendapatkan nilai Gain, dicari terlebih dahulu nilai Entropy.

Entropy adalah informasi mengenai proporsi pembagian kelas, nilai entropy berkisar mulai dari 0 sampai dengan 1, jika nilai entropy = 0, maka menandakan jumlah sampel hanya berada di salah satu kelas, sedangkan jika nilai entropy = 1, maka menandakan jumlah sampel berada di masing-masing kelas dengan jumlah yang sama. Adapun rumus dasar dari perhitungan Entropy adalah sebagai berikut:

$$\text{Entropy}(S) = - \sum_{i=1}^n P_i \log_2 P_i \quad (2)$$

dengan:

S : Himpunan Kasus

n : Jumlah partisi S

p_i : Proporsi dari S_i terhadap S

2. Penentuan cabang untuk masing-masing nilai Untuk penentuan cabang sama seperti mencari variabel akar yaitu didasarkan pada nilai gain tertinggi dari variabel-variabel yang ada.

4. Proses diulang untuk masing-masing cabang sampai semua kelas pada cabang memiliki kelasnya masing-masing. Dalam melakukan perbandingan hasil klasifikasi dan hasil prediksi maka dilakukan perhitungan tingkat akurasi. Untuk mendapatkan tingkat akurasi aplikasi dari hasil prediksi maka digunakan rumus berikut (David. 2014):

$$\text{Persentase} = \frac{\text{jumlah prediksi benar}}{\text{jumlah data}} \times 100\%$$

Algoritma C4.5 dapat menangani berbagai macam contoh kasus. Salah satu contoh dari pengklasifikasian Algoritma C4.5 yaitu tentang masa studi mahasiswa.

b. Design Aplikasi

1. Tampilan login



Gambar 1. Tampilan Login

Pada halaman ini admin dapat melakukan login dengan memasukan username dan password

2. Tampilan Halaman Utama



Aplikasi Data MIning Memprediksi Masa Studi Mahasiswa
dengan Metode Decision Tree C4.5

Gambar 2. Tampilan Halaman Utama

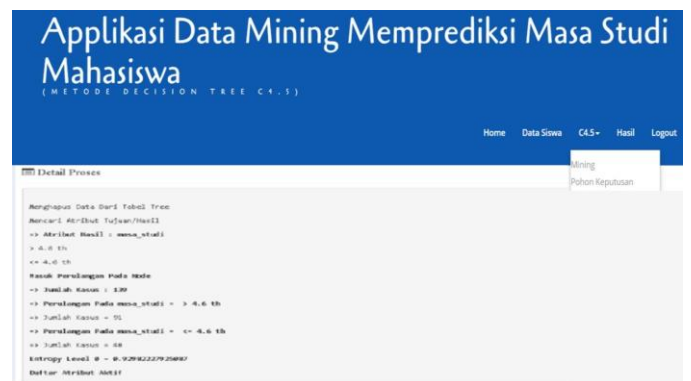
3. Tampilan Inputan



Gambar 3. Tampilan inputan

Pada halaman ini admin atau user dapat menginputkan data pribadi masing masing dengan atribut NIM,Username,jeniskelamin,IPK,penjurusan setelah itu Save

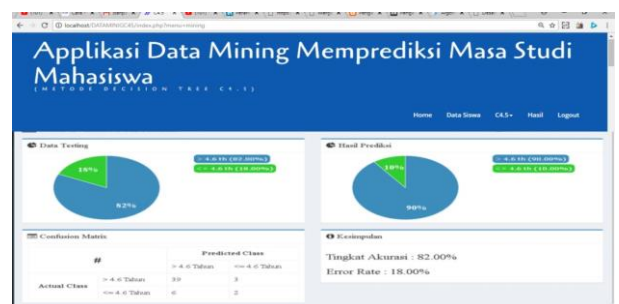
4. Tampilan Halaman Proses Mining



Gambar 4. Tampilan Halaman Proses Mining

Hasil implementasi perancangan antarmuka halaman proses mining ditunjukan pada Gambar 4. Pada halaman ini admin dapat memulai proses mining terhadap data kasus.

5. Tampilan Halaman Hasil Evaluasi Pola Aturan



Gambar 5 Halaman Hasil Evaluasi Pola Aturan

Berdasarkan hasil evaluasi yang dilakukan terhadap perangkat lunak yang dibangun, dapat ditarik kesimpulan bahwa pola aturan yang dibangun algoritme C4.5 mampu mengklasifikasikan mahasiswa ke dalam kelas-kelas yang telah ditentukan sebelumnya dengan nilai error sebesar 27.75% dan nilai akurasi diatas 50% yaitu 72.25%. Dalam penelitian ini, pola aturan yang berhasil diekstrak sebanyak 22 pola aturan, dengan 17 pola aturan mengelompokkan data ke klas '> 4.6 th', dan 5 pola aturan mengelompokkan data ke klas '<= 4.6 th'. Sesuai data hasil evaluasi, maka tingkat akurasi yang berhasil dicapai oleh algoritme c4.5 sebesar 72.25%. Hasil Pola aturan terlihat pada pohon keputusan pada Gambar 6.

c. Decision Tree



Gambar 6 Pohon Keputusan Algoritme C4.5

V. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, maka kesimpulan yang dapat diambil adalah:

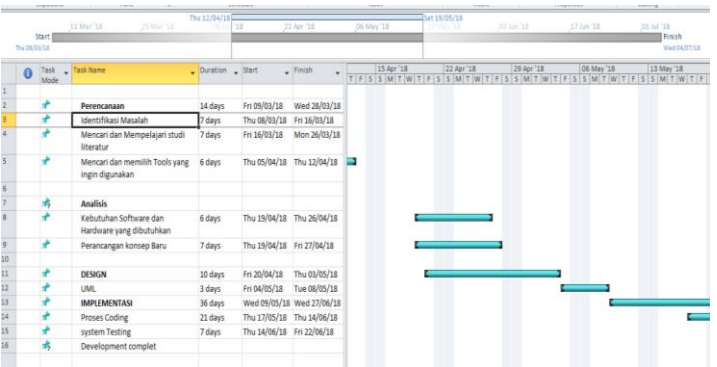
1. Dalam penelitian ini, pola aturan yang berhasil diekstrak sebanyak 22 pola aturan, dengan 17 pola aturan mengelompokkan data ke kelas '> 4.6 th',

dan 5 pola aturan mengelompokkan data ke klas '<= 4.6 th'

2. Hasil evaluasi pola aturan menggunakan k-fold cross validation menunjukkan tingkat akurasi 72.25% dengan tingkat error 27.75% sehingga dapat disimpulkan bahwa data mining dengan algoritme C4.5 dapat diimplementasikan untuk memprediksi masa studi mahasiswa dengan empat atribut yaitu data jumlah SKS yang sudah diselesaikan saat semester 7, IPK mahasiswa saat semester 7, matakuliah konsentrasi yang diambil mahasiswa dan asal sekolah mahasiswa.
3. Pohon keputusan yang dihasilkan dari algoritme c4.5 menunjukkan bahwa atribut IPK dan asal sekolah merupakan atribut yang paling dominan diantara atribut yang lain dalam menentukan masa studi mahasiswa. Hal ini terlihat dari root node pohon keputusan yang merupakan atribut IPK dan internal node ke-2 yang memiliki terminal node '<= 4.6 th' merupakan atribut asal sekolah.

Jadwal Penelitian

Penelitian ini dilaksanakan pada rentang waktu bulan Maret 2018 sampai dengan bulan Juni 2018. Adapun rincian kegiatan sebagai berikut:



Gambar 7 Jadwal Penelitian Menggunakan Microsoft Project

Saran

Hasil dari penelitian akhir metode c45 ini masih terdapat beberapa kekurangan, antara lain:

1. Dapat dikombinasikan dengan metode lain agar hasil lebih sempurna

DAFTAR ACUAN

- [1] Riadi Muchlisin, 2017, "Pengertian, Fungsi, Proses dan Tahapan Data Mining" diakses dari <https://www.kajianpustaka.com/2017/09/datamining.html> pada tanggal 27 Juni 2018
- [2] Chendrawira Jennifer, 2016, "Makalah Data Mining" diakses dari https://www.academia.edu/29553365/MAKALAH_DATA_MINING-SEM_3_PSI pada tanggal 27 juni 2018.
- [3] Quinlan, J.R. 1993. C4.5 : Programs For Machine Learning. Morgan Kaufmann Publisher, Inc. San Mateo
- [4] Ilmu skripsi, 2016, "Algoritma C45" diakses dari <https://www.ilmuskripsi.com/2016/07/algoritma-c45.html> pada tanggal 27 juni 2018
- [5] Pramudiono. (2017, February 26). iko-datamining. Retrieved from ilmukomputer: <http://www.ilmukomputer.org/wpcontent/uploads/2006/08/iko-datamining.zip>
- [6] Rokach, L and Maimon, O. 2008. Data Mining With Decision Trees : Theory and Applications. World Scientific Publishing Co. Pte. Ltd. Singapura.
- [7] Sunjaya. 2010. Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree. Seminar Nasional Aplikasi Teknologi Informasi 2010. Yogyakarta.
- [8] Informatikalogi . 13 July 2017 Algoritma C4.5 . diakses dari <https://informatikalogi.com/algoritma-c4-5/> pada tanggal 1 juni 2018.