

REKOMENDASI PEMILIHAN PROGRAM STUDI PESERTA DIDIK PADA UNIVERSITAS TARUMANGARA MENGGUNAKAN METODE DECISION TREE DENGAN ALGORITMA C4.5

Reynaldo¹, Bagus Mulyawan², Tri Sutrisno³

^{1,2,3} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: ¹aldoboy97@gmail.com, ²bagus@fti.untar.ac.id, ³tris@fti.untar.ac.id

Abstrak

Pendidikan merupakan suatu hal yang penting dalam kehidupan seseorang, penjurusan dalam jenjang pendidikan sekolah dasar sampai sekolah menengah atas masih belum membingungkan. Pada saat ingin masuk perguruan tinggi yang memiliki jurusan yang cukup banyak jika di bandingkan pada tingkat SMA atau SMK, hal ini yang membuat para siswa yang akan berubah statusnya menjadi mahasiswa bingung dan sering kali juga salah dalam memilih jurusan yang akan diambil dalam perguruan tinggi tersebut.

Sistem rekomendasi jurusan ini menggunakan Data Mining yaitu metode Decision Tree dengan algoritma C4.5 dan dibantu dengan Confusion Matrix yang berfungsi untuk mengetahui tingkat akurasi yang dihasilkan.

Hasil pengujian dari beberapa skema percobaan yang sudah dilakukan menunjukkan bahwa, fakultas Teknologi Informasi memiliki tingkat akurasi yang paling besar yaitu 80.76% yang dihitung dengan menggunakan rumus akurasi dari Confusion Matrix jika dibandingkan dengan percobaan pada fakultas lain. Dari pengujian tersebut dapat diketahui bahwa faktor nilai masukan atau faktor x dan juga nilai keluaran yang akan dihasilkan atau faktor y mempengaruhi presentase yang akan dihasilkan. Semakin banyak variabel nilai masukan yang diproses jika dibandingkan dengan variabel nilai keluaran yang dihasilkan maka hasil presentase akan semakin besar, dan juga sebaliknya.

Kata kunci—Decision Tree, Algoritma C4.5, Confusion Matrix, Rekomendasi, Program Studi

Abstract

Education is an important thing in a person's life, majors in the level of elementary school to high school education that has not been closed. When you want to enter college that has quite a lot of majors at high school level or vocational high school, this makes students who will change their status to be confused and often mistaken in choosing the direction to be taken at the college.

This department's recommendation system uses Data Mining, namely the Decision Tree method with the C4.5 algorithm and is assisted by the Confusion Matrix needed to obtain the level of accuracy produced.

80.76% which is calculated using the accuracy formula of the Confusion Matrix when compared to experiments in other faculties.

From the test, it can be seen that the factor entered or factor x and also the value that will be produced or the factors that will increase the percentage that will be produced. The more values entered that need to be compared with the value issued, the greater the percentage will be generated, and vice versa.

Keyword Decision Tree, Algorithm C4.5, Confusion Matrix, Recommendation, Accurac

1. PENDAHULUAN

Sistem yang akan dirancang adalah sistem rekomendasi program studi peserta didik yang akan masuk ke Universitas Tarumanagara, sistem ini dirancang untuk menghasilkan keluaran atau hasil akhir berupa rekomendasi program studi untuk calon mahasiswa tersebut. Hasil rekomendasi itu diperoleh dan juga melalui tahapan awal yaitu, calon mahasiswa harus memasukkan nilai yang mereka dapatkan pada saat di bangku SMA atau SMK, nilai yang diperlukan adalah nilai Bahasa Indonesia, Bahasa Inggris dan Matematika dari semester 1 sampai dengan semester 4 atau dari kelas 10 sampai kelas 11.

Metode yang digunakan dalam sistem ini yaitu *Decision Tree*. *Decision Tree* atau pohon keputusan adalah suatu metode klasifikasi yang menggunakan struktur pohon yang setiap node merepresentasikan atribut dan cabangnya merepresentasikan nilai dari atribut, sedangkan daunnya digunakan untuk merepresentasikan kelas. *Decision Tree* digunakan untuk menghasilkan aturan yang akan digunakan sebagai dasar acuan penilaian dalam memberikan hasil rekomendasi program studi. *Decision Tree* memiliki berbagai algoritma yang dapat digunakan yaitu CART, ID3, C4.5, dan C5.0. Pada sistem ini diterapkan algoritma C4.5.

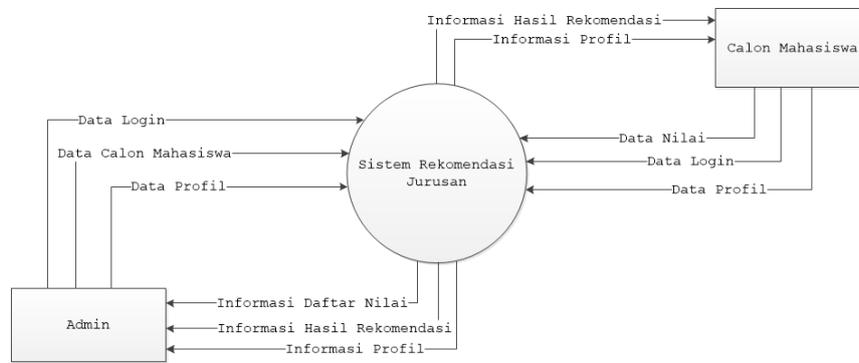
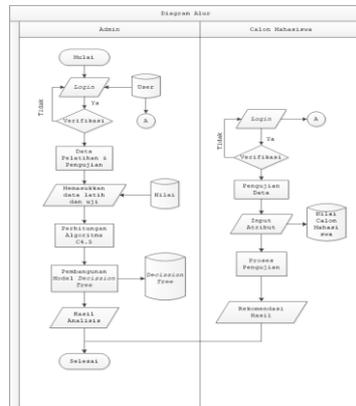
Sistem yang dirancang menggunakan *Decision tree* dengan algoritma C4.5. Pemilihan metode dilakukan berdasarkan penelitian yang dilakukan oleh Hamidah Jantan. Pada penelitiannya dalam memprediksi kinerja seseorang didapatkan kesimpulan bahwa *Decision tree* dengan algoritma C4.5 memiliki potensi yang sangat baik dalam memprediksi kinerja seseorang dibandingkan algoritma lainnya.[2] Selain itu, Anitha Mary Florence juga menyatakan bahwa klasifikasi *Decision tree* dengan algoritma C4.5 memiliki tingkat akurasi yang tinggi dengan error rates sebesar 0.02.[1] Dengan menggunakan metode tersebut, penilaian sikap dan prestasi kerja karyawan dapat dilakukan secara otomatis, mempermudah proses penilaian serta dengan hasil yang akurat dan sesuai dengan pendapat direksi.

Naskah ditulis dengan Times New Roman ukuran font 11, rata kiri dan kanan (*justify alignment*), pada kertas A4 (210 x 297 mm) margin atas, kanan, kiri, dan bawah adalah 3 cm. Naskah termasuk isi grafik, tabel, dan referensi harus antara 8 sampai 12 halaman. Naskah ditulis dalam Bahasa Indonesia yang baku. Daftar pustaka diurutkan sesuai kemunculan pada naskah. Penulisan rujukan adalah kalimat kemudian diakhir oleh nomor rujukan [1].

2. METODE PENELITIAN

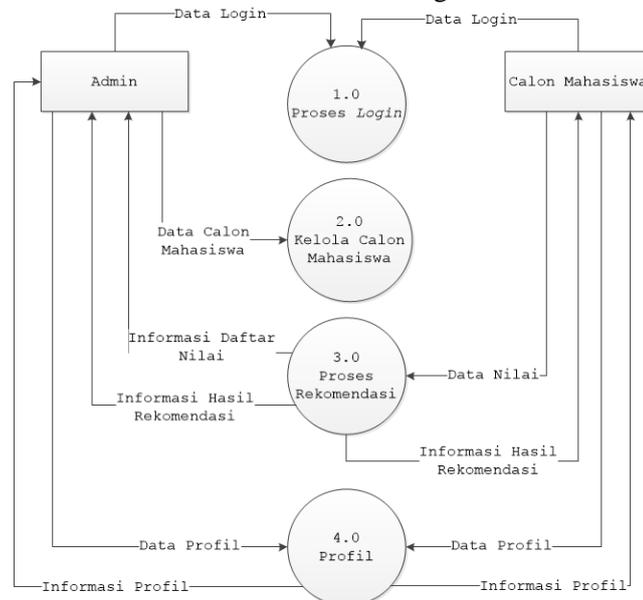
Aplikasi ini berbasis website dan dibangun dengan menggunakan bahasa pemrograman Python untuk pengolahan *Data Mining* dan PHP untuk pembuatan website, dengan basis data dan server menggunakan MySQL yang sudah terdapat di dalam XAMPP. Rekayasa perangkat lunak menggunakan *System Development Life Cycle* (SDLC) yang berisi tahapan *Planning Defining, Designing, Building, Testing, Deployment*. Tahapan awal yang dilakukan dalam

Pembuatan program aplikasi ini ialah merancang sistem basis data, selanjutnya lanjut ke tahap pembuatan *flowchart* program dan *flowchart* metode *Decision Tree* dengan Algoritma C4.5 yang dapat dilihat pada gambar 1 dan juga beserta pembuatan *Contex Diagram* yang dapat dilihat pada gambar 2 dan *Data Flow Diagram* pada gambar 3 yang digunakan untuk membantu perancangan ini agar terbentuk lebih efektif dan efisien.



Gambar 1 Flowchart Program

Gambar 2 Context Diagram



Gambar 3 Data Flow Diagram

Langkah selanjutnya yaitu pembuatan basis data sesuai dengan rancangan yang sudah dibentuk dengan menggunakan MySQL yang sudah terhubung dengan PHP MyAdmin. Setelah

tahap perancangan sistem selesai, langkah selanjutnya melakukan proses *Data Mining* yaitu *preprocessing* untuk memisahkan data-data yang tidak dipakai untuk kelangsungan pembuatan program

2.1 Data Mining

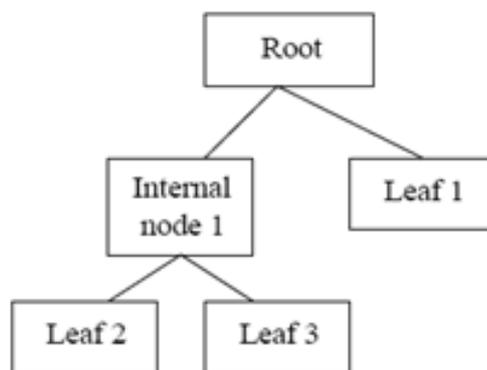
Data Mining merupakan sebuah analisa dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya dan metode baru untuk data agar mudah dipahami serta kegunaannya untuk pemilik data tersebut[3]. Metode, teknik, dan algoritma dalam *Data Mining* memiliki banyak model atau variasi yang berbeda-beda. *Data Mining* dapat menganalisa data yang besar menjadi hasil akhir berupa informasi yang berbentuk pola dimana pola tersebut memiliki arti atau kesimpulan akhir bagi pendukung keputusan atau rekomendasi. Lalu, hasil dari pengolahan *Data Mining* tersebut dievaluasi dan diteliti untuk menemukan suatu informasi baru yang bernilai dan kemudian divisualisasikan agar mempermudah *user* untuk menganalisa informasi sebagai pendukung keputusan.

2.1.1 Decision Tree

Decision tree atau pohon keputusan adalah suatu metode klasifikasi yang menggunakan struktur pohon yang setiap *node* merepresentasikan atribut dan cabangnya merepresentasikan nilai dari atribut, sedangkan daunnya digunakan untuk merepresentasikan kelas. Pohon dalam analisis pemecahan masalah pengambilan keputusan adalah pemetaan mengenai alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. Gambar pohon keputusan dapat dilihat pada gambar 4.

Pohon keputusan sendiri memiliki 3 jenis bagian node, yaitu sebagai berikut : [4]

1. *Root Node*
Root node merupakan node teratas. Pada *node* ini, tidak ada *input* atau simpul akar tidak memiliki *input* cabang yang masuk dan memiliki cabang lebih dari satu.
2. *Internal Node*
Internal node atau simpul internal hanya memiliki 1 cabang yang masuk dan dapat memiliki lebih dari 1 cabang yang keluar.
3. *Leaf Node*
Leaf node atau simpul daun merupakan simpul akhir yang hanya memiliki 1 cabang yang masuk, tidak memiliki cabang sama sekali, dan menandakan bahwa simpul tersebut merupakan label kelas.



2.2.1 Decision Tree

Algoritma C4.5 merupakan kelompok algoritma pohon keputusan. Algoritma ini mempunyai *input* berupa *training samples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya, sedangkan *samples* adalah *field* data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data.

Langkah kerja dari algoritma C4.5 untuk menghitung nilai *entropy*, nilai *information gain*, dijelaskan di bawah ini: [5]

1. Entropy

Entropy digunakan untuk menghitung kemiripan data atau *impurity* pada data latih.

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

2. Information Gain

Information Gain digunakan untuk menentukan berapa banyak informasi yang dapat diberikan oleh *attribute* terhadap kelas yang ada

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

3. HASIL DAN PEMBAHASAN

Pengujian yang dilakukan terhadap program dilakukan dengan 2 tahap, yaitu pengujian terhadap modul dan pengujian terhadap modul dan terhadap data. Pengujian terhadap modul ini bertujuan untuk menguji apakah modul-modul yang telah dirancang berfungsi sesuai dengan rancangan. Tahap pengujian data dilakukan pengujian pada proses perhitungan akurasi, bertujuan untuk mengevaluasi proses perhitungan pada program dengan hasil rekomendasi akhir yang akan diberikan sudah sesuai atau belum.

Pada tahap pengujian modul ini menunjukkan bahwa semua modul yang dijalankan dapat berfungsi dan berjalan sesuai dengan rancangan, percobaan terhadap modul yang telah dibuat yaitu modul *Sign In*, *Input* data nilai rapor, cek data mahasiswa, *profile* dan *Log Out*.



Data yang digunakan merupakan data asli mahasiswa Universitas Tarumanagara tahun angkatan 2016 sampai 2018 sebanyak 9797 data yang dapat dilihat pada tabel 1, yang digunakan untuk masuk ketahap *Data Mining* lalu mengolah menjadi *data training* dan *data*

testing. Pada tahap pengujian ini dilakukan beberapa skema percobaan pada setiap fakultas yaitu fakultas Teknik, Teknologi Informasi, Seni Rupa dan Desain, Ekonomi dan semua fakultas yang teradapat di Universitas Tarumanagara kecuali fakultas kedokteran.

Pengujian akurasi hasil akhir yang dihasilkan terhadap proporsi data dilakukan untuk mengetahui apakah sistem dapat berjalan sesuai dengan yang sudah ditentukan. Aplikasi Rekomendasi Program Studi Peserta Didik Pada Universitas Tarumanagara Menggunakan *Decision Tree* Dengan Algoritma C4.5 memiliki beberapa tahapan perhitungan sampai dapat membuat suatu pohon keputusan. Rumus perhitungan yang digunakan yaitu *Entropy* dan *Information Gain*. Nilai rapor yang akan diproses diantaranya nilai Bahasa Indonesia, Bahasa Inggris dan Matematika dari semester 1 sampai 4 siswa yang diperoleh pada saat SMA atau SMK.

Total data keseluruhan sebanyak 9797 data. Data yang digunakan pada percobaan fakultas Teknologi Informasi sebanyak 519 data dan dapat dilihat pada Tabel 3. Dengan pembagian proporsi data yang berbeda-beda yang dapat dilihat pada Tabel 2. Perhitungan hasil akurasi menggunakan metode *Confusion Matrix* dapat dilihat pada tabel 4.

Tabel 1 Jumlah Data Program Studi

Daftar Program Studi	Jumlah Data
Ilmu Psikologi	934
Ilmu Hukum	733
Ilmu Komunikasi	711
Desain Interior	492
Desain Komunikasi Visual	401
Teknik Informatika	375
Sistem Informasi	144
Manajemen Bisnis	1767
Akuntansi Bisnis	2012
Arsitektur	862
Perencanaan Tata Kota & Real Estat	92
Teknik Sipil	685
Teknik Elektro	89
Teknik Mesin	142

Pengujian ke-	Pembagian Data	Training	Testing	Benar	Salah	Akurasi
1	60:40	311	208	146	61	70.19%
2	70:30	363	156	116	40	74.35%
3	80:20	415	104	80	24	76.92%
4	90:10	467	52	42	10	80.76%

Tabel 3 Jumlah Data Fakultas Teknologi Informasi

Program Studi	Jumlah
Teknik Informatika	375
Sistem Informasi	144
Total	519

Tabel 4 *Confusion Matrix* Pengujian Fakultas Teknologi Informasi

Kelas Sebenarnya	Sistem Informasi	Teknik Informatika	Jumlah
	Sistem Informasi	3	10
Teknik Informatika	0	39	39
			52

Hasil yang ditampilkan diatas merupakan hasil percobaan dengan perolehan nilai akurasi yang terbesar yaitu Fakultas Teknologi Informasi. Skema percobaan tersebut dilakukan pada setiap fakultas dengan cara yang sama tetapi dengan menggunakan data latih dan data uji yang berbeda pada setiap percobaannya menyesuaikan data yang terdapat di setiap fakultas.

4. KESIMPULAN

Kesimpulan yang dapat diperoleh berdasarkan pembuatan dan pengujian dari sistem ini adalah sebagai berikut :

Teknik Industri	358
Total	9797

Tabel 2 Pengujian Akurasi Terhadap Proporsi Data Fakultas Teknologi Informasi

1. Hasil pengujian modul menggunakan metode blackbox *testing* sudah sesuai. Fitur dan menu yang terdapat di dalam setiap modul dapat berjalan dengan seharusnya dan dapat dipergunakan sebagaimana mestinya.
2. Hasil pengujian percobaan 1 sampai percobaan 5 dengan beberapa skema percobaan yang berbeda-beda, mendapatkan hasil akurasi terbesar 80.76% pada fakultas Teknologi Informasi. Untuk akurasi terbesar pada setiap pengujian fakultas yaitu untuk fakultas Teknik sebesar 43.94%, 70.90% untuk fakultas Ekonomi, 70.0% untuk fakultas Seni Rupa dan Desain, dan 26.93% untuk semua program studi yang ada.
3. Semakin banyak *input* (masukkan) atau variabel *x* yang dimiliki dan diproses jika dibandingkan *output* (keluaran) atau variabel *y* yang akan dihasilkan, akan mendapatkan hasil akhir dengan akurasi yang lebih baik dan juga sebaliknya. Banyaknya penggunaan data untuk membentuk pohon keputusan juga mempengaruhi hasil akurasi akhir yang akan didapatkan.

UCAPAN TERIMA KASIH

Terima kasih penulis ucapkan kepada Tuhan Yang Maha Esa karena atas Kasih dan Rahmat-Nya penulis dapat membuat naskah ini, juga kepada pembimbing yang sudah membantu dalam merangkai naskah ini, serta kepada seluruh pihak Universitas Tarumanagara yang sudah membantu penulis dalam proses perkuliahan. Terima kasih juga terhadap segenap keluarga, teman-teman dan pendahulu penulis yang selalu memberikan dukungan moral dan materil dalam pembuatan naskah ini.

DAFTAR PUSTAKA

- [1] Anita Mary Florence and Ms.Savithri. *Talent Knowledge Acquisition Using C4.5 Classification Algorithm. International Journal Of Emerging Technologies in Computational and Applied Sciences (IJETCAS)*. 2013
- [2] Hamidah, 2014. Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik. Seminar Nasional Aplikasi Teknologi Informasi 2014. 2014
- [3] David Hand, Heikki Mannila dan Padhraic Smith, Principles of data mining, (London: A Bradford book The MIT Press, 2001), h. 104.
- [4] Sunjana , 2010, Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree, Seminar Nasional Aplikasi Teknologi Informasi, Yogyakarta.
- [5] Saedsayad, Decision Tree-Classcation https://www.saedsayad.com/decision_tree.html