

IMPLEMENTASI ALGORITMA C4.5 UNTUK MEMPREDIKSI MASA STUDI MAHASISWA TEKNIK INFORMATIKA BERBASIS WEB

Tiefani Permata Sarie¹, Dian Novianti², Agust Isa Martinus³

¹²³Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Cirebon Jl.
Fatahillah, Watubelah, Kec. Sumber, Cirebon, Jawa Barat, Indonesia, 45611

¹tiefanipermatasari06@gmail.com, ²dian.novianti@umc.ac.id, ³agust.isa@umc.ac.id

ABSTRACT

Higher education is a continuation of secondary education which is held to prepare students to become members of society Higher education is a continuation of secondary education which is held to prepare students to become members of society who have academic and professional abilities, Regulation of the Minister of Education and Culture of the Republic of Indonesia No. undergraduate programs ", and paragraph 3 " 4 years to 5 years for four and undergraduate diploma programs ". Every college and every faculty needs to pay attention to the level of student study duration to improve the quality of both the College itself and the Faculty. The implementation of the C4.5 Algorithm in predicting the Student Study Period can be used as a medium to help find out the length of study that will be taken. The C4.5 algorithm is a classification algorithm that produces a rule or decision tree. The C4.5 algorithm is a classification algorithm, namely an algorithm that uses data with a target / class / label in the form of a categorical value (nominal) which aims to determine the prediction of the study period of the Informatics Engineering Student at Muhammadiyah University of Cirebon. The parameters used in this study were the values of Ips_1, Ips_2, Ips_3, Ips_4, Ips_5, study time, and gender. This system produces information about the study period of Informatics Engineering Students, and produces an accuracy value using a confusion matrix, the results of the study also show that the level of accuracy of the C4.5 algorithm method using 15 data has a level of accuracy, rate error, sensitivity, and specificity of 73, 33%, 26.67%, 0%, and 78.57%, using 10 data has a level of accuracy, the success rate of the system, and the accuracy of the decision results of 60%, 40%, 0%, and 66.67%, and while using 6 random data from training data has a level of accuracy, rate error, sensitivity, and specificity of 71.43%, 28.57%, 100%, and 60%.

Keyword : Student Study Period Prediction informatics engineering, C4.5 Algorithm, Data Mining

ABSTRAK

Perguruan tinggi merupakan kelanjutan pendidikan menengah yang diselenggarakan untuk mempersiapkan peserta didik menjadi anggota masyarakat yang memiliki kemampuan akademis dan profesional, Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia N0.49 Tahun 2014 Tentang Standar Nasional Pendidikan Tinggi pasal 17 ayat 2 "144 sks untuk diploma empat dan program sarjana", dan ayat 3 "4 tahun sampai 5 tahun untuk program diploma empat dan sarjana". Setiap Perguruan Tinggi dan setiap Fakultas perlu memperhatikan jenjang Lama Studi Mahasiswa untuk meningkatkan kualitas baik dari Perguruan itu sendiri dan Fakultasnya. Implementasi Algoritma C4.5 dalam memprediksi Masa Studi Mahasiswa dapat dijadikan suatu media untuk membantu mengetahui jenjang lama Studi yang akan ditempuh. Algoritma C4.5 merupakan Algoritma Klasifikasi yang menghasilkan rule atau pohon keputusan. Algoritma C4.5 merupakan algoritma klasifikasi yaitu algoritma yang menggunakan data dengan target/class/label berupa nilai kategorikal (nominal) bertujuan untuk menentukan prediksi masa Studi Mahasiswa Teknik Informatika Universitas Muhammadiyah Cirebon. parameter yang digunakan dalam penelitian ini adalah nilai Ips_1, Ips_2, Ips_3, Ips_4, Ips_5, waktu studi, dan jenis kelamin. Sistem ini menghasilkan informasi mengenai masa studi Mahasiswa Teknik Informatika, dan menghasilkan nilai akurasi dengan menggunakan confusion matrix, hasil penelitian juga menunjukkan bahwa tingkat keakuratan metode algoritma C4.5 dengan menggunakan 15 data memiliki tingkat akurasi, laju error, sensitivitas, dan spesifitas sebesar 73,33%, 26,67%, 0%, dan 78,57%, menggunakan 10 data memiliki tingkat akurasi, tingkat keberhasilan sistem, dan tingkat ketepatan hasil keputusan sebesar 60%, 40%, 0%, dan 66,67%, dan sedankan menggunakan 6 data random dari data training memiliki tingkat akurasi, laju error, sensitivitas, dan spesifitas sebesar 71,43%, 28,57%, 100%, dan 60%.

Kata Kunci : Prediksi Masa Studi Mahasiswa Teknik Informatika , Algoritma C4.5 Data Mining

1. PENDAHULUAN

Pada Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia N0.49 tahun 2014 Tentang Standar Nasional Pendidikan Tinggi pasal 17 ayat 2 “144 sks untuk diploma empat dan program sarjana”, dan ayat 3 “4 tahun sampai 5 tahun untuk program diploma empat dan sarjana”. Dan peraturan akademik yang berada pada buku pedoman mahasiswa Universitas Muhammadiyah Cirebon tahun 2016 “seseorang dinyatakan lulus Program Sarjana(S1) Teknik Informatika apabila telah menempuh jumlah kredit sekurang-kurangnya 148 SKS (satuan kredit semester) yang dijadwalkan untuk 8 semester dan dapat ditempuh dalam waktu kurang dari 8 semester dan paling lama 14 semester.

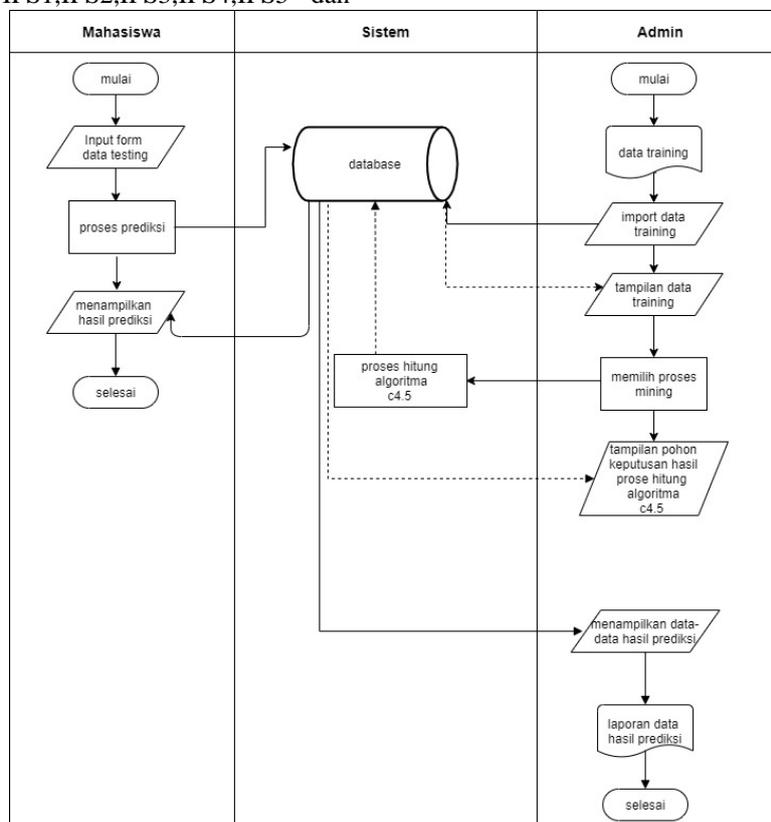
Universitas Muhammadiyah Cirebon memiliki banyak jurusan salah satunya yaitu Jurusan Teknik Informatika hanya beberapa orang mahasiswa yang lulus dalam 8 semester. Hal ini menunjukkan bahwa masih banyak mahasiswa Program Sarjana(S1) reguler maupun non reguler di jurusan Teknik Informatika yang menempuh lama Studi lebih dari 8 semester. Melihat kondisi di atas tersebut diperlukan penelitian untuk menggali data yang dimiliki oleh Jurusan Teknik Informatika.

Pada Penelitian ini dasar penentuan parameter yang digunakan yaitu IPS1,IPS2,IPS3,IPS4,IPS5 dan

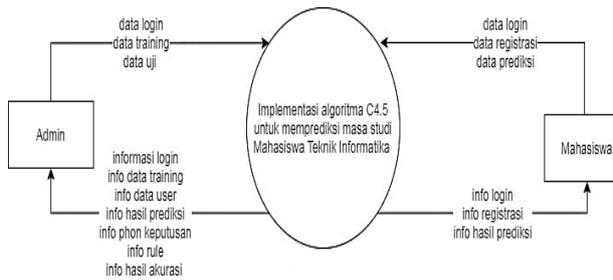
jenis kelamin. Parameter tersebut digunakan untuk proses data training dan dikelola sehingga didapatkan hasil untuk membaca data testing. Minimnya pengetahuan Mahasiswa yang masih aktif mengenai masa studi masing-masing Mahasiswa, Jika masa studi Mahasiswa dapat diketahui lebih dini maka pihak Jurusan Teknik Informatika dapat melakukan tindakan-tindakan yang diperlukan, agar Mahasiswa dapat lulus kurang dari 8 semester atau sama dengan 8 semester.

2. METODE

Data yang akan dimanfaatkan disini adalah data nilai Akademik Alumni Mahasiswa Angkatan tahun 2013-2016 yang telah lulus jumlah data yang digunakan 50, 35 data digunakan sebagai data training, dan 15 data digunakan sebagai data validasi. Dan data Mahasiswa yang belum lulus digunakan sebagai data uji untuk memprediksi lama Studi masing-masing Mahasiswa. Dibutuhkan suatu Teknik klasifikasi yang merupakan salah satu teknik dari data mining untuk menganalisis data jurusan Teknik Informatika tersebut dengan menerapkan Algoritma C4.5, untuk memprediksi Masa Studi Mahasiswa dan melihat kemungkinan Mahasiswa yang lulus lebih dari 8 semester yang sedang menempuh perkuliahan.



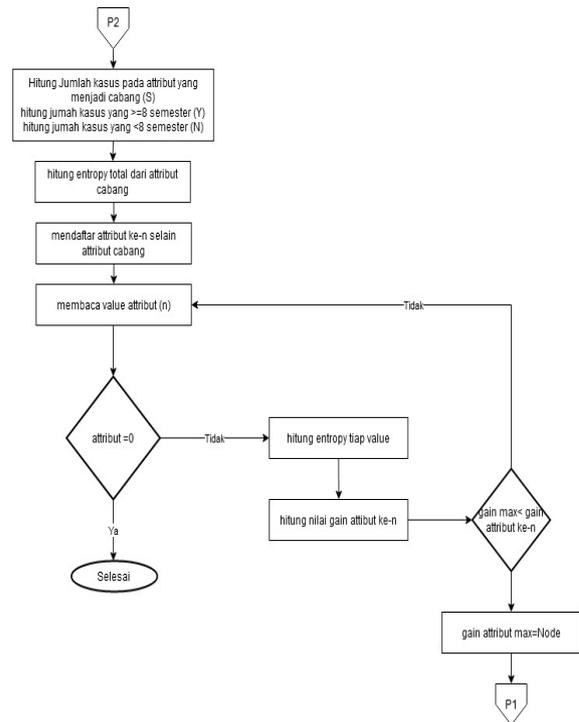
Gambar 1 Perancangan Sistem



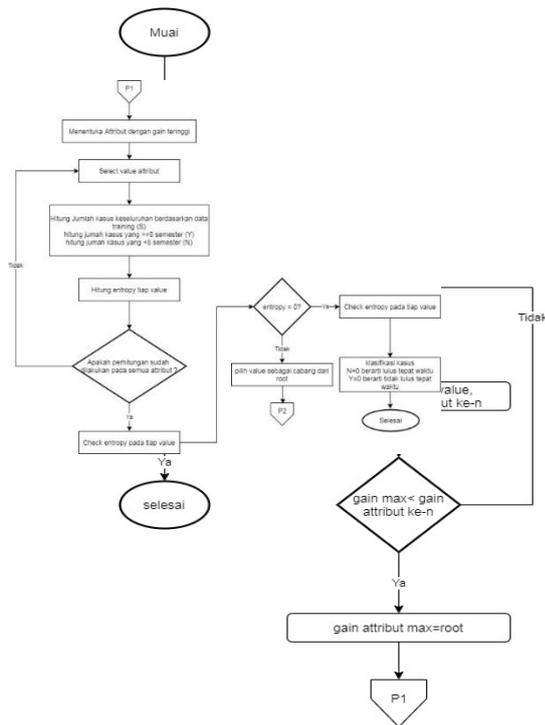
Gambar 2 Diagram Konteks

Algoritma C4.5

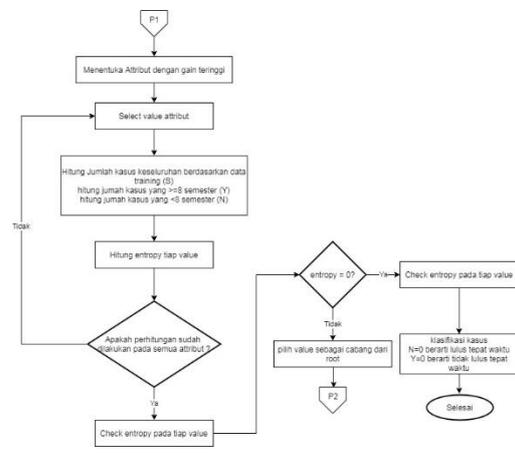
Algoritma C4.5 yaitu sebuah algoritma yang digunakan untuk membangun decision tree (pengambilan keputusan). Algoritma C.45 adalah salah satu algoritma induksi pohon keputusan yaitu ID3 (Iterative Dichotomiser 3). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel training, label training dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3[3].



Gambar 4 Penentuan node

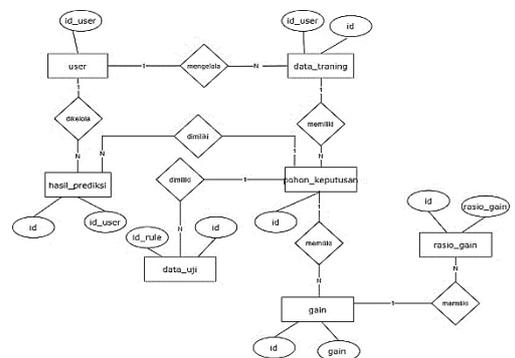


Gambar 3 Penentuan Akar



Gambar 5 Penentuan Cabang

Entity Relationship Diagram merupakan model data yang dipergunakan sebagai alat untuk mendefinisikan kebutuhan-kebutuhan tentang database suatu organisasi. Suatu teknik untuk mengorganisasikan dan mendokumentasikan data sistem[4].



Gambar 6 ERD

3. HASIL DAN PEMBAHASAN

Menghitung entropy dengan data training yang Vtela diimport dan diketahui nilai attributnya. nilai attribut ke-1 dibagi dengan jumlah data training *log2 + nilai attribut ke-2 dibagi dengan jumlah data training.

```
function hitung_entropy($nilai1, $nilai2){
    global $koneksi;
    $jumlah_data = $nilai1 + $nilai2;
    //jika salah satu nilai 0, maka entropy 0
    if($nilai1==0 or $nilai2==0){
        $entropy = 0;
    }else{
        $entropy= (-
        ($nilai1/$jumlah_data)*(log(($nilai1/$jumlah_data),2
        ))) + (-
        ($nilai2/$jumlah_data)*(log(($nilai2/$jumlah_data),2
        )));
    }
    //desimal 3 angka dibelakang koma
    $entropy = round($entropy, 3);

    return $entropy;
}
```

Menghitung nilai akurasi

```
$tepatwaktu=($TP+$TN);
$tidaktepatwaktu=($FP+$FN+$kosong);
$akurasi=($tepatwaktu/$jumlah)*100;
$laju_error=($tidaktepatwaktu/$jumlah)*100;
$sensitivitas=($TP/($TP+$FN))*100;
$spesifisitas=($TN/($FP+$TN))*100;
```

Pengujian

Pengujian sistem dilakukan dengan cara menguji setiap proses pada sistem sesuai dengan rencana pengujian yang telah ditentukan. Pengujian proses perhitungan dilakukan dengan cara menguji perhitungan secara manual dan perhitungan secara sistem dimana hasil akhir dari keduanya harus sesuai. Pengujian ini bertujuan untuk melihat kesesuaian algoritma perhitungan pada sistem dengan rumus yang digunakan pada perhitungan manual. Hasil dari pengujian sesuai dengan yang diharapkan yaitu sistem mampu menghasilkan nilai entropy sesuai dengan yang diinputkan.

Nilai Atribut	Jumlah data	Jumlah Tepat waktu	Jumlah Tidak tepat waktu	Entropy	Gain
ip_1=rendah	12	0	12	0	
ip_1=tinggi	14	10	4	0.863	
ip_1=sedang	9	0	9	0	0.318
ip_2=rendah	10	0	10	0	
ip_2=sedang	11	2	9	0.694	
ip_2=tinggi	14	8	6	0.985	0.254
ip_3=rendah	9	0	9	0	
ip_3=sedang	18	5	13	0.832	
ip_3=tinggi	8	5	3	0.914	0.307
ip_4=rendah	13	3	10	0.779	
ip_4=tinggi	8	4	4	0.811	
ip_4=sedang	14	1	13	0.371	0.24

Gambar 7 Hasil Pengujian

Pengujian Akurasi

	waktu	waktu									
12	12	tinggi	tinggi	tinggi	tinggi	L	tidak tepat waktu	tepat waktu	5	Salah	
13	13	sedang	sedang	sedang	sedang	P	tidak tepat waktu	tidak tepat waktu	2	Benar	
14	14	Sedang	tinggi	sedang	sedang	L	tidak tepat waktu	tepat waktu	3	Salah	
15	15	tinggi	tinggi	tinggi	tinggi	sedang	P	tidak tepat waktu	tidak tepat waktu	4	Benar

Jumlah data yang diprediksi: 15
 Jumlah data yang diprediksi tepat waktu: 11
 Jumlah data yang diprediksi tidak tepat waktu: 4
 AKURASI = 73.33 %
 LAJU ERROR = 26.67 %
 TP: 0 | TN: 11 | FP: 3 | FN: 1
 SENSITIVITAS = 0 %
 SPESIFISITAS = 78.57 %

Gambar 8 Hasil Pengujian Akurasi

4. PENUTUP

Berdasarkan uraian dan pembahasan sebelumnya, sistem mampu menghasilkan nilai entropy dan gain untuk menghasilkan pohon keputusan atau rule. ditampilkan oleh sistem secara detail, dan juga hasil perhitungan manual. Adapun saran untuk pengembangan sistem ini kedepannya yaitu sistem ini masih menggunakan metode confusion matrix untuk menghitung nilai akurasi sehingga dapat dikembangkan menggunakan dua metode.

DAFTAR PUSTAKA

[1] Kristanto, Andi. (2018). Perancangan Sistem Informasi dan Aplikasi. GAVA MEDIA. Yogyakarta.
 [2] Trisyanto. (2017). Analisis & Perancangan Sistem Basis Data. Surabaya: CV. Garuda Mas Sejahtera.
 [3] Amin, S., Adahati, A., & Inyomusi, E. I. E. (2020). Penerapan Algoritma Data Mining C4. 5 Untuk Klasifikasi Kualitas Layanan Publik Kantor Dinas Kependudukan dan Pencatatan Sipil Kabupaten Manokwari Selatan. DINAMIS, 17(1. Juli), 46-59.
 [4] Yanto, Robi. (2016). Manajemen Basis Data Menggunakan MySQL. Yogyakarta: Deepublish.