

PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS DI UNIVERSITAS PERADABAN)

Ade Fatma Ayu Rahman¹, Sorikhi², Sri Wartulas³

¹Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Peradaban,

²Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Peradaban,

³Program Studi Teknik Elektro, Fakultas Sains dan Teknologi, Universitas Peradaban.

¹adhefatma21@gmail.com, ²soric2001@gmail.com, ³swartulas@gmail.com

Jl. Raya Pagojengan KM 03 Paguyangan Brebes

Abstrak

Kata Kunci:

*Algoritma C4.5,
Data Mining,
Decision Tree*

Mahasiswa merupakan parameter penting dalam penyelenggaraan program studi, banyak mahasiswa yang tingkat kelulusannya berbeda, tepat waktu dan tidak tepat waktu. Hal ini menjadi kendala Universitas pada umumnya karena jumlah kelulusan setiap tahun menjadi salah satu faktor yang akan dinilai untuk mengajukan akreditasi. Penelitian ini bertujuan untuk mengetahui tingkat akurasi algoritma C4.5 dalam prediksi kelulusan mahasiswa, menggunakan metode decision tree algoritma C4.5. Data mahasiswa yang digunakan sebagai atribut adalah IPK, SKS, Umur, Jenis Kelamin. Hasil analisis pada penelitian ini menunjukkan bahwa algoritma C4.5 dapat memprediksi dengan nilai accuracy sebesar 88,74% precision sebesar 91,79%, dan recall sebesar 95,34%.

Abstract

Abstract:

*C4.5 algorithm, Data
Mining,
DecisionTree*

Students are an important parameter in implementing a study program, many students have different graduation rates, are on time and are not on time. This is an obstacle for the University in general because the number of graduations each year is one of the factors that will be assessed for applying for accreditation. This study aims to determine the level of accuracy of the C4.5 algorithm in predicting student graduation, using the C4.5 decision tree algorithm method. Student data used as attributes are IPK, credits, age, gender. The results of the analysis in this study indicate that the C4.5 algorithm can predict with an accuracy value of 88.07% with a precision of 91.79% and a recall of 95.34%.

Pendahuluan

Mahasiswa merupakan parameter penting dalam penyelenggaraan program studi yang berkaitan dengan prestasi, kompetensi dan presensi mahasiswa seharusnya mendapatkan perhatian yang lebih serius dalam evaluasi kinerja mahasiswa. Bentuk kinerja yang baik sebagai mahasiswa adalah dengan lulus tepat waktu dan tidak terancam sanksi drop out atau mengundurkan diri. Peraturan Menteri Ristek Dikti nomor 44 tahun 2005 menyebutkan standar proses pembelajaran maksimum dalam masa studi adalah 7 tahun, IPK diatas 2.0, minimum SKS adalah 144 untuk program Sarjana [1]. Tingginya tingkat persaingan pada dunia Pendidikan membuat setiap individu maupun perguruan tinggi terus mengembangkan kompetensi bakat kemampuannya (skill). Pentingnya kemampuan mahasiswa menjadi tolak ukur keberhasilan mahasiswa itu sendiri.

Sebagaimana halnya yang terjadi pada perguruan tinggi yang peneliti lakukan di Universitas Peradaban Bumiayu, peneliti menemukan masalah mahasiswa tidak lulus tepat waktu yang tidak sesuai standar pada masa studi. Hal ini dikarenakan mahasiswa mengikuti aktivitas kampus, dan banyak mahasiswa yang beranggapan bahwa kuliah hanya ingin mendapatkan ijazah sarjana saja, namun disisi lain, perguruan tinggi menekankan kepada setiap mahasiswa harus memiliki skill atau kemampuan dalam bidang yang diambilnya. Faktor lain yang membuat mahasiswa tidak lulus tepat waktu seperti sibuk bekerja, banyak mengulang mata kuliah, masalah dengan kampus, dan menjadi menjadi aktivis atau

kegiatan kemahasiswaan [2]. Hal tersebut menjadi kendala Universitas pada umumnya karena banyaknya mahasiswa yang lulus tidak tepat waktu. Jumlah kelulusan setiap tahun menjadi salah satu faktor yang dinilai ketika fakultas atau program studi Universitas mengajukan akreditasi [3].

Mengatasi hal tersebut perlu ada teknik yang bisa melakukan prediksi terhadap kelulusan mahasiswa. Teknik yang sering digunakan adalah dengan data mining. Data mining merupakan proses dalam menganalisa data dari sudut pandang yang berbeda dan membentuknya menjadi informasi yang berguna sehingga pengguna dapat memahami substansi hubungan antar data. Data mining memiliki beberapa metode salah satunya ialah metode klasifikasi merupakan teknik data mining yang memetakan data ke kelompok yang telah ditetapkan. Metode klasifikasi dianggap penting karena metode ini dapat memudahkan pengelompokan terhadap data yang akan diolah.[4]. Penggunaan metode klasifikasi dapat memanfaatkan beberapa algoritma di antaranya algoritma C4.5, yang terjadi proses pengklasifikasian data dalam bentuk tingkatan yang berbeda mulai dari akar ke daun secara hirarki, proses ini berlanjut hingga mencapai node yang tidak dapat dibagi lagi.

Algoritma C4.5 digunakan karena dalam algoritma klasifikasi data dengan teknik pohon keputusan dapat mengolah data numerik dan diskrit, dapat menangani nilai atribut yang hilang (missing value), melanjutkan data (continue data), dan menghasilkan aturan-aturan yang mudah diinterpretasikan. Pada algoritma C4.5 dapat melakukan prediksi dengan memberikan nilai akurasi, dengan menerapkan teknik ini akan dibangun pohon keputusan (decision tree) untuk menentukan kelulusan mahasiswa di Universitas Peradaban.

Pada penelitian sebelumnya metode pohon keputusan yang dibangun dengan algoritma C4.5 dalam memprediksi kelulusan mahasiswa Prodi Teknik Informatika dengan kriteria yang digunakan adalah jenis kelamin, asal daerah, IPK dan TOEFL. Atribut yang paling dominan dalam kelulusan mahasiswa adalah IPK, kedua TOEFL, ketiga asal daerah, yang keempat adalah jenis kelamin. Dan menghasilkan nilai rata-rata precision 63.93%, recall 60.73% dan akurasi 60.52% [5]. Berdasarkan pengujian pada penelitian tersebut, algoritma C4.5 memiliki performa yang lebih baik dan menghasilkan nilai akurasi yang besar.

Berdasarkan penjelasan di atas, maka peneliti tertarik untuk melakukan penelitian yang berjudul "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus Di Universitas Peradaban)". Penelitian tersebut bertujuan untuk mengetahui tingkat akurasi algoritma C4.5 dalam memprediksi kelulusan mahasiswa di Universitas Peradaban sehingga mampu mendukung operasional Universitas yang diharapkan dapat tercapai.

Landasan Teori

A. Data Mining

Data mining yaitu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi informasi pengetahuan potensial berguna yang bermanfaat yang tersimpan dalam *database* besar [1].

B. Pohon Keputusan (Decision Tree)

Pohon keputusan atau *decision tree* adalah proses mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*. Pohon keputusan tersebut hasil dari proses perhitungan *entropy* dan *information gain* [6].

C. Algoritma C4.5

Algoritma C4.5 adalah salah satu metode untuk membuat *decision tree* berdasarkan *training* data yang telah disediakan. algoritma C4.5 merupakan pengembangan dari ID3 [7].

D. Entropy dan Information Gain

Entropy adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *imurity* dan *homogeneity* dari kumpulan data. *Information gain* adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada *tree* [8].

E. Confusion Matrix

Confusion matrix merupakan sebuah tabel yang terdiri atas banyaknya baris data uji yang diprediksi benar (*positive*) dan tidak benar (*negative*) oleh model klasifikasi [9].

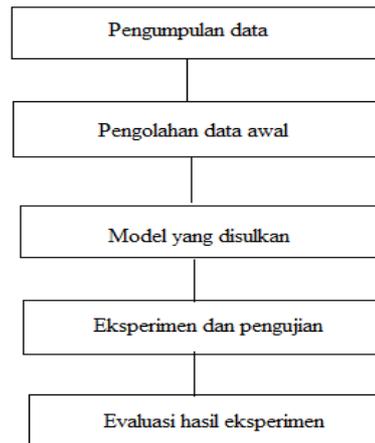
F. Python

Python merupakan bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Fitur pada *python* sebagai bahasa pemrograman dinamis lainnya, *python* umumnya digunakan sebagai bahasa *script* meski pada praktiknya penggunaan bahasa

ini lebih luas mencakup konteks pemanfaatan yang umumnya dilakukan dengan bahasa *script*. *Python* dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan diberbagai platform sistem operasi [10].

Metode Penelitian

Metode penelitian dapat diartikan sebagai prosedur/tahapan dalam menyelesaikan suatu permasalahan yang dibahas pada penelitian ini. Metode penelitian ini akan menghasilkan sebuah alur atau tahapan yang akan dijadikan sebagai acuan dalam pelaksanaan penelitian, mulai dari pengumpulan data sampai pengujian sistem yang telah dibuat. Tahap-tahap penelitian yang digunakan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Metode penelitian.

A. Pengumpulan Data

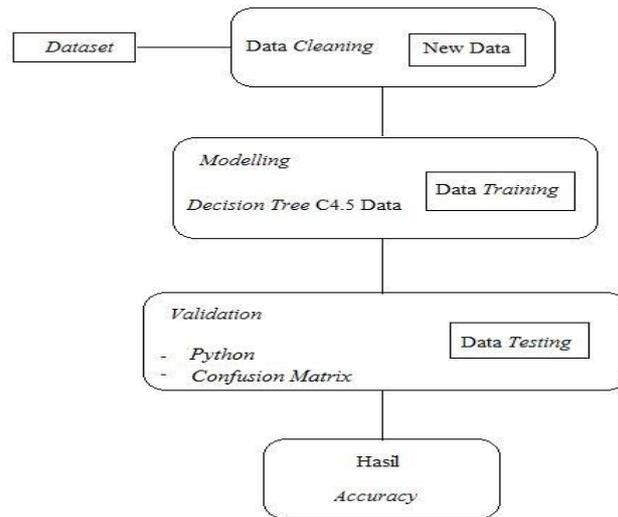
Pengumpulan data dilakukan dengan cara meminta data mahasiswa yang sudah lulus Tahun 2016 di Universitas Peradaban, data yang dipakai yaitu data kelulusan jurusan Pendidikan Guru Sekolah Dasar (PGSD). Data ini mencakup beberapa variabel, antara lain Nim, Nama, Jenis Kelamin, Asal Daerah, Prodi, SKS, IPK dan Umur, yang akan diolah menggunakan algoritma C4.5.

B. Pengolahan Data

Tahap pengolahan data awal ini yaitu menetapkan atribut sebagai variabel keputusan dan *dataset* yang masih berisi informasi pendukung perlu dilakukan pengolahan dengan membersihkan data informasi pendukung tersebut. Transformasi data pada masing-masing *dataset* dilakukan ke dalam bentuk yang dibutuhkan pada penelitian ini yaitu menggunakan metode klasifikasi. Berikut ini salah satu *dataset* kelulusan mahasiswa yang telah ditransformasikan dengan menghilangkan informasi pendukung seperti Nama, Asal Daerah dan Prodi yang akan diwakili oleh atribut Nim sehingga informasi yang dibutuhkan untuk selanjutnya dilakukan proses klasifikasi adalah *record* dan atribut dari IPK, SKS, Umur dan Jenis Kelamin.

C. Metode Yang Diusulkan

Metode yang diusulkan pada penelitian ini adalah eksperimen dan tahap-tahap penelitian ini dapat dilihat pada Gambar 2.



Gambar 2. Metode yang diusulkan

1. Mengumpulkan data mahasiswa (*Dataset*)
2. Melakukan proses *cleaning* atau pembersihan data pendukung yang tidak dipakai kemudian menentukan data yang akan digunakan sebagai atribut untuk melakukan perhitungan sehingga menghasilkan *dataset* baru.
3. Perhitungan algoritma C4.5 pada *modelling decision tree*, proses pembagian *dataset* dibagi menjadi 2 yaitu data *training* dan data *testing*
4. *Dataset* diolah menggunakan *python* menggunakan algoritma C4.5 dengan format data *cvs* untuk memperoleh pohon keputusan atau *rule* dan untuk mengetahui nilai akurasi.

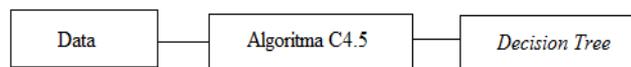
D. Eksperimen dan Pengujian

1. Menyiapkan *dataset* kelulusan mahasiswa
2. Melakukan pengolahan data awal untuk menghasilkan *dataset* baru
3. Dari hasil *dataset* yang baru diolah menggunakan algoritma C4.5 untuk menghasilkan suatu model atau *rule*.
4. Dalam perhitungan algoritma C4.5 proses pembagian *dataset* dibagi menjadi dua *subset* yaitu data *training* dan data *testing*. Data *training* untuk pembentukan model dan data *testing* untuk pengujian model.
5. Melakukan evaluasi menggunakan *confusion matrix*.

E. Evaluasi dan Hasil

Tahap ini model yang terbentuk akan diuji dengan *confusion matrix* untuk mengetahui tingkat akurasi. *Confusion matrix* akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar, prediksi positif salah, prediksi *negative* yang benar dan prediksi *negative* yang salah.

PERHITUNGAN / PENGOLAHAN DATA MENGGUNAKAN ALGORITMA C4.5



Gambar 3. Alur Pengolahan Data

a. Data

Menyiapkan data *training* untuk proses perhitungan yang sudah dikelompokkan dalam kelas-kelas tertentu.

b. Algoritma C4.5

- 1) Memilih atribut sebagai akar, berdasarkan nilai *gain* tertinggi dari atribut-atribut yang ada.
- 2) Menentukan akar dari pohon. Akar tersebut diambil dari atribut, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* tertinggi akan dijadikan akar pertama. Sebelum menentukan nilai *gain* hitung terlebih dahulu nilai *entropy* menggunakan rumus sebagai berikut :

$$Entropy(S) = -\sum_{j=1}^k p_j \log_2 p_j$$

Keterangan :

S = kumpulan data

k = banyaknya kelas dalam S

p_j = probabilitas kelas S

3) Menentukan nilai *gain*, menggunakan rumus sebagai berikut :

$$Gain(A) = entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} * entropy(S_i)$$

Keterangan :

S = kumpulan data

A = atribut

A_i = nilai atribut ke-i

$|S_i|$ = jumlah data untuk A_i

$|S|$ = jumlah data dalam S

k = jumlah nilai atribut

4) Ulangi langkah ke-2 hingga semua kasus dalam cabang memiliki kelas yang sama, ulangi langkah ke-3 untuk masing-masing cabang kasus sampai tidak bisa lagi dilakukan proses perhitungan.

c. Decision Tree

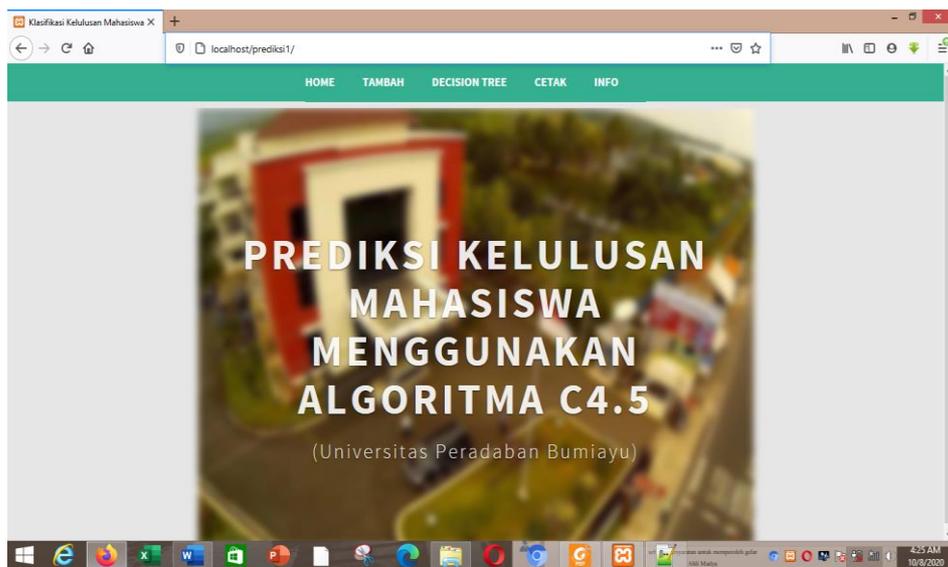
Setelah semua proses perhitungan selesai maka proses selanjutnya adalah pembentukan pohon keputusan sehingga mendapatkan sebuah *rule*.

Hasil dan Pembahasan

Hasil dan pembahasan berisi tentang pembahasan pengujian atau hasil apa yang telah dicapai dari penelitian ini. Berikut adalah hasil yang dicapai dari penelitian ini :

A. Hasil Tampilan Sistem

1) Tampilan Menu Home



Gambar 4. Tampilan menu *home*.

2) Tampilan Menu Tambah Data

Tambah Data Mahasiswa

NIM

Nama Mahasiswa

Jenis Kelamin

Umur

IPK

SKS

Gambar 5. Tampilan menu tambah data

3) Tampilan Menu Edit Data Mahasiswa

Edit Data Mahasiswa

NIM
40211162

Nama Mahasiswa
Yadi Subakti

Jenis Kelamin
L

Umur
Tua

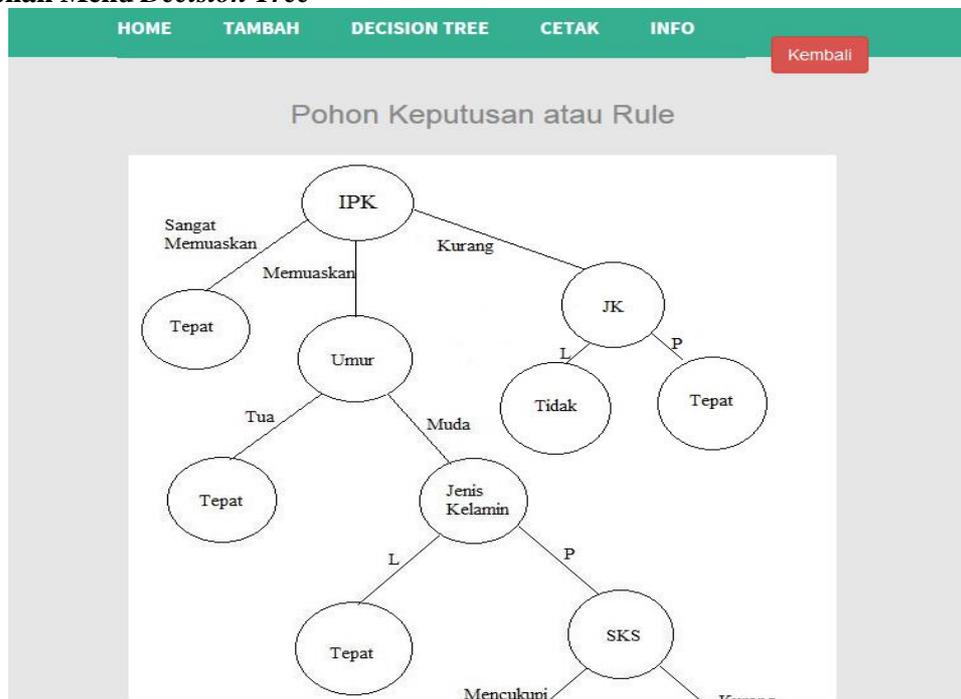
IPK
1.90

SKS
Kurang

Edit Kembali

Gambar 6. Tampilan menu edit data mahasiswa

4) Tampilan Menu Decision Tree



Gambar 7. Tampilan menu decision tree

5) Tampilan Menu Cetak Data

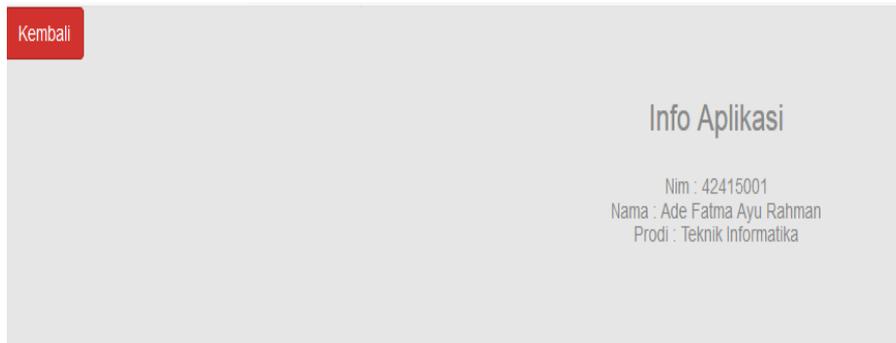
HOME TAMBAH DECISION TREE CETAK INFO

Laporan Data Kelulusan Mahasiswa Universitas Peradaban

No	Nim	Nama Mahasiswa	IPK	SKS	Umur	Jenis Kelamin	Hasil
1	40211162	Yadi Subakti	1.90	Kurang	Tua	L	Tidak Tepat
2	40212147	Dayaturrohman	3.93	Mencukupi	Muda	L	Tepat

Gambar 8. Tampilan menu cetak data

6) Tampilan Menu Informasi



Gambar 9. Tampilan menu informasi

B. Hasil Evaluasi

Dalam tahap pengujian dilakukan untuk mengetahui tingkat akurasi dari algoritma C4.5 dengan menggunakan *dataset* 151 data mahasiswa, penelitian ini *gain* tertinggi terletak pada IPK, sehingga menjadi *root* pada pohon keputusan, kemudian Umur, Jenis Kelamin, dan SKS.

Uji coba menggunakan *python* dilakukan sebanyak tiga kali dengan jumlah data sampel yang berbeda-beda. Pertama menggunakan 84 data mahasiswa memperoleh akurasi 82.14%, uji coba kedua menggunakan 115 data mahasiswa memperoleh akurasi 86.95%, dan uji coba ketiga menggunakan 151 data mahasiswa memperoleh akurasi 88.74%.

Untuk mengetahui kinerja algoritma C4.5 dalam melakukan klasifikasi terhadap kelas yang telah ditentukan dalam penelitian, untuk menghitung nilai akurasi dilakukan dengan menggunakan *confusion matrix*, penelitian ini perhitungan nilai akurasinya menggunakan 151 *dataset* mahasiswa menghasilkan nilai akurasi sebagai berikut :

Tabel 1. Hasil Perhitungan *Confusion Matrix*

Akurasi	<i>Precision</i>	<i>Recall</i>
88,74%	91,79%	95,34%

Kesimpulan dan Saran

Kesimpulan

Berdasarkan hasil penelitian yang dilakukan oleh peneliti, maka dapat disimpulkan bahwa kelulusan mahasiswa dapat diprediksi dan dievaluasi dengan memanfaatkan Teknik *data mining* menggunakan algoritma *decision tree* C4.5 untuk memprediksi (menentukan kelas) dari data *training* yang telah diperoleh. Pada penelitian ini uji coba dilakukan sebanyak tiga kali dan menghasilkan nilai akurasi yang berbeda-beda. Uji coba pertama menggunakan 84 data mahasiswa memperoleh nilai akurasi sebesar 82.14%. Uji coba kedua menggunakan 115 data mahasiswa memperoleh nilai akurasi sebesar 86.95%, dan uji coba ketiga menggunakan 151 data mahasiswa memperoleh nilai akurasi sebesar 88.74%.

Saran

Saran dari penulis untuk penelitian lebih lanjut agar mampu meningkatkan kinerja dan menyempurnakan penelitian yang telah dibuat yaitu:

- Pada penelitian selanjutnya jumlah *record* data yang digunakan untuk proses *training* sebaiknya ditingkatkan lagi karena jumlah data *training* mempengaruhi nilai akurasi, kemudian pemilihan atribut yang akan digunakan dalam melakukan prediksi juga sangat berpengaruh.
- Penelitian ini perlu dilakukan penelitian lebih lanjut untuk membandingkan hasil prediksi, melakukan perbandingan pada algoritma lain agar dapat menghasilkan prediksi yang lebih baik.

Referensi

- [1] Y.D. d. Atma, "Perbandingan Algoritma C4.5 dan K-NN Dalam Identifikasi Mahasiswa Berpotensi Drop Out," *Metik Jurnal*, vol.2, No.2, 2018.
- [2] Risnawati, "Analisis Kelulusan Mahasiswa Menggunakan Algoritma C4.5," *Jurnal Mantik Penusa*, vol. 2, No.1, 2018.
- [3] I. P. Astuti, "Prediksi Ketetapan Waktu Kelulusan Dengan Algoritma Data Mining C4.5," *Fountain Of Informatic Journal*, vol.2, No.2, 2017.
- [4] E. P. Rohmawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network," *Jurnal Ilmiah Matrik*, vol.20, No.01, 2018.
- [5] R. P. d. I. P. Sari, "Penerapan Algoritma C4.5 Pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Teknik Informatika," *Ilmu Komputer Dan Informatika*, vol.4, No.1, 2018.
- [6] P. Harmianty, "Aplikasi Prediksi Kelulusan Tepat Waktu Mahasiswa Menggunakan Algoritma C4.5," *Skripsi*, 2017.
- [7] P. K. O. Elsa, dkk, "Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *Jurnal Sistem dan Teknologi Informasi*, vol.7, No.4, 2019.
- [8] R. H. Laksamana, dkk, "Analisa Prediksi Mahasiswa Drop Out Menggunakan Metode Decision Tree Dengan Algoritma ID3 dan C4.5," *Transformatika*, vol.17, No.2, pp 97-106, 2019.
- [9] A. d. Romadhona, "Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree," *Teknologi Informasi*, vol.13, No.1, 2017.
- [10] A. N. d. T. K. Syahrudin, "Input Output Pada Bahasa Pemrograman Python," *Dasar Pemrograman STIMIK*, 2018.