

Penerapan Model *Algoritma C4.5* dengan *Tool Weka* Untuk Memprediksi Kelulusan Mahasiswa

Daryanto¹, Rahma Wahyuningsih², Miftahur Rahman³

^{1,2,3}*Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember*
Email: ¹*daryanto@unmuhjember.ac.id*, ²*rahma.wahyu@gmail.com*,
³*miftahurrahman@unmuhjember.ac.id*

(Naskah masuk: 5 Agustus 2022, diterima untuk diterbitkan: 26 Agustus 2022, Terbit: 28 Agustus 2022)

ABSTRAK

Fakultas Teknik Universitas Muhammadiyah Jember merupakan Fakultas yang memiliki jumlah mahasiswa yang cukup banyak pada setiap tahunnya, akan tetapi mahasiswa yang dapat lulus tepat waktu per-tahun kurang dari 50%, sedangkan jumlah mahasiswa baru pada setiap tahunnya semakin meningkat. Sehingga mengakibatkan penumpukan jumlah mahasiswa yang belum lulus sesuai dengan masa periode kelulusannya. Berdasarkan latar belakang tersebut penelitian ini dilakukan untuk membuat sebuah *system* menggunakan metode klasifikasi yang dapat memprediksi persentase kelulusan mahasiswa. Sistem ini menggunakan metode C4.5. Pada penelitian ini menggunakan data Mahasiswa Fakultas Teknik Tahun Akademik 2016/2017 sampai dengan 2018/2019 sebanyak 1220, terdiri dari 5 Prodi yaitu: Teknik Elektro 132, Teknik Mesin 176, Manajemen Informatika 42, Teknik Sipil 388 dan Teknik Informatika 482. Dari jumlah mahasiswa tersebut digunakan sebanyak 50% untuk dijadikan *dataset*, kemudian dihitung menggunakan *weka explorer* versi 3.8.5. Inputan yang digunakan sebagai pengujian berupa atribut dari data mahasiswa yang meliputi Jenis Kelamin, Asal Sekolah, Jalur Masuk, *Grade* IPS dan lulus tepat waktu. Hasil perhitungan *weka* Fakultas Teknik didapatkan persentase akurasi 96%, presisi 96% dan *recall* 4%. Dari perhitungan yang dilakukan peneliti memprediksi tingkat kelulusan mahasiswa Fakultas Teknik adalah 70% lulus tepat waktu dan 30% tidak lulus tepat waktu.

Kata kunci: Prediksi, Klasifikasi, *Algoritma C4.5*, *Weka V 3.8.5*

ABSTRACT

The Faculty of Engineering, Muhammadiyah University of Jember is a faculty that has a large number of students each year, but less than 50% of students can graduate on time per year, while the number of new students is increasing every year. This results in a buildup of the number of students who have not graduated in accordance with their graduation period. Based on this background, this research was conducted to create a system using a classification method that can predict the percentage of student graduation. This system uses the C4.5 method. In this study, 1220 students of the Engineering Faculty Academic Year 2016/2017 to 2018/2019 were used, consisting of 5 study programs, namely: Electrical Engineering 132, Mechanical Engineering 176, Information Management 42, Civil Engineering 388 and Informatics Engineering 482. Of the total number of students 50% is used as a dataset, then calculated using Weka Explorer version 3.8.5. The input used as a test is in the form of attributes from student data which include Gender, School Origin, Entry Path, Social Studies Grade and graduating on time. The results of the calculation of the Weka Faculty of Engineering obtained the percentage of accuracy of 96%, precision of 96% and recall of 4%. From the calculations carried out, the researcher predicts that the graduation rate of students from the Faculty of Engineering is 70% graduating on time and 30% not graduating on time.

Keywords: Prediction, Classification, Algorithm C4.5, Weka V 3.8.5

1. PENDAHULUAN

Mahasiswa merupakan salah satu elemen dalam sebuah universitas atau perguruan tinggi. Setiap mahasiswa memiliki data yang beragam, mulai dari informasi identitas diri seperti asal daerah, jenis sekolah asal (SMA, SMK, MA, dan MAK), jenis kelas tempat siswa belajar dan lain sebagainya. Data akademik perguruan tinggi yang telah disimpan dapat digunakan sebagai bahan analisis antara lain untuk mengetahui hubungan prestasi dan sekolah asal, SKS yang diambil dengan topik peminatan, jenis sekolah asal dengan ketepatan waktu kelulusan dan begitu seterusnya (Astuti, 2017).

Secara umum jumlah mahasiswa pada jenjang kelulusan bervariasi, ada yang tepat waktu dan ada yang tidak tepat waktu. Permasalahan secara umum di Fakultas Teknik Universitas Muhammadiyah Jember khususnya adalah banyaknya mahasiswa yang lulus tidak tepat waktu. Padahal jumlah kelulusan mahasiswa merupakan salah satu faktor penilaian ketika suatu fakultas atau program studi pada suatu universitas mengajukan akreditasi.

Dilihat dari data kelulusan Fakultas Teknik tersebut bahwa dalam kurun waktu 3 tahun, dari tahun akademik 2016/2017 hingga tahun 2018/2019 terdapat 1220 mahasiswa Fakultas Teknik yang berbeda pada setiap tahun ajaran baru. Hal ini berdampak pada bertambahnya jumlah data akademik dari seluruh mahasiswa yang masih terdaftar.

Jumlah mahasiswa baru dan yang lulus tiap tahun tidak seimbang. Oleh karena itu diperlukan suatu sistem untuk memprediksi kelulusan mahasiswa. Sistem ini membutuhkan informasi yang ada untuk dapat mengetahui apakah mahasiswa dapat lulus tepat waktu atau tidak. Dengan

sistem ini diharapkan civitas akademika dapat mengetahui terlebih dahulu persentase mahasiswa yang lulus tiap tahunnya dan dapat menerapkan kebijakan khusus untuk meminimalisir jumlah mahasiswa yang tidak lulus tepat waktu, agar mahasiswa yang masuk dan yang sudah lulus bisa seimbang.

Sistem ini menggunakan metode Data Mining untuk mengelola data yang ada, guna mendapatkan hasil yang diinginkan. Data mining adalah istilah yang digunakan untuk menggambarkan penemuan pengetahuan dalam *database*. Penambangan data adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi yang berguna dan pengetahuan terkait dari berbagai database besar. (Kusrini & Emha Taufiq Luthfi, 2009:25).

Decision tree (algoritma C4.5) merupakan salah satu metode data mining adalah sebuah sistem atau cara yang manusia kembangkan untuk membantu mencari dan membuat keputusan untuk masalah-masalah tersebut dan dengan memperhitungkan berbagai macam faktor yang ada di dalam lingkup masalah tersebut (Basuki, A., & Syarif, I., 2003).

Secara umum, pohon keputusan adalah suatu gambaran permodelan dari suatu persoalan yang terdiri dari serangkaian keputusan yang mengarah ke solusi (Kusrini & Emha Taufiq Luthfi, 2009:25). Pada usulan metode ini diharapkan mampu memprediksi tingkat ketelitian kelulusan mahasiswa sesuai dengan masa perkuliahan yang ditentukan oleh instansi terkait yaitu Universitas Muhammadiyah Jember.

Penelitian terdahulu yang pernah dilakukan oleh Nanda Dimas Prayoga (2018) tentang penerapan algoritma C.45 dalam memprediksi kelulusan tepat waktu

pada perguruan tinggi. Data yang digunakan sebanyak 146 dengan 10 *atribute* yaitu meliputi NIM, nama, jenis kelamin, asal daerah, tanggal lahir, jumlah sks, IPK, predikat, dan lama studi. Hasil pengujian pada tools *Rapidminer* 5 menunjukkan sebuah hasil akurasi dengan tingkat kinerja Algoritma C.45 yaitu 92,60% +/- 1.60%. Sedangkan penelitian yang dilakukan oleh Mujib Ridwan (2017) tentang sistem rekomendasi proses kelulusan mahasiswa menggunakan algoritma C4.5. Data yang digunakan sebanyak 100 dengan 9 *atribute* yaitu meliputi jenis kelamin, asal sekolah, jalur masuk, nilai UNAS, gaji orang tua, IPS, IPK, keterangan lulus, dan nilai matakuliah hasil dari pengujian klasifikasi menggunakan algoritma C4.5 menunjukkan nilai akurasi dalam kategori *fair classification*, hal ini disebabkan karena faktor penentu kelulusan mahasiswa tepat atau tidak tepat waktu pada kenyataannya memiliki nilai yang tidak konsisten. Pengujian pada data training menghasilkan nilai presisi 100%, recall 50%, dan akurasi 75%,

2. METODE PENELITIAN

Tahapan-Tahapan metode yang dilakukan pada penelitian ini adalah sebagai berikut (Rahman, M. 2019):



Gambar 1. Metode Penelitian

a. Identifikasi Masalah

Adalah tahapan yang pertama dalam melakukan penelitian dengan cara menganalisa masalah yang ada dan menawarkan solusi yang diharapkan dapat digunakan dalam menyelesaikan masalah dimaksud.

b. Studi Pustaka

Tahapan membaca, mempelajari baik buku, jurnal, artikel dan lainnya sebagai bahan rujukan/referensi dalam mendukung penelitian ini.

c. Pengumpulan Data

Adapun *dataset* yang digunakan dalam penelitian ini adalah mahasiswa Fakultas Teknik Universitas Muhammadiyah Jember tahun akademik 2016/2017 hingga 2018/2019 dengan total keseluruhan 1220 *record*.

d. Eksperimen dan Pengujian

Tahapan ini menggunakan metode Algoritma C4.5 yang dibantu dengan *tool Weka*

e. Evaluasi

Pada tahapan ini nantinya diperoleh hasil berupa akurasi, presisi, dan *recall*

2.1. Data Mining

2.2. Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari ID3, beberapa pengembangan yang dilakukan pada C 4.5 antara lain mengatasi nilai yang hilang, mampu menangani data lanjutan, dan pemangkasan. Proses dalam pohon keputusan adalah mengubah data (tabel) menjadi model pohon, mengubah model pohon menjadi aturan, dan menyederhanakan aturan. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Dwi Meliani Achmad, Budanis, Slamet. 2012):

1. Pilih atribut sebagai akar.
2. Buat cabang untuk nilai masing-masing.

3. Membagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang hingga semua kasus di cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai *root* berdasarkan nilai perolehan tertinggi dari atribut yang ada. Untuk menghitung profit digunakan rumus seperti di bawah ini (Prayoga, 2018).

$$Gain(S, A) =$$

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

Keterangan:

Gain: Proses perhitungan untuk mendapatkan nilai *Entropy* yang terbesar

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai Gain yaitu mencari nilai Entropy. Entropi digunakan untuk menentukan setiap atribut masukan informatif untuk menghasilkan suatu atribut. Rumus dasar Entropi adalah sebagai berikut (Prayoga 2018).

$$Entropy(S) = \sum_{i=1}^n - p_i * \log^2 p_i \quad (2)$$

Keterangan:

Entropy : Proses perhitungan untuk mengukur ketidakpastian variable acak

S : Himpunan kasus

n : Jumlah partisi S

p_i : Proporsi dari S_i terhadap S

Kemudian menghitung presentase dengan rumus sebagai berikut:

$$Akurasi = \frac{TP+TN}{JML\ DATA} \quad (3)$$

$$Presisi = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Keterangan:

Akurasi: Tingkat kedekatan hasil prediksi dengan hasil fakta.

Presisi: Tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh system.

Recall: Tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

TP (True Positif) ialah data dengan hasil benar dan prediksinya juga benar.

FP (False Positif) ialah data dengan hasil salah namun prediksinya tidak benar.

FN (False Negatif) ialah data dengan hasil salah namun prediksinya salah.

TN (True Negatif) ialah data dengan hasil benar dan prediksinya salah.

2.3. Weka

Weka adalah aplikasi penambahan data *open source* berbasis *Java*. Aplikasi ini pertama kali dikembangkan oleh University of Waikato di Selandia Baru sebelum menjadi bagian dari Pentaho. Weka terdiri dari kumpulan algoritma pembelajaran mesin yang dapat digunakan untuk menggeneralisasi/merumuskan kumpulan data pengambilan sampel. Meski kekuatan Weka terletak pada algoritme yang lebih lengkap dan canggih, data mining masih terletak pada pengetahuan komunitas pelaksananya. Tugas mengumpulkan data dan pengetahuan berkualitas tinggi tentang pemodelan dan penggunaan algoritma yang tepat diperlukan untuk memastikan formulasi yang diharapkan (Aswendy, 2016).

3. HASIL DAN PEMBAHASAN

Hasil pengujian adalah hasil akhir yang berarti penelitian telah selesai dilakukan, semua yang berkaitan dengan pengujian akan dibahas. Mulai dari pengolahan *dataset* hingga pengujian data pada *tool* yang digunakan.

3.1. Analisis Data

Pada penelitian ini menggunakan data Mahasiswa Fakultas Teknik Tahun Akademik 2016/2017 sampai dengan 2018/2019 sebanyak 1220 *record*, terdiri dari 5 Prodi yaitu: Teknik Elektro 132, Teknik Mesin 176, Manajemen Informatika 42, Teknik Sipil 388 dan Teknik Informatika 482, *sample* data ditunjukkan pada Tabel 1. Dari jumlah mahasiswa tersebut digunakan sebanyak 50 % untuk dijadikan *dataset*, kemudian di hitung menggunakan *weka explorer versi 3.8.5*.

Tabel 1. Data Awal (sample)

Angkatan	NIM	Nama	Asal Sekolah	Alamat	Jalur Masuk	Jenis Kelamin	Grade IPS	Lulus & Tidak
2016	1810851063	HAMDHAN ASHARI	SMA	Jember	TES	L	SB	YA
2016	1810851072	ADE YUGO DWINDAVIT	SMK	Sibubondo	PMDK	L	SB	YA
2016	1810851068	MEKIT MAHAROKA	SMK	Jember	TES	L	B	TDK
2016	1810851054	ADINDA PUTRI RENGGANIS	SMA	Jember	TES	P	SB	YA
2016	1810851053	FENI SHOFTYA	SMA	Sibubondo	TES	P	SB	YA
2016	1810851052	YUKA REKSA NUGROHO	SMA	Jember	TES	L	SB	YA
2016	1810851051	TAUFIK TRI WILANARJO	SMK	Jember	TES	L	B	YA
2016	1810851050	REZA CHAIDIR	SMA	Jember	TES	L	B	YA
2016	1810851054	ADINDA PUTRI RENGGANIS	SMA	Jember	TES	P	SB	YA
2016	1810851053	FENI SHOFTYA	SMA	Sibubondo	TES	P	SB	YA
2016	1810851052	YUKA REKSA NUGROHO	SMA	Jember	TES	L	SB	YA

Data training yang digunakan penelitian ini adalah data mahasiswa tahun akademik 2016/2017 sampai dengan 2018/2019, yang mana dalam pemilihan data dilakukan secara acak (*random*). Dengan jumlah data sebanyak 50 % dari 1220 data.

Berikut adalah contoh perhitungan entropy dan gain:

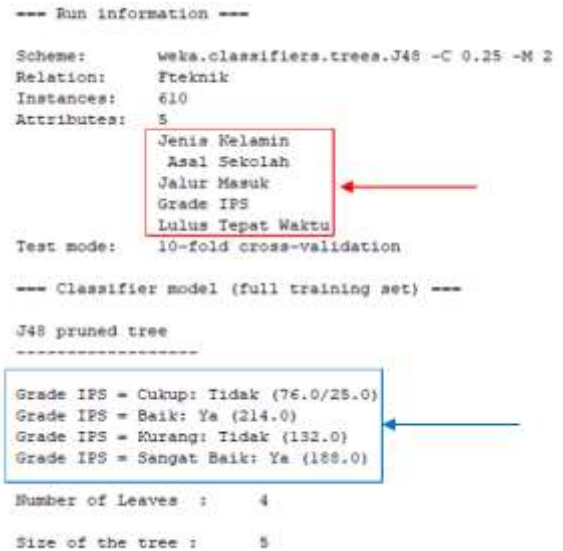
$$\text{Entropy}(\text{Total}) = \left(-\frac{12}{20} * \log_2\left(\frac{12}{20}\right)\right) + \left(-\frac{8}{20} * \log_2\left(\frac{8}{20}\right)\right) = 0.970950594$$

$$\text{Gain}() = (0.970950594) - \left(\frac{16}{20} * 0.988699408\right) - \left(\frac{4}{20} * 0.811278124\right) = 0.01773544$$

Selanjutnya perhitungan dilakukan menggunakan *tool Weka*.

3.2. Pengujian Algoritma dengan Tool Weka

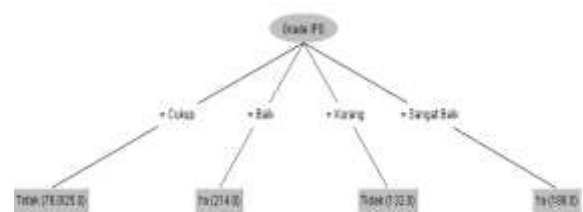
Hasil klasifikasi menggunakan *tool Weka* ditunjukkan pada Gambar 2 yang ditandai dengan panah merah menunjukkan beberapa *atribut* yang ada pada *dataset* uji. Seperti *atribut* Jenis Kelamin, Asal Sekolah, Jalur Masuk, Grade IPS dan Lulus Tepat Waktu. Sedangkan gambar yang ditandai dengan garis biru adalah *Rule* yang tercipta dari hasil klasifikasi prediksi kelulusan mahasiswa menggunakan *Algoritma C4.5* pada *tool Weka*.



Gambar 2. Hasil Pengujian

Adapun *Rule* yang terbentuk berdasarkan klasifikasi prediksi kelulusan mahasiswa adalah sebagai berikut:

- Jika Grade IPS = Cukup, maka Lulus Tepat Waktu = Tidak
- Jika Grade IPS = Baik, maka Lulus Tepat Waktu = Ya
- Jika Grade IPS = Kurang, maka Lulus Tepat Waktu = Tidak
- Jika Grade IPS = Sangat Baik, maka Lulus Tepat Waktu = Ya
- Setelah semua tahap dilakukan, tahap selanjutnya adalah melihat hasil pohon keputusan yang terbentuk dengan menggunakan *algoritma C4.5* yang diterapkan dengan *tool Weka*. Gambar pohon keputusan ditunjukkan pada Gambar 3.



Gambar 3. Hasil Pohon Keputusan

Gambar 3 adalah pohon keputusan (*tree*) yang terbentuk dari proses klasifikasi prediksi kelulusan mahasiswa menggunakan *algoritma C4.5*. Pada pohon

keputusan tersebut atribut yang menjadi cabang akar. Pesrentase yang dihasilkan sebagai berikut:

```

Time taken to build model: 0.01 seconds

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

Correctly Classified Instances      585      95.9216 %
Incorrectly Classified Instances    35       4.9984 %
Kappa statistic                    0.9561
Mean absolute error                0.0438
Root mean squared error            0.2115
Relative absolute error            13.2256 %
Root relative squared error       36.5441 %
Total Number of Instances         610

=== Detailed Accuracy By Class ===

  TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
-----
1.000  0.000  1.000  0.941  0.970  0.910  0.989  0.988  Ya
1.000  0.000  1.000  1.000  1.000  0.936  0.910  0.909  Tidak
Weighted Avg.  0.955  0.018  0.964  0.955  0.960  0.910  0.909  0.988

=== Confusion Matrix ===
  a  b  c ← classified as
427 25 1 ← a = Ya
  183 183 0 ← b = Tidak
    
```

Gambar 4. Hasil Perhitungan dengan Weka

Berikut Tabel 2 adalah hasil persentase secara keseluruhan:

Tabel 2. Hasil Persentase

Dataset	Akurasi	Presisi	Recall
Fakultas Teknik	96%	96%	96%
Teknik Elektro	97%	97%	97%
Teknik Mesin	93%	93%	93%
Manajemen Informatika	100%	100%	100%
Teknik Sipil	94%	95%	94%
Teknik Informatika	96%	97%	96%

3.3. Persentase Kelulusan

Persentase kelulusan berdasarkan perhitungan pada *dataset* Fakultas Teknik dengan Jumlah Kasus 610, [Si] Ya 427, dan [Si] Tidak 183, berikut detail perhitungannya:

$$\text{Lulus} = \frac{427}{610} \times 100 = 70\%$$

$$\text{Tidak Lulus} = \frac{183}{610} \times 100 = 30\%$$

Bahwa *dataset* Fakultas Teknik yang lulus tepat waktu 70% dan yang tidak lulus tepat waktu 30% dari 610 data, hal ini juga dilakukan perhitungan pada data per prodi masing-masing, berikut disajikan pada Tabel 3.

Tabel 3. Prediksi Kelulusan

Dataset	Jumlah Data	Lulus	Tidak lulus
Fakultas Teknik	610	70%	30%
Teknik Informatika	241	66%	34%
Teknik Sipil	194	78%	22%
Teknik Mesin	88	70%	30%
Manajemen Informatika	21	52%	48%
Teknik Elektro	66	67%	33%

4. KESIMPULAN DAN SARAN

Kesimpulan yang dipat disajikan bahwa berdasarkan hasil pengujian klasifikasi menggunakan Algoritma C4.5 pada *dataset* 50% dari 1220 jumlah mahasiswa Fakultas Teknik, peneliti menggunakan perhitungan *weka*. Hasil perhitungan *weka* Fakultas Teknik didapatkan presentase akurasi 96%, presisi 96% dan recall 96% Dari perhitungan yang dilakukan peneliti memprediksi tingkat kelulusan mahasiswa Fakultas Teknik adalah 70% lulus tepat waktu dan 30 % tidak lulus tepat waktu.

Saran yang dapat diberikan sebagai *future plan* terhadap penelitian ini adalah bagi peneliti-peneliti selanjutnya yang akan melakukan penelitian hampir serupa: (1) Mencoba menggunakan *tool* selain *Weka* dalam analisa data dan mencoba menggunakan algoritma lain selain *algoritma C4.5*. (2) Mencoba lebih banyak *record, attribute* dan *parameter* dalam pemrosesan data.

DAFTAR PUSTAKA

Astuti, I. P. (2017). Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5. Fountain of Informatics Journal, 2(2), 5. <https://doi.org/10.21111/fij.v2i2.106>

- Aswendy. 2016. "Analisis Data Iklim Indonesia Menggunakan Aplikasi Weka Dengan Metode Klasifikasi." *Teknologi Rekayasa* 21 (3): 217–28.
- Basuki, A., & Syarif, I. (2003). *Modul Ajar Decision Tree*. Surabaya: PENS-ITS.
- Dwi Meliani Achmad, Budanis, Slamet, Fauzi. 2012. "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree." *Jurnal IPTEK* 16 (1): 18–23.
- Prayoga, Nanda Dimas (2018). "Penerapan Algoritma C.45 Dalam Memprediksi Kelulusan Tepat Waktu Pada Perguruan Tinggi (Studi Kasus : Stmik Royal Kisaran)." <https://doi.org/10.31227/osf.io/unqt4>
- Kusrini, dan Emha Taufiq Luthfi. 2009. *Algoritma data Mining*. Yogyakarta: Andi Publisher.
- Rahman, M. (2019). Prediksi Pembayaran Tagihan Listrik Menggunakan Model Artificial Neural Network. *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, 4(1), 7-12.
<https://doi.org/10.32528/justindo.v4i1.2417>
- Ridwan, M. (2017). Sistem Rekomendasi Proses Kelulusan Mahasiswa Berbasis Algoritma Klasifikasi C4.5. *Jurnal Ilmiah Informatika*. 2(1), 105-111.
- Tool Weka. (2012). *Data Mining Menggunakan Weka*. di akses [<http://www.erdisusanto.com/2012/06/data-mining-menggunakan-weka.html>] (tanggal 29 Desember 2016).
- Tool: <http://www.cs.waikato.ac.nz/ml/weka/> (akses tool weka).