

## Penerapan Algoritma *Decision Tree* dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa

M Riski Qisthiano<sup>1✉</sup>, Putri Armilia Prayesy<sup>2</sup>, Istiana Ruswita<sup>3</sup>

<sup>1,2,3</sup> Program Studi Ilmu Komputer, Institut Teknologi dan Bisnis Nasional (ITBN)

### Informasi Artikel

#### Riwayat Artikel

**Diserahkan** : 07-11-2022

**Direvisi** : 18-11-2022

**Diterima** : 22-11-2022

### ABSTRAK

Dalam melakukan proses klasifikasi terhadap prediksi kelulusan mahasiswa, ada banyak faktor dan kriteria dalam mengukur kelulusan mahasiswa tersebut, serta menentukan mahasiswa tersebut tepat atau tidaknya menyelesaikan studi. Oleh sebab itu, maka dibutuhkan suatu metode klasifikasi untuk melakukan pengukuran terhadap data prediksi kelulusan tepat waktu, penulis menggunakan *dataset* yang berasal dari beberapa perguruan tinggi tersebar di Kota Palembang. Model yang digunakan ini menggunakan *Decision Tree* yang berfungsi sebagai salah satu metode untuk melakukan klasifikasi. *Dataset* yang digunakan adalah data alumni yang sudah dikumpulkan berasal dari perguruan tinggi di Kota Palembang, sedangkan kriteria untuk melakukan proses klasifikasi adalah jurusan, perguruan tinggi setiap mahasiswa, jenis kelas pilihan, dan nilai setiap mahasiswa yang diambil mulai dari semester awal sampai semester ke 4, lalu ada data tahun kelulusan mahasiswa tersebut, beserta data tahun masuk dari mahasiswa tersebut. Setelah peneliti menentukan atribut data yang akan menjadi bagian dari proses klasifikasi, peneliti memilih menggunakan alat bantu Rapidminer dalam mengelola data klasifikasi dengan model *decision tree*. Proses berikutnya pengujian menggunakan 5 kali proses uji *K-Fold Validation* dengan membagi *dataset* ke dalam *training* dan *testing*. Hasil penelitian ini merupakan akurasi dari hasil klasifikasi terhadap prediksi yang didapat dari alat bantu Rapidminer dan metode *Decision Tree* yang memiliki hasil akurasi sebesar 87.93%.

### Kata Kunci:

Klasifikasi, Kelulusan, Mahasiswa, *Decision Tree*.

### Keywords :

*Classification, Graduation, Student, Decision Tree.*

### ABSTRACT

*In carrying out the classification process for student graduation predictions, there are many factors and criteria in measuring the student's graduation and determining whether the student is right or not to complete the study. Therefore, a classification method is needed to measure the predicted graduation data on time, the authors use datasets from several universities spread across the city of Palembang. The model used uses a Decision Tree which serves as a model for classifying. The dataset used is alumni data that has been collected from universities in the city of Palembang, while the criteria for carrying out the classification process are the majors, each student's university, the type of class chosen and the grades of each student taken from the first semester to the fourth semester, then there is data on the year the student graduated, along with data on the year the student entered. After the researcher determines the data attributes that will be part of the classification process, the researcher chooses to use the rapidminer tool in managing classification data with the decision tree model. The next process the examiner uses 5 times the K-Fold Validation test process by dividing the dataset into training and testing. The results of this study are the accuracy of the classification results against the predictions obtained from the Rapidminer tool and the Decision Tree model which has an accuracy of 87.93%.*

---

**Corresponding Author :**

M Riski Qisthiano

Program Studi Ilmu Komputer, Institut Teknologi dan Bisnis Nasional (ITBN)

Jl. Cahaya Berlian No.81 Pangkalan Balai, Kecamatan Banyuasin III, Banyuasin, Sumatera Selatan

Email: [thiano72@gmail.com](mailto:thiano72@gmail.com)

---

---

**PENDAHULUAN**

Ada banyak cara dalam menentukan setiap mahasiswa dapat menyelesaikan studi, baik yang sesuai jadwal dan tepat waktu, atau tidak tepat waktu, bahkan melewati waktu yang sudah dibatasi untuk setiap mahasiswa. Proses penyelesaian mahasiswa secara tepat waktu merupakan bagian dari penilaian untuk perguruan tinggi (Rohmawan, 2018). Banyaknya mahasiswa yang tidak menyelesaikan studi dengan tepat waktu, membuat metrik lulusan perguruan tinggi menjadi semakin tidak relevan. Setelah mencapai APS 4.0 dalam hal standar kelulusan, mahasiswa memiliki standar kelulusan 50% atau lebih dari prestasi mahasiswa (Kinasih et al., 2021).

Penelitian yang berjudul implementasi *data mining* dengan menggunakan algoritma *Decision Tree C4.5* untuk prediksi kelulusan mahasiswa di Universitas Pandaran menggunakan data latih sebanyak 170 dataset. Penelitian tersebut menghasilkan pola/model kelulusan mahasiswa yang dapat dimanfaatkan untuk sebuah keputusan di perguruan tinggi. Hasil dari penelitian tersebut menghasilkan 10 *rules* dengan nilai akurasi 65,98%, nilai AUC 0,874, dan dapat dikategorikan sebagai klasifikasi data yang baik. Maka hasil tersebut sangat penting untuk dijadikan pengambilan keputusan dalam lembaga (Rohman & Rufiyanto, 2019).

Sedangkan penelitian yang berjudul penerapan metode klasifikasi *Decision Tree* untuk memprediksi kelulusan tepat waktu menggunakan data dari atribut *cluster* asal daerah berdasarkan PDRB, program studi, jalur masuk, asal sekolah, bidang keahlian, IPK semester ganjil 2019/2020, dan status kelulusan TA 1 dan seminar proposal. Sumber data pada penelitian tersebut diperoleh dari bagian Penerimaan Mahasiswa Baru (PMB), Akademik Institusi, dan SISFO. Metode yang digunakan pada penelitian tersebut adalah metode klasifikasi *decision tree*. Penelitian ini menghasilkan 63 *rules* dengan akurasi 68,49%, dengan nilai *precision* 79,63%, dan nilai *recall* 55,13% (Ma'sum et al., 2021).

Pada penelitian yang berjudul penerapan algoritma *Support Vector Machine* (SVM) untuk model prediksi kelulusan mahasiswa tepat waktu, hasil pengujian dengan jumlah data *training* sebanyak 90%, dan data *testing* sebanyak 10% menunjukkan bahwa algoritma SVM memberikan nilai akurasi yang sangat baik yaitu 94.4 (Haryatmi & Pramita Hervianti, 2021).

Pemecahan masalah yang akan dilakukan penulis adalah dengan melakukan klasifikasi terhadap data prediksi kelulusan mahasiswa dengan metode klasifikasi *Decision Tree*. Metode tersebut merupakan salah satu algoritma klasifikasi yang banyak digunakan untuk memperoleh hasil klasifikasi non-biner. Dibanding algoritma sejenis, *Decision Tree 4.5* memiliki kelebihan pada kemampuan untuk mengelola data dalam berbagai format (Supangat et al., 2018).

Penelitian yang berjudul prediksi ketepatan waktu kelulusan dengan algoritma *data mining* C4.5 menggunakan data mahasiswa dengan atribut *input* diantaranya adalah jenis sekolah asal, daerah asal, pekerjaan orang tua, dan kelas. Atribut *output* yang digunakan untuk mengklasifikasikan data mahasiswa adalah status, yang terdiri dari “lulus tepat waktu” dan “lulus tidak tepat waktu”. Hasil analisis menunjukkan bahwa, kasus pada penelitian tersebut algoritma C4.5 dapat memprediksi dengan nilai akurasi hanya sebesar 82% (Astuti, 2017). Metode ini digunakan karena keunggulannya dalam hal kecepatan, dan klasifikasi sederhana sehingga mudah untuk diinterpretasikan oleh manusia (Solehuddin et al., 2022).

Berdasarkan referensi yang dijadikan penulis sebagai referensi dalam menjalankan penelitian ini. Maka penelitian dengan menggunakan metode *decision tree* merupakan salah satu metode yang tepat dalam mengelola proses klasifikasi. Algoritma C4.5 merupakan kelompok

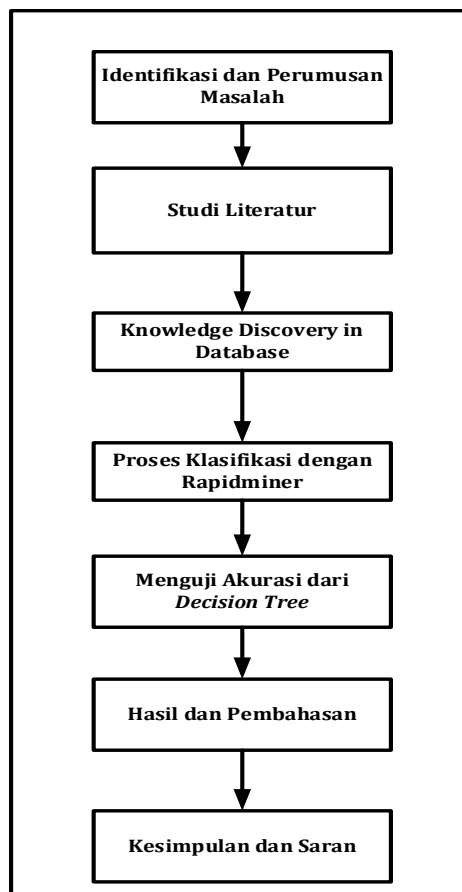
algoritma pohon keputusan (*decision tree*). Algoritma tersebut memiliki masukan yang diuji dengan contoh-contoh yang dimiliki. Contoh data yang digunakan pada data yang digunakan dalam membuat sebuah pohon yang sudah teruji kebenarannya. Sedangkan *samples* merupakan *field-field* (Sianturi, 2018). Alat bantu yang digunakan dalam memproses data ini adalah Rapidminer yang mana Rapidminer merupakan perangkat lunak yang dapat diakses oleh siapa saja dan bersifat terbuka (*open source*). Rapidminer ini dijadikan sebuah solusi untuk menganalisa terhadap data *processing*. Pada Rapidminer ini digunakan berbagai teknik seperti teknik deskriptif dan prediksi. Rapidminer menggunakan bahasa Java untuk pengoperasiannya (Setio et al., 2020).

Berdasarkan fenomena tersebut, maka peneliti ingin melakukan penelitian mengenai klasifikasi terhadap data prediksi kelulusan mahasiswa dengan menggunakan metode *decision tree* dan *tools* seperti rapidminer dalam mengelola data mahasiswa yang menyelesaikan proses studi pada perguruan tinggi. Sehingga dengan ini dapat membantu pihak perguruan tinggi untuk mengevaluasi di mana letak kekurangan mahasiswa tersebut agar bisa menyelesaikan studi lebih tepat waktu, sedangkan data yang akan digunakan oleh peneliti adalah data mahasiswa yang masih aktif dalam menyelesaikan studi dengan atribut data yang sudah dijelaskan oleh peneliti pada fenomena di atas.

## METODE PENELITIAN

### Tahapan Penelitian

Adapun desain penelitian dimulai dari beberapa tahapan yang akan dijalankan oleh penulis pada penelitian ini. Pendekatan penelitian ini menggunakan *Decision Tree* sebagai penentu dalam melakukan proses klasifikasi data prediksi terhadap tingkat kelulusan mahasiswa tepat waktu. *Tools* yang digunakan adalah Rapidminer sebagai alat untuk mengelola data. Tahapan penelitian ini dapat dilihat pada Gambar 1 berikut ini:



Gambar 1. Tahapan Penelitian

1. Tahap pertama melakukan penentuan identifikasi dan perumusan masalah pada penelitian ini.
2. Proses selanjutnya melakukan studi pustaka untuk mempelajari penelitian-penelitian terkait terutama mengenai klasifikasi data terhadap kelulusan mahasiswa tepat waktu hingga algoritma *Decision Tree*.
3. Proses selanjutnya melakukan proses pengolahan data dengan menggunakan tahapan *knowledge discovery in database* atau KDD dari proses seleksi data, pembersihan data, tranformasi data, hingga data itu siap untuk diolah dengan alat bantu Rapidminer dengan algoritma *Decision Tree*.
4. Proses selanjutnya melakukan uji coba terhadap algoritma *Decision Tree* dengan membagi data *training* dan data *testing*, uji coba dilakukan sebanyak 5 kali untuk mendapatkan hasil akurasi tertinggi.
5. Hasil dan pembahasan, merupakan tahap pembahasan hasil dari penelitian yang dijalankan oleh peneliti, yang mana hasil ini merupakan akurasi dari klasifikasi terhadap data prediksi kelulusan mahasiswa tepat waktu.
6. Kesimpulan, proses di mana peneliti menarik kesimpulan terhadap hasil akhir dari penelitian yang dijalankan oleh peneliti.

## Data Penelitian

Untuk menunjang penelitian ini maka diperlukannya data penelitian. Data penelitian terdiri dari data *sample*, tata cara pengumpulan data, hingga jenis data yang digunakan oleh peneliti. Data yang digunakan pada penelitian ini adalah sampel data dari data alumni yang didapat dari beberapa kampus, atau perguruan tinggi yang ada di Kota Palembang. Jumlah *record data* akan menggunakan sebanyak 1739 record data.

## Pengumpulan Data

Adapun teknik atau metode pengumpulan data yang digunakan penulis pada penelitian sebagai berikut ini:

1. Melakukan proses studi literatur atau studi pustaka untuk mengumpulkan referensi penelitian terkait dengan penelitian yang sedang dijalankan oleh peneliti.
2. Proses pengumpulan data diambil dengan mengumpulkan data alumni yang mana data ini akan menjadi landasan terhadap data mahasiswa yang masih aktif, dan menyesuaikan dengan atribut data yang digunakan.

## HASIL DAN PEMBAHASAN

### Pengumpulan Data

Setelah melakukan identifikasi dan perumusan masalah, peneliti melakukan pengumpulan *dataset* sebelum data ini diolah dengan menggunakan model *Decision Tree*. Dengan menggunakan data alumni untuk dijadikan landasan dalam pemberian label status kelulusan mahasiswa yang masih aktif saat ini, serta data yang sudah dikumpulkan oleh penulis yang berasal dari beberapa perguruan tinggi yang ada di Kota Palembang maka terkumpul sebanyak 1739 data dengan rincian data sebagai berikut ini pada Tabel 1.

Tabel 1. Tabel Jumlah Dataset

No	Perguruan Tinggi	Jumlah Data
1	UIN Raden Fatah Palembang	513
2	Universitas Bina Darma Palembang	700
3	Universitas PGRI Palembang	90
4	Stisipol Candradimuka Palembang	436

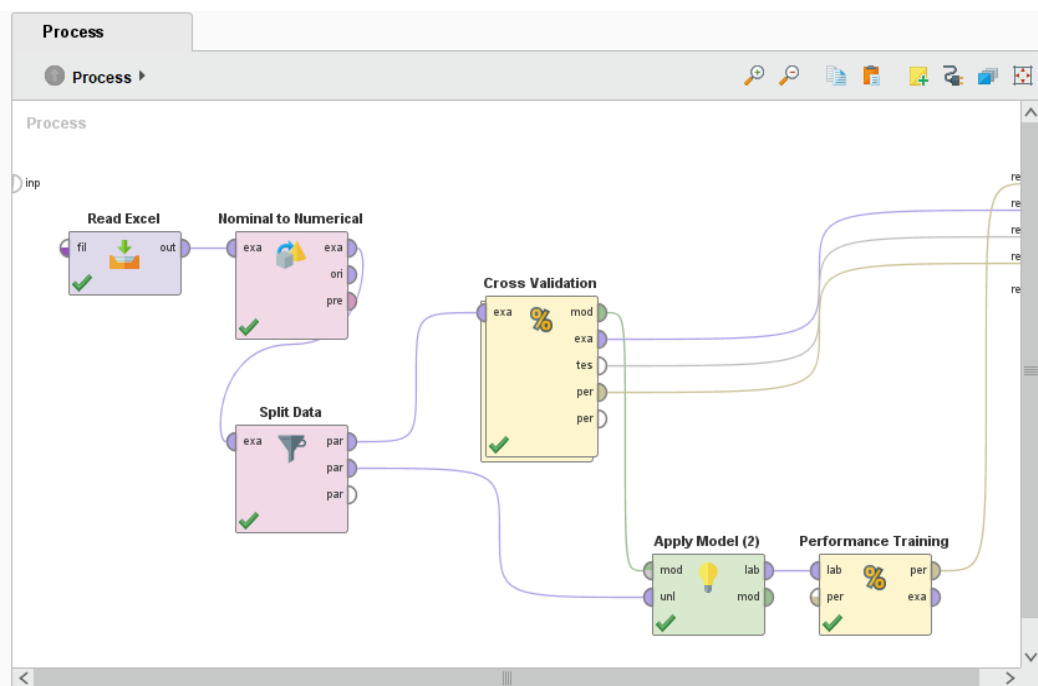
Jumlah data alumni yang telah dikumpulkan berjumlah 1739 *record* data yang sudah dikumpulkan oleh penulis, dari beberapa perguruan tinggi, untuk data mahasiswa yang lulus tepat waktu berjumlah 909, sedangkan yang tidak tepat waktu berjumlah 830, berikut merupakan grafik perbandingan dari data kelulusan mahasiswa.



Gambar 2. Grafik Jumlah Dataset

### Pengujian dengan Rapidminer

Berikut ini uji coba dengan menggunakan *tools* untuk melihat akurasi dari algoritma *Decision Tree* dalam melakukan klasifikasi data prediksi kelulusan mahasiswa dan 1739 *record data*, adapun operator-operator yang digunakan oleh peneliti pada penelitian ini dapat dilihat pada Gambar 3 berikut ini:



Gambar 3. Pengujian Dengan Rapidminer

### Hasil Pengujian Data

Berikut ini merupakan hasil dari uji coba dengan menggunakan *tools* Rapidminer pada penelitian penerapan algoritma *Decision Tree* dalam klasifikasi data prediksi kelulusan mahasiswa yang mana hasil tersebut merupakan hasil uji data dengan bantuan *tools* Rapidminer yang dapat dilihat pada Tabel 2 berikut ini:

**Tabel 2. Tabel Jumlah Dataset**

	Tepat Waktu	Tidak Tepat Waktu	Precision
Pred. Tepat Waktu	74	4	94.87%
Pred. Tidak Tepat Waktu	17	79	82.29%
Recall	81.32%	95.18%	

### Hasil Uji Validasi

Berikut ini merupakan hasil dari pengujian dengan bantuan alat bantu Rapidminer seperti berikut ini dengan hasil dari uji coba klasifikasi terhadap prediksi kelulusan mahasiswa dengan menggunakan metode *Decision Tree* dengan menggunakan *tools* Rapidminer dengan 1739 sampel *dataset* yang telah dilakukan pengujian sebanyak 5 kali untuk mendapatkan hasil *confussion matrix* yang dapat dilihat pada Gambar 4.

accuracy: 87.88% +/- 5.85% (micro average: 87.93%)

	true Tepat Waktu	true Tidak Tepat Waktu	class precision
pred. Tepat Waktu	74	4	94.87%
pred. Tidak Tepat Waktu	17	79	82.29%
class recall	81.32%	95.18%	

**Gambar 4. Hasil Pengujian dengan Rapidminer**

Berdasarkan gambar hasil pengolahan data di atas maka berikut ini merupakan hasil pengujian terhadap 1739 *record* data yang telah dilakukan pengujian sebanyak 5 kali untuk mendapatkan hasil pengujian tertinggi, yang dapat dilihat pada Tabel 3 *confussion matrix* berikut ini:

**Tabel 3. Tabel K-Fold Validation atau confusion Matrix**

No	Record Data	Data Training	Data Testing	Pembagian Data	Hasil Akurasi	Hasil Presentase
1	1739	1565	174	90:10	0.8793	87.93%
2	1739	1391	348	80:20	0.8276	82.76%
3	1739	1217	522	70:30	0.8084	80.84%
4	1739	1043	696	60:40	0.8621	86.21%
5	1739	870	870	50:50	0.8368	83.68%

Berikut ini merupakan hasil dari ujicoba yang menghasilkan hasil akurasi tertinggi sebesar 87.93% yang mana hasil ini didapat dari pembagian data sebesar 10% untuk data *testing* dan 90% sebagai data *training*. Uji coba dilakukan sebanyak 5 kali terhadap jumlah *record* data yang sama yaitu 1739 data. Berikut ini merupakan hasil perhitungan manual terhadap hasil akurasi tertinggi yang dapat dilihat sebagai berikut ini:

a. Hasil Recall



$$\text{Tepat Waktu} = \frac{74}{74+17} = \frac{74}{91} = 0.8132 \text{ (81.32\%)}$$

$$\text{Tidak Tepat Waktu} = \frac{79}{79+4} = \frac{79}{83} = 0.9518 \text{ (95.18\%)}$$

b. Hasil *Precision*

$$\text{Tepat Waktu} = \frac{74}{74+4} = \frac{74}{78} = 0.9487 \text{ (94.87\%)}$$

$$\text{Tidak Tepat Waktu} = \frac{79}{79+17} = \frac{79}{96} = 0.8229 \text{ (82.29\%)}$$

c. Hasil Akurasi

$$\text{Accuracy} = \frac{74+79}{74+4+79+17}$$

$$\text{Accuracy} = \frac{153}{174}$$

$$\text{Accuracy} = 0.8793 \text{ (87.93\%)}$$

## KESIMPULAN DAN SARAN

### Kesimpulan

Kesimpulan dari hasil akhir penelitian yang sudah dijalankan oleh peneliti pada penerapan algoritma *decision tree* terhadap data klasifikasi terhadap penggunaan data prediksi mahasiswa yang dapat lulus tepat waktu atau tidak dengan memakai data alumni sebagai landasan tolak ukur kelulusan setiap mahasiswa apakah tepat waktu atau sebaliknya, sampel data berasal dari data alumni yang telah dikumpulkan dari beberapa perguruan tinggi di Kota Palembang. Berdasarkan pengujian didapat hasil akurasi tertinggi yaitu sebesar 0.8739, atau sebesar 87.93% dengan pembagian data 90% data *training*, dan 10% data *testing* yang digunakan dari 1739 *record* data dengan menggunakan metode *Decision Tree* menghasilkan akurasi sebesar 87.93%.

### Saran

Untuk pengembangan penelitian selanjutnya diharapkan adanya pengembangan atau peningkatan terhadap akurasi dengan pembagian data dan jumlah *dataset* yang berbeda sehingga akan menghasilkan penelitian yang bervariasi guna menambah referensi penelitian terutama penelitian mengenai *decision tree* atau mengenai klasifikasi.

## REFERENSI

- 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.v2i2.1067>
- Haryatmi, E., & Pramita Hervianti, S. (2021). Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 386–392. <https://doi.org/10.29207/resti.v5i2.3007>
- Kinasih, H. W., Prajanto, A., & Sartika, M. (2021). Peran Dosen Pembimbing Dalam Lulus Tepat Waktu Mahasiswa : Study Pada Mahasiswa Akuntansi Universitas X. *Proceeding SENDU*, 208–214.
- Ma'sum, J., Febriani, A., & Rachmawaty, D. (2021). Penerapan Metode Klasifikasi Decision Tree Untuk Memprediksi Kelulusan Tepat Waktu. *Journal Of Industrial Engineering And Technology (Jointech)*, 1(2), 52–60.
- Rohman, A., & Rufiyanto, A. (2019). Implementasi Data Mining Dengan Algoritma Decision Tree C4 . 5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandaran. *Proceeding SINTAK 2019*, 134–139.

- Rohmawan, E. P. (2018). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Desicion Treedan Artificial Neural Network. *Jurnal Ilmiah MATRIK*, 20(1), 21–30.
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA (Prosiding Seminar Nasional Matematika)*, 3.
- Sianturi, F. A. (2018). Analisa Decision Tree Dalam Pengolahan Data Siswa. *MEANS (Media Informasi Analisa Dan Sistem)*, 3(2), 166–172. <https://doi.org/10.54367/means.v3i2.287>
- Solehuddin, M., Syafei, W. A., & Gernowo, R. (2022). Metode Decision Tree Untuk Meningkatkan Kualitas Rencana Pelaksanaan Pembelajaran dengan Algoritma C4.5. *Jurnal Penelitian Dan Pengembangan Pendidikan (JPPP)*, 6(3).
- Supangat, S., Amna, A. R., & Rahmawati, T. (2018). Implementasi Decision Tree C4.5 Untuk Menentukan Status Berat Badan dan Kebutuhan Energi Pada Anak Usia 7-12 Tahun. *Teknika*, 7(2), 73–78. <https://doi.org/10.34148/teknika.v7i2.90>