

Penerapan Algoritma C4.5 dalam Memprediksi Minat Siswa SMKS Unggulan Wakatobi untuk Melanjutkan Pendidikan Tinggi

Sigrid Tamriesfatno ^{1*}, Muhammad Ikhwan Mardin ², La Ode Muh. Armadi. AM³, M. Noor Fuad⁴, Muhammad As'ad⁵

¹ Institut Kesehatan dan Teknologi Buton Raya; email : sigridaaris@gmail.com

² Institut Kesehatan dan Teknologi Buton Raya; email : muhammadikhwan034@gmail.com

³ Institut Kesehatan dan Teknologi Buton Raya; email : armadiam18@gmail.com

⁴ Institut Kesehatan dan Teknologi Buton Raya; email : mnoorfuad2@gmail.com

⁵ Institut Kesehatan dan Teknologi Buton Raya; email : achmad.asad147@gmail.com

Abstract: Students' interest in Vocational High Schools (SMK) to pursue higher education is influenced by various complex factors, ranging from economic conditions to personal backgrounds. This study aims to predict the interest of students at SMKS Unggulan Wakatobi in continuing their education to the tertiary level, providing a basis for school evaluations in developing counseling guidance strategies. The method used in this research is Data Mining by applying the C4.5 algorithm to form a decision tree as a classification model. The analyzed data includes variables such as parents' economic capability and students' gender. The results indicate that the C4.5 algorithm successfully maps student interest patterns, where the parents' economic capability serves as the root node or primary determinant. A specific finding shows that students with low economic capability actually have a high tendency to pursue higher education. Meanwhile, in the moderate and high economic categories, gender becomes the final determining factor, where female students tend to be interested in continuing their studies, while male students do not. This model is expected to serve as an instrument for SMKS Unggulan Wakatobi to increase the number of graduates pursuing higher education, thereby supporting the improvement of human resource quality in the Wakatobi region.

Keywords: C4.5 Algorithm; Data Mining; Educational Interest; Decision Tree; SMKS Unggulan Wakatobi

Abstrak: Minat siswa Sekolah Menengah Kejuruan (SMK) untuk melanjutkan pendidikan ke jenjang perguruan tinggi dipengaruhi oleh berbagai faktor kompleks, mulai dari kondisi ekonomi hingga latar belakang personal. Penelitian ini bertujuan untuk memprediksi minat siswa SMKS Unggulan Wakatobi dalam melanjutkan pendidikan tinggi guna memberikan dasar evaluasi bagi pihak sekolah dalam menyusun strategi bimbingan konseling. Metode yang digunakan dalam penelitian ini adalah Data Mining dengan menerapkan algoritma C4.5 untuk membentuk pohon keputusan (decision tree) sebagai model klasifikasi. Data yang dianalisis mencakup variabel kemampuan ekonomi orang tua dan jenis kelamin siswa. Hasil penelitian menunjukkan bahwa algoritma C4.5 berhasil memetakan pola minat siswa, di mana faktor kemampuan ekonomi orang tua menjadi atribut akar (root node) atau penentu utama. Temuan spesifik menunjukkan bahwa siswa dengan kemampuan ekonomi rendah justru memiliki kecenderungan tinggi untuk melanjutkan pendidikan. Sementara pada kategori ekonomi sedang dan tinggi, jenis kelamin menjadi faktor penentu akhir, di mana siswa perempuan cenderung berminat melanjutkan studi, sedangkan siswa laki-laki cenderung tidak. Model ini diharapkan dapat menjadi instrumen bagi SMKS Unggulan Wakatobi dalam meningkatkan angka lulusan yang menempuh pendidikan tinggi demi penguatan kualitas sumber daya manusia di daerah Wakatobi.

Kata kunci: Algoritma C4.5; Data Mining; Minat Pendidikan; Pohon Keputusan; SMKS Unggulan Wakatobi.

Diterima: 18 September 2025

Direvisi: 22 September 2025

Diterima: 25 September 2025

Diterbitkan: 1 November 2025

Versi sekarang: Januari 2026



Hak cipta: © 2025 oleh penulis.
Diserahkan untuk kemungkinan
publikasi akses terbuka
berdasarkan syarat dan ketentuan
lisensi Creative Commons
Attribution (CC BY SA) (
<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Pendahuluan

Pendidikan merupakan salah satu faktor penting dalam meningkatkan kualitas sumber daya manusia. Pendidikan merupakan cara untuk mewujudkan sesuatu pewarisan budaya dari satu generasi ke generasi yang lain (Abd Rahman, 2022). Semakin tinggi tingkat pendidikan seseorang, semakin besar pula peluangnya untuk memperoleh keterampilan, pengetahuan, dan kesempatan kerja yang lebih baik. Namun, tidak semua siswa pada jenjang Sekolah Menengah Kejuruan (SMK) memiliki minat yang sama untuk melanjutkan pendidikan ke jenjang yang lebih tinggi. Sebagian siswa memilih langsung bekerja setelah lulus, sementara sebagian lainnya berkeinginan melanjutkan studi ke perguruan tinggi. Hal ini menunjukkan perlunya analisis mendalam mengenai faktor-faktor yang memengaruhi minat siswa dalam melanjutkan pendidikan.

SMKS Unggulan Wakatobi sebagai salah satu lembaga pendidikan menengah kejuruan memiliki peran penting dalam menyiapkan lulusan yang berkualitas. Akan tetapi, fenomena yang sering ditemui adalah adanya perbedaan minat siswa setelah lulus. Beberapa siswa memiliki motivasi tinggi untuk melanjutkan pendidikan, sedangkan yang lain lebih memilih langsung bekerja atau bahkan berhenti melanjutkan pendidikan. Faktor-faktor yang memengaruhi keputusan ini dapat berupa latar belakang keluarga, kondisi ekonomi, prestasi akademik, motivasi pribadi, serta lingkungan sosial. Oleh karena itu, dibutuhkan suatu metode analisis yang mampu memprediksi kecenderungan siswa untuk melanjutkan pendidikan tinggi.

Salah satu pendekatan yang dapat digunakan dalam menganalisis minat siswa adalah metode data mining. Data mining memungkinkan peneliti untuk menemukan pola tersembunyi dari data siswa yang selama ini tersimpan, seperti nilai akademik, absensi, prestasi non-akademik, dan data sosial-ekonomi. Di antara berbagai algoritma yang tersedia, algoritma C4.5 merupakan salah satu metode klasifikasi yang banyak digunakan karena kemampuannya menghasilkan pohon keputusan yang mudah dipahami. Pohon keputusan ini dapat membantu pihak sekolah dalam memetakan siswa yang berpotensi melanjutkan pendidikan tinggi maupun yang tidak.

Dengan penerapan algoritma C4.5, diharapkan pihak sekolah dapat memperoleh informasi yang lebih akurat mengenai faktor-faktor penentu minat siswa untuk melanjutkan studi. Hasil prediksi ini dapat dijadikan bahan evaluasi dan pertimbangan dalam menyusun strategi bimbingan konseling, program motivasi, serta kerja sama dengan perguruan tinggi. Penelitian ini diharapkan mampu memberikan kontribusi nyata dalam meningkatkan jumlah lulusan SMKS Unggulan Wakatobi yang melanjutkan pendidikan ke jenjang lebih tinggi, sehingga dapat mendukung peningkatan kualitas sumber daya manusia di daerah Wakatobi.

2. Tinjauan Literatur

2.1. Pendidikan Tinggi dan Minat Melanjutkan Studi

Status sosial ekonomi orang tua, prestasi belajar, persepsi terhadap biaya pendidikan, serta efikasi diri terbukti tidak memberikan pengaruh yang signifikan terhadap minat siswa untuk melanjutkan studi ke perguruan tinggi (Effendi dkk, 2024). Temuan ini menunjukkan bahwa keputusan siswa dalam melanjutkan pendidikan ke jenjang yang lebih tinggi tidak semata-mata ditentukan oleh faktor ekonomi, kemampuan akademik, maupun keyakinan terhadap kemampuan diri sendiri. Oleh karena itu, pihak sekolah disarankan untuk berperan aktif dalam memberikan motivasi, bimbingan, serta penyediaan fasilitas pendukung, seperti layanan konseling dan informasi pendidikan lanjutan, guna mendorong dan memfasilitasi siswa agar memiliki minat dan kesiapan yang lebih kuat untuk melanjutkan pendidikan ke perguruan tinggi.

2.2. Data Mining

Data Mining merupakan bidang yang mempelajari proses pengumpulan, pembersihan, pengolahan, serta analisis data untuk mengekstraksi informasi dan wawasan yang bernilai. Dalam penerapannya di dunia nyata, data mining menghadapi beragam variasi, mulai dari jenis permasalahan, area aplikasi, bentuk perumusan masalah, hingga representasi data yang digunakan. Oleh sebab itu, istilah penambangan data digunakan sebagai sebutan umum untuk mewakili keseluruhan proses dan berbagai tahapan dalam pengolahan data tersebut (S. Mita). Data Mining adalah istilah yang digunakan untuk menggambarkan proses menjelajahi basis data besar untuk mencari pola-pola menarik dan yang sebelumnya tidak diketahui (T. Turap).

2.3 Klasifikasi

Klasifikasi merupakan proses pengelompokan objek atau entitas ke dalam kategori atau kelas tertentu berdasarkan kesamaan karakteristik atau atribut yang dimiliki. Tujuan dari klasifikasi adalah untuk menyederhanakan serta mengorganisasi informasi sehingga dapat mempermudah proses pemahaman, analisis, dan pengambilan keputusan. Metode ini banyak digunakan dalam berbagai bidang.

Dalam konteks rekrutmen, klasifikasi mengacu pada proses pengelompokan dan pengorganisasian kandidat berdasarkan kriteria yang relevan dengan kebutuhan perusahaan. Proses klasifikasi menjadi salah satu aspek penting untuk memastikan rekrutmen berjalan secara efisien dan efektif, serta membantu dalam menentukan kandidat yang paling sesuai untuk suatu posisi. Pengklasifikasi yang dilatih dengan baik mampu memprediksi pola atau perilaku data baru yang belum pernah ditemui sebelumnya dalam permasalahan serupa, yang merupakan inti dari sistem klasifikasi otomatis (R. Goebel, 2002)

2.3 Pohon Keputusan

Decision tree merupakan struktur berbentuk diagram alur yang menyerupai pohon. Pada struktur ini, setiap simpul internal menunjukkan pengujian terhadap suatu atribut, setiap cabang menggambarkan hasil dari pengujian tersebut, dan simpul daun menunjukkan kelas atau distribusi kelas (Rismayani dkk, 2016). Pohon keputusan merupakan salah satu metode klasifikasi yang populer dan memiliki kemampuan yang kuat. Metode Decision Tree mengolah sejumlah besar fakta menjadi struktur pohon yang merepresentasikan aturan-aturan, sehingga aturan tersebut lebih mudah dipahami oleh manusia. Selain itu, Decision Tree juga dapat digunakan untuk mengeksplorasi data serta mengidentifikasi hubungan tersembunyi antara berbagai variabel input dan variabel target (Baiq, 2018).

2.4 Algoritma C4.5

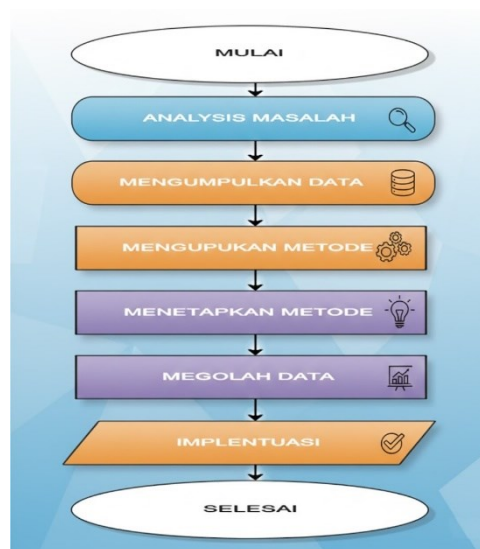
Algoritma C4.5 merupakan salah satu algoritma klasifikasi berbasis pohon keputusan yang dikembangkan oleh Quinlan. Prinsip utama algoritma ini adalah memilih atribut dengan nilai gain ratio tertinggi untuk dijadikan akar pohon, kemudian membagi dataset ke dalam cabang-cabang hingga menghasilkan aturan klasifikasi.

Dalam penelitian di bidang pendidikan, C4.5 banyak digunakan untuk memprediksi hasil belajar, minat siswa, serta faktor-faktor yang memengaruhi kelulusan. Hasil berupa pohon keputusan memudahkan pihak sekolah atau guru untuk memahami pola data yang ada serta mengambil kebijakan yang tepat.

Secara umum metode C4.5 untuk membangun decision tree adalah sebagai berikut:

2.5 Rancangan Penelitian

Pada rancangan penelitian ini penulis menguraikan metodologi dan kerangka penelitian kerja yang digunakan dalam menyelesaikan masalah peneliti. Rancangan penelitian dapat dilihat dalam rancangan flowchart pada gambar 1.



Gambar 1. *Flowchart Rancangan Penelitian*

Pada Gambar 1. menjelaskan Alur penelitian ini dimulai dengan mengidentifikasi masalah dan mengumpulkan data siswa SMKS Unggulan Wakatobi. Setelah data siap, selanjutnya menetapkan penggunaan Algoritma C4.5 yang berfungsi untuk memetakan faktor-faktor penentu tersebut ke dalam sebuah pohon keputusan (decision tree). Data tersebut diolah melalui perhitungan teknis untuk melihat atribut mana yang paling berpengaruh terhadap keputusan siswa.

3. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining*, klasifikasi. Tujuan dari penelitian ini adalah membangun model prediksi minat siswa SMKS Unggulan Wakatobi dalam melanjutkan pendidikan tinggi.

3.1. Sumber dan Jenis Data

Data penelitian menggunakan data primer dan data sekunder. Data primer diperoleh melalui wawancara dan kuesioner kepada siswa SMKS Unggulan Wakatobi terkait faktor yang memengaruhi minat melanjutkan pendidikan tinggi, seperti motivasi, dukungan orang tua, kondisi ekonomi, dan lingkungan sosial. Data sekunder berupa data akademik siswa yang diperoleh dari pihak sekolah, meliputi nilai raport, absensi, prestasi non-akademik, serta data administrasi siswa.

3.2 Tahapan Penelitian dengan Algoritma C4.5

Proses penelitian ini dilakukan melalui beberapa tahapan utama, yaitu:

3.2.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui beberapa metode, yaitu observasi, wawancara, kuesioner, dan dokumentasi. Observasi dilakukan untuk mengamati kondisi akademik dan non-akademik siswa/i secara langsung. Wawancara dilakukan kepada guru dan wali kelas guna memperoleh informasi mengenai faktor-faktor yang memengaruhi minat siswa dalam melanjutkan pendidikan. Selain itu, kuesioner dibagikan kepada siswa kelas XII untuk mengumpulkan data terkait nilai rata-rata akademik, penghasilan orang tua, pendidikan orang tua, motivasi pribadi, dukungan keluarga, serta minat melanjutkan kuliah (Ya/Tidak) yang digunakan sebagai variabel target.

3.2.2. Preprocessing Data

Tahap preprocessing bertujuan untuk meningkatkan kualitas data sebelum dilakukan proses klasifikasi. Tahapan meliputi:

a. Data Cleaning

Data Cleaning adalah proses membersihkan dan menyiapkan data agar siap dianalisis atau disimpan dalam basis data. Tujuannya adalah memastikan data akurat,

konsisten, dan berkualitas. Data cleaning meliputi tiga tahap, yaitu: menghapus data duplikat, menangani missing value dan memperbaiki inkonsistensi data.

- b. Data Transformation
Data Transformation adalah proses mengubah bentuk, format, atau skala data agar data lebih sesuai, konsisten, dan siap digunakan untuk analisis, pemodelan, atau penyimpanan dalam basis data. Tujuannya yaitu menyesuaikan data dengan kebutuhan analisis, meningkatkan akurasi dan kinerja proses pengolahan data serta membuat data lebih mudah dipahami dan diolah
- c. Data Selection
Data Selection adalah proses memilih data yang relevan dan diperlukan dari keseluruhan data yang tersedia untuk digunakan dalam analisis atau pengolahan selanjutnya. Data Selection bertujuan untuk: mengambil data yang sesuai dengan tujuan analisis, mengurangi data yang tidak diperlukan dan meningkatkan efisiensi pemrosesan data
- d. Data Splitting
Data Splitting adalah proses membagi dataset menjadi beberapa bagian dengan tujuan melatih, menguji, dan mengevaluasi model atau sistem pengolahan data. Data Splitting dilakukan agar hasil analisis atau pemodelan lebih objektif dan tidak bias, karena data yang digunakan untuk pelatihan dipisahkan dari data pengujian.

3.2.3. Pembentukan Pohon Keputusan dengan Algoritma C4.5

- a. Menghitung Entropy
Entropy digunakan untuk mengukur tingkat ketidakpastian atau ketidakteraturan dalam suatu data.
$$\text{entropy}(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Keterangan
S : himpunan kasus
A : atribut
n : jumlah partisi S
pi : proporsi dari Si terhadap S
- b. Menghitung Information Gain
$$\text{Gain}(S, A) = \sum_{i=1}^n * \frac{S_i}{S} * \text{entropy}(S_i)$$

Keterangan:
S : himpunan kasus
A : atribut
n : jumlah partisi atribut A
|Si| : jumlah kasus pada partisi ke-i (Si)
|S| : jumlah kasus dalam S
- c. Memilih atribut dengan Gain tertinggi sebagai akar pohon
Setelah nilai information gain untuk setiap atribut dihitung, atribut yang memiliki nilai gain tertinggi dipilih sebagai akar (root) dari pohon keputusan. Atribut ini dianggap paling mampu memisahkan data ke dalam kelompok yang lebih seragam berdasarkan kelasnya. Pemilihan akar yang tepat sangat penting karena akan memengaruhi struktur dan akurasi pohon keputusan yang dihasilkan.
- d. Membentuk cabang berdasarkan nilai atribut
Setelah atribut akar ditentukan, pohon keputusan akan membentuk cabang-cabang sesuai dengan nilai-nilai yang dimiliki oleh atribut tersebut. Setiap cabang merepresentasikan satu nilai atribut dan berisi subset data yang memenuhi kondisi tersebut. Proses ini bertujuan untuk membagi data ke dalam kelompok yang lebih kecil dan lebih homogen sehingga memudahkan proses klasifikasi lanjutan.
- e. Mengulangi proses hingga semua data terklasifikasi

Proses perhitungan entropy, information gain, pemilihan atribut, dan pembentukan cabang diulangi pada setiap cabang pohon menggunakan atribut yang tersisa. Proses ini berlanjut hingga semua data dalam cabang berada pada satu kelas yang sama atau tidak ada lagi atribut yang dapat digunakan. Pada tahap akhir, setiap daun (leaf) pada pohon keputusan merepresentasikan hasil klasifikasi akhir dari data.

4. Hasil dan Pembahasan

Dalam penelitian ini data yang digunakan akan diolah dari hasil kuisisioner, dengan jumlah kuisisioner kepada siswa dengan jumlah 100 sampel data. Setiap faktor atau variabel terdiri dari pertanyaan yang diajukan ke siswa yang diberikan.

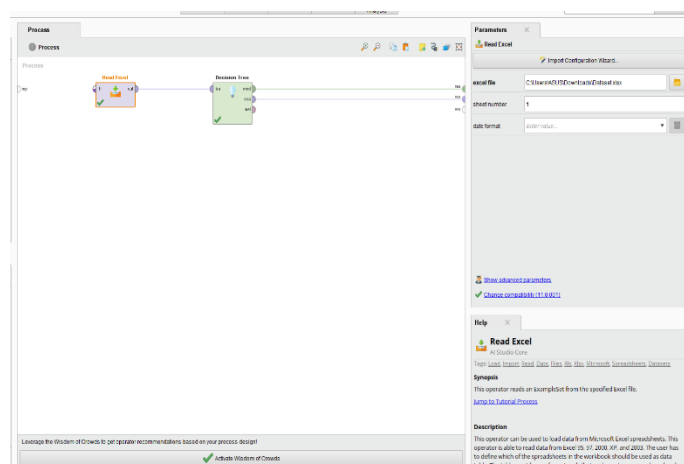
Data yang digunakan merupakan jenis statistik deskriptif dengan siswa/i SMA Unggulan Wakatobi. Data hasil dari kuesioner selanjutnya diolah menggunakan RapidMiner. Berikut ini sampel data hasil rekapitulasi kuesioner yang diolah di Microsoft Excel pada Tabel 1:

user	jenis kelamin	kemampuan ekonomi orang tua/wali	tulang punggung keluarga	Label
siswa001	P	sedang	orangtua	TIDAK
siswa002	P	sedang	orangtua	TIDAK
siswa003	P	sedang	orangtua	YA
siswa004	P	sedang	orangtua	YA
siswa005	P	sedang	orangtua	YA
.....				
siswa100	P	rendah	saudara	YA

Tabel 1. Data Hasil Rekapitulasi Kuesioner Penelitian

4.1 . Penerapan Algoritma C4.5

Penerapan algoritma untuk membentuk pohon keputusan dilakukan dengan menggunakan tools Altair AI Studio, sebagai berikut:



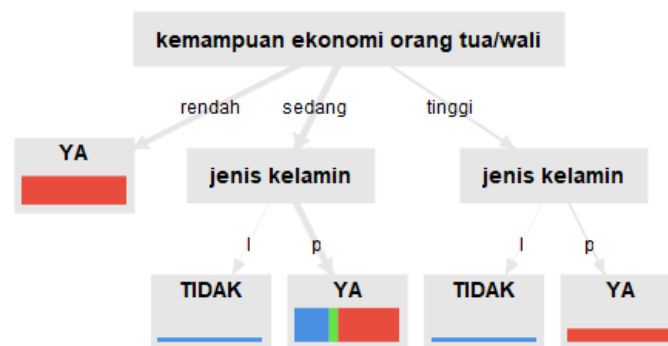
Gambar 2. Tools Altair AI Studio

Gambar di atas menunjukkan rancangan proses analisis data menggunakan RapidMiner Studio yang terdiri atas dua operator utama, yaitu Read Excel dan Decision Tree. Operator Read Excel digunakan untuk mengimpor dataset dari berkas Microsoft Excel dengan menentukan lokasi file dan nomor lembar kerja sebagai sumber data, sehingga data dapat dibaca dan diubah menjadi ExampleSet yang siap diproses. Selanjutnya, dataset tersebut dihubungkan langsung ke operator Decision Tree yang berfungsi untuk membangun model klasifikasi berbasis pohon keputusan dengan mempelajari pola dan hubungan antaratribut

dalam data. Hasil dari proses ini berupa model Decision Tree serta data yang dapat ditampilkan pada panel result untuk dianalisis lebih lanjut.

4.2 Pembentukan Pohon Keputusan

Berikut merupakan hasil pembentukan pohon keputusan yang dilakukan menggunakan Altair AI Studio:



Gambar 3. Pembentukan Pohon Keputusan

Deskripsi

kemampuan ekonomi orang tua/wali = rendah : YA {TIDAK=0, YA =0, YA=30}

kemampuan ekonomi orang tua/wali = sedang

| jenis kelamin = l: TIDAK {TIDAK=3, YA =0, YA=0}

| jenis kelamin = p: YA {TIDAK=12, YA =3, YA=21}

kemampuan ekonomi orang tua/wali = tinggi

| jenis kelamin = l: TIDAK {TIDAK=3, YA =0, YA=0}

| jenis kelamin = p: YA {TIDAK=0, YA =0, YA=13}

5. Kesimpulan

Pohon keputusan menunjukkan bahwa faktor utama penentu hasil adalah kemampuan ekonomi orang tua/wali. Jika kemampuan ekonomi tergolong rendah, maka seluruh kasus langsung diklasifikasikan sebagai ya, sedangkan jika kemampuan ekonomi sedang, maka keputusan dipengaruhi oleh jenis kelamin: untuk laki-laki diklasifikasikan sebagai tidak dan untuk perempuan diklasifikasikan sebagai ya. Sementara itu, pada kemampuan ekonomi tinggi, pola yang sama muncul seperti kategori sedang, di mana laki-laki diklasifikasikan sebagai tidak dan perempuan sebagai ya. Alur pohon ini menunjukkan bahwa kondisi ekonomi menjadi pemisah utama, kemudian pada kategori ekonomi sedang dan tinggi, jenis kelamin menjadi faktor penentu akhir, dengan perempuan cenderung menghasilkan keputusan ya dan laki-laki tidak.

Referensi

- [1] Effendi, A. K., Susanti, E., & Widiatami, A. K. (2024). Determinan Minat Melanjutkan Studi Perguruan Tinggi Pada Siswa Akuntansi Smk Negeri 1 Batang. *Business and Economic Analysis Journal*, 4(1), 14-28.
- [2] S. Mita, Y. Yamazoe, T. Kamataki, dan R. Kato, Metabolic activation of a tryptophan pyrolysis product, 3-amino-1-methyl-5H-pyrido[4,3-b]indole(Trp-P-2) by isolated rat liver nuclei, vol. 14, no. 3. 1981. doi: 10.1016/0304-3835(81)90152-X
- [3] T. Turap, T. B. Merupakan, T. B. Lebih, dan T. D. Turap, Covariance Structure Analysis of Health-Related Indicators in Home-Dwelling Elderly Focused on Subjective Health Perceptions.
- [4] Rismayanti, "Implementasi algoritma c4.5 untuk menentukan penerima beasiswa di stt harapan medan," vol. 12, no. 2, pp. 116–120, 2016.
- [5] Baiq Andriska Candra, "Prediksi Kepuasan Mahasiswa Terhadap Tingkat Pelayanan Menggunakan Algoritma C4.5 (Decision Tree) (Studi Kasus : Fakultas Teknik Universitas Hamzanwadi) Baiq," vol. 1, no. 1, pp. 32–39, 2018
- [6] R. Goebel, J. Siekmann, dan W. Wahlster, *Learning Classifier Systems*. 2002
- [7] Abd Rahman, B. P., Munandar, S. A., Fitriani, A., Karlina, Y., & Yumriani, Y. (2022). Pengertian pendidikan, ilmu pendidikan dan unsur-unsur pendidikan. *Al-Urwatul Wutsqa: Kajian Pendidikan Islam*, 2(1), 1-8.