

Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Metode Decision Tree Cart Pada Stti Sony Sugema

Arisman Ismail^{1*}, Muhamad Rizki², Ridwan Efendi Rai³, Sabaruddin⁴, Enen Nurbaeti⁵, Ary Gunawan⁶, Erik Perdana Ibrahim⁷

^{1,2,3,4,5,6,7} Prodi Teknik Informatika, Sekolah Tinggi Teknologi Informatika Sony Sugema, Karawang, Indonesia

^{1*} arisman.sttis@gmail.com ² muhamadrizki382@gmail.com ³ yogiefather@gmail.com ⁴ sbdr69@gmail.com
⁵ enensyifanurbaeti@gmail.com ⁶ erikperdanaibrahum@gmail.com ⁷

Abstrak

Prediksi masa studi mahasiswa merupakan salah satu permasalahan klasifikasi yang dapat diselesaikan menggunakan teknik data mining. Kemampuan memprediksi potensi keterlambatan kelulusan sejak dini dapat membantu perguruan tinggi dalam menyusun strategi pembinaan akademik yang lebih efektif berbasis data. Penelitian ini bertujuan membangun model prediksi masa studi mahasiswa menggunakan algoritma Classification and Regression Tree (CART) berdasarkan data akademik dan non-akademik mahasiswa Program Studi Teknik Informatika Sekolah Tinggi Teknologi Informatika (STTI) Sony Sugema. Dataset yang digunakan merupakan data mahasiswa yang telah menyelesaikan studi, dengan atribut seperti indeks prestasi akademik, status pekerjaan, lokasi tempat tinggal, serta karakteristik lain yang berpengaruh terhadap masa studi. Tahapan penelitian meliputi proses data preprocessing, pemilihan atribut, pembentukan model klasifikasi menggunakan algoritma CART, serta evaluasi performa model menggunakan metrik accuracy, precision, recall, dan F1-score. Model prediksi kemudian diimplementasikan dalam aplikasi berbasis web menggunakan bahasa pemrograman Python dengan framework Flask sebagai media pengujian dan pemanfaatan hasil prediksi. Hasil evaluasi menunjukkan bahwa model CART mampu menghasilkan akurasi sebesar 90,14%, precision sebesar 95,96%, recall sebesar 92,96%, dan F1-score sebesar 94,42%. Nilai tersebut menunjukkan bahwa algoritma CART memiliki kemampuan yang baik dalam mengklasifikasikan masa studi mahasiswa. Sistem yang dikembangkan diharapkan dapat mendukung pengambilan keputusan akademik melalui identifikasi dini mahasiswa yang berpotensi mengalami keterlambatan kelulusan sehingga upaya pendampingan dapat dilakukan secara lebih tepat sasaran.

Kata Kunci: *Data Mining, Decision Tree, CART, Klasifikasi, Flask*

1. Pendahuluan

Pendidikan tinggi memiliki peran strategis dalam menghasilkan sumber daya manusia yang kompeten dan berdaya saing. Salah satu indikator keberhasilan penyelenggaraan pendidikan tinggi adalah kemampuan perguruan tinggi dalam menghasilkan lulusan yang mampu menyelesaikan studi sesuai dengan masa studi yang telah ditetapkan. Ketepatan waktu kelulusan tidak hanya mencerminkan efektivitas proses pembelajaran, tetapi juga menjadi salah satu indikator mutu penyelenggaraan pendidikan tinggi dan akreditasi program studi (Jananto et al., 2021).

Berdasarkan *Statistik Pendidikan Tinggi Indonesia 2023* yang diterbitkan oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek), tingkat kelulusan mahasiswa program sarjana (S1) tepat waktu masih belum optimal. Sebagian mahasiswa dapat menyelesaikan studi dalam delapan semester, sedangkan sebagian lainnya membutuhkan waktu lebih lama atau bahkan tidak menyelesaikan studinya. Kondisi tersebut menunjukkan bahwa keterlambatan masa studi masih menjadi salah satu permasalahan yang perlu mendapat perhatian dalam pengelolaan pendidikan tinggi sehingga diperlukan sistem prediksi untuk membantu proses pengambilan keputusan akademik.

Permasalahan serupa juga dijumpai pada berbagai perguruan tinggi swasta, termasuk Sekolah Tinggi Teknologi Informatika (STTI) Sony Sugema. Berdasarkan hasil pengamatan terhadap data akademik mahasiswa, masih terdapat mahasiswa yang menyelesaikan studi melebihi masa studi ideal. Keterlambatan tersebut diduga dipengaruhi oleh berbagai faktor, seperti latar belakang pendidikan, status pekerjaan, lokasi tempat tinggal, serta capaian akademik yang tercermin dari Indeks Prestasi (IP) pada setiap semester. Identifikasi dini terhadap faktor-faktor tersebut menjadi penting agar perguruan tinggi dapat menyusun strategi pembinaan akademik secara lebih efektif.

Perkembangan teknologi informasi, khususnya pada bidang *data mining*, memberikan peluang untuk mengolah data akademik menjadi informasi yang bermanfaat dalam mendukung pengambilan keputusan. *Data mining* merupakan proses mengekstraksi pola atau pengetahuan dari kumpulan data berukuran besar menggunakan teknik statistik, pembelajaran mesin *machine learning*, maupun kecerdasan buatan. Dalam bidang pendidikan, penerapan *Educational Data Mining* telah banyak dimanfaatkan untuk memprediksi prestasi akademik, mengidentifikasi mahasiswa yang berisiko mengalami kegagalan studi, serta memprediksi ketepatan waktu kelulusan. Salah satu metode klasifikasi yang banyak digunakan dalam *data mining* adalah algoritma *Classification and Regression Tree*

(CART). Algoritma ini membentuk pohon keputusan berdasarkan atribut yang memiliki kemampuan terbaik dalam membedakan kelas melalui pengukuran *Gini Index*. Keunggulan CART adalah mampu menghasilkan model yang mudah dipahami sekaligus memiliki performa klasifikasi yang baik pada data akademik mahasiswa (Prabawati et al., 2019).

Salah satu metode klasifikasi yang banyak digunakan dalam *data mining* adalah algoritma *Classification and Regression Tree* (CART). Algoritma CART membangun model klasifikasi dalam bentuk pohon keputusan (*decision tree*) dengan membagi data berdasarkan atribut yang memiliki kemampuan terbaik dalam membedakan kelas data. Selain memiliki tingkat akurasi yang baik, CART juga menghasilkan model yang mudah diinterpretasikan sehingga faktor-faktor yang memengaruhi hasil klasifikasi dapat dipahami oleh pengambil kebijakan.

Meskipun berbagai penelitian telah menerapkan algoritma CART untuk klasifikasi pada berbagai bidang, penerapannya dalam memprediksi masa studi mahasiswa dengan memanfaatkan kombinasi variabel akademik dan nonakademik pada STTI Sony Sugema masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk membangun model klasifikasi masa studi mahasiswa menggunakan algoritma CART dengan memanfaatkan atribut seperti status pekerjaan, tempat tinggal, asal sekolah, serta Indeks Prestasi (IP) semester 1 sampai semester 6. Penelitian serupa juga menunjukkan bahwa algoritma CART efektif dalam mengidentifikasi mahasiswa yang berisiko mengalami gagal studi berdasarkan karakteristik akademik maupun nonakademik (Rafika et al., 2024).

Penelitian ini bertujuan untuk (1) Membangun model klasifikasi masa studi mahasiswa di STTI Sony Sugema menggunakan algoritma Classification and Regression Tree (CART) berdasarkan data akademik dan nonakademik mahasiswa, (2) Menganalisis faktor-faktor yang paling berpengaruh terhadap klasifikasi masa studi mahasiswa berdasarkan atribut status pekerjaan, tempat tinggal, asal sekolah, dan Indeks Prestasi (IP) semester 1 sampai semester 6 menggunakan algoritma CART, (3) Mengevaluasi kinerja model klasifikasi algoritma CART dalam mengklasifikasikan masa studi mahasiswa berdasarkan hasil pengujian model, seperti akurasi, presisi, recall, dan F1-score, (4) Memberikan rekomendasi kepada STTI Sony Sugema berdasarkan hasil klasifikasi sebagai dasar pengambilan keputusan untuk mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan studi serta mendukung peningkatan persentase kelulusan tepat waktu.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining, yang bertujuan membangun model klasifikasi masa studi mahasiswa menggunakan algoritma Classification and Regression Tree (CART). Pendekatan ini dipilih karena mampu mengidentifikasi pola hubungan antarvariabel dan menghasilkan aturan klasifikasi (*decision tree*) yang mudah diinterpretasikan sebagai dasar pengambilan keputusan. Tahapan penelitian mengacu pada proses Knowledge Discovery in Databases (KDD) yang meliputi seleksi data, prapemrosesan (*preprocessing*), transformasi data, proses data mining menggunakan algoritma CART, serta evaluasi dan interpretasi hasil model.

Data yang dikumpulkan bersumber dari data yang dikumpulkan berupa data akademik mahasiswa yang sudah dinyatakan lulus oleh STTI Sony Sugema yang memiliki atribut berupa data awal masuk mahasiswa dan data predikat akademik mahasiswa yaitu tahun angkatan, provinsi asal, status pekerjaan, domisili tinggal, dan status asal sekolah. Sedangkan untuk atribut pada data predikat akademik mahasiswa yang dikumpulkan yaitu Indeks Prestasi (IP) semester 1 s.d 6 dan satuan kredit semester (SKS) semester 1 s.d 6.

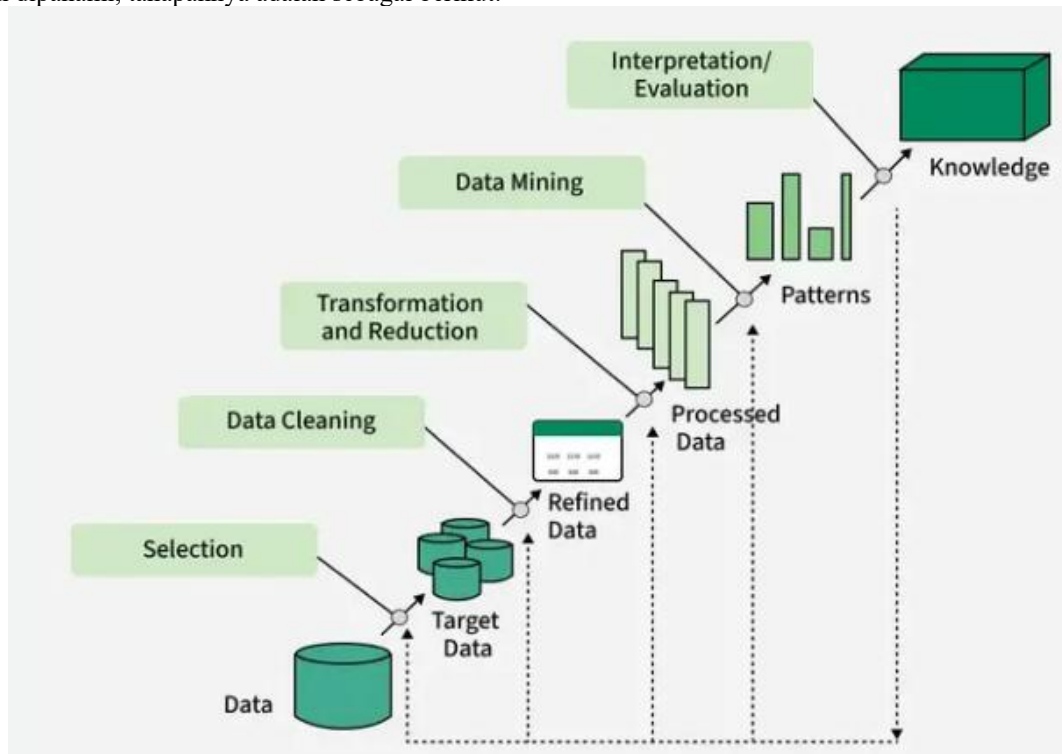
Tahun Angkatan	Provinsi Asal	Domisili Tinggal	Status Sekolah	Status Bekerja	Semester 1		Semester 2		Semester 3		Semester 4		Semester 5		Semester 6	
					SKS	IP	SKS	IP	SKS	IP	SKS	IP	SKS	IP	SKS	IP
2020	Jawa Barat	Karawang	SMK	Ya	11	3.53	19	3.56	16	3.52	17	3.57	19	3.76	16	3.68
2020	Jawa Barat	Karawang	SMK	Tidak	16	3.85	20	3.79	22	3.65	22	3.71	22	3.88	17	3.75
2020	Jawa Barat	Cirebon	SMA	Ya	16	3.4	13	3.47	12	3.41	12	3.48	17	3.44	14	3.42
2020	Jawa Barat	Karawang	SMK	Tidak	21	3.5	20	3.57	22	3.4	22	3.49	22	3.48	22	3.51
2020	Jawa Barat	Karawang	SMK	Tidak	16	3.44	21	3.42	16	3.43	20	3.55	20	3.54	21	3.48
2020	Jawa Barat	Karawang	SMK	Tidak	21	3.42	14	3.51	17	3.58	18	3.4	16	3.51	17	3.62
2020	Jawa Barat	Karawang	SMK	Tidak	21	3.55	13	3.65	18	3.42	11	3.47	17	3.51	13	3.44
2020	Jawa Barat	Karawang	SMK	Ya	15	3.49	12	3.7	12	3.65	11	3.42	11	3.51	13	3.4
2020	Jawa Barat	Karawang	SMK	Tidak	16	3.66	15	3.54	11	3.41	16	3.47	16	3.4	17	3.47
2020	Jawa Barat	Karawang	SMK	Tidak	17	3.55	19	3.67	13	3.82	16	3.4	19	3.49	19	3.68
2020	Jawa Barat	Karawang	SMK	Tidak	17	3.55	20	3.44	17	3.57	16	3.44	21	3.67	20	3.58
2020	Jawa Barat	Karawang	SMK	Tidak	21	3.4	20	3.59	22	3.51	22	3.4	22	3.76	22	3.63
2020	Jawa Barat	Karawang	SMK	Tidak	14	3.46	15	3.49	17	3.58	17	3.54	21	3.61	17	3.71
2020	Jawa Barat	Karawang	SMK	Tidak	21	3.45	13	3.57	21	3.63	21	3.55	16	3.48	20	3.71
2020	Jawa Barat	Karawang	SMK	Tidak	21	3.45	16	3.68	22	3.5	23	3.68	22	3.5	22	3.43
2020	Jawa Barat	Karawang	SMK	Tidak	17	3.6	11	3.68	17	3.73	21	3.45	14	3.62	12	3.5
2020	Jawa Barat	Karawang	SMK	Tidak	21	3.4	12	3.45	22	3.6	22	3.65	21	3.65	23	3.48
2020	Jawa Barat	Karawang	SMK	Tidak	20	3.63	11	3.4	23	3.5	22	3.64	22	3.65	23	3.59
2020	Jawa Barat	Karawang	SMK	Tidak	21	3.57	21	3.6	22	3.64	22	3.7	23	3.52	17	3.48

Gambar 1. Screenshot Data Akademik Mahasiswa S1 STTI Sony Sugema

DOI: <https://doi.org/10.69693/ijmst.v4i2.12433>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Knowledge Discovery in Databases (KDD) merupakan proses untuk mencari dan mengidentifikasi pola atau informasi yang bermanfaat dalam data, di mana pola dan informasi yang ditemukan harus sah, baru, berguna, dan mudah dipahami, tahapannya adalah sebagai berikut:



Gambar 2. Tahapan proses KDD

Dataset adalah aspek krusial dalam penelitian ini, sehingga pemilihan *dataset* telah dilakukan dengan seksama. *Dataset* kemudian disesuaikan dengan kebutuhan penelitian, ambil kolom-kolom yang diperlukan agar sesuai dengan kebutuhan pemodelan data menggunakan algoritma CART.

Bekerja	DomisiliTinggal	StatusSekolah	IP Smt 1	IP Smt 2	IP Smt 3	IP Smt 4	IP Smt 5	IP Smt 6	SKS Smt 1	SKS Smt 2	SKS Smt 3	SKS Smt 4	SKS Smt 5	SKS Smt 6	MasaStudi
Ya	Karawang	SMK	3.53	3.56	3.52	3.57	3.76	3.68	1	4	3	3	4	3	Tepat
Tidak	Karawang	SMK	3.85	3.79	3.65	3.71	3.88	3.75	3	4	5	5	5	3	Telat
Ya	Luar Karawang	SMA	3.4	3.47	3.41	3.48	3.44	3.42	3	2	1	1	3	2	Tepat
Tidak	Karawang	SMK	3.5	3.57	3.4	3.49	3.48	3.51	4	4	5	5	5	5	Tepat
Tidak	Karawang	SMK	3.44	3.42	3.43	3.55	3.54	3.48	3	4	3	4	4	4	Tepat
Tidak	Karawang	SMK	3.42	3.51	3.58	3.4	3.51	3.62	4	2	3	3	3	3	Tepat
Tidak	Karawang	SMK	3.55	3.65	3.42	3.47	3.51	3.44	4	2	3	1	3	2	Tepat
Ya	Karawang	SMK	3.49	3.7	3.65	3.42	3.51	3.4	2	1	1	1	1	2	Telat
Tidak	Karawang	SMK	3.66	3.54	3.41	3.47	3.4	3.47	3	2	1	3	3	3	Tepat
Tidak	Karawang	SMK	3.55	3.67	3.82	3.4	3.49	3.68	3	4	2	3	4	4	Tepat
Tidak	Karawang	SMK	3.55	3.44	3.57	3.44	3.67	3.58	3	4	3	3	4	4	Tepat
Tidak	Karawang	SMK	3.4	3.59	3.51	3.4	3.76	3.63	4	4	5	5	5	5	Cepat
Tidak	Karawang	SMK	3.46	3.49	3.58	3.54	3.61	3.71	2	2	3	3	4	3	Tepat
Tidak	Karawang	SMK	3.45	3.57	3.63	3.55	3.48	3.71	4	2	4	4	3	4	Tepat
Tidak	Karawang	SMK	3.45	3.68	3.5	3.68	3.5	3.43	4	3	5	5	5	5	Tepat
Tidak	Karawang	SMK	3.6	3.68	3.73	3.45	3.62	3.5	3	1	3	4	2	1	Tepat
Tidak	Karawang	SMK	3.4	3.45	3.6	3.65	3.65	3.48	4	1	5	5	4	5	Tepat
Tidak	Karawang	SMK	3.63	3.4	3.5	3.64	3.65	3.60	4	1	5	5	5	5	Tepat

Gambar 3. Screenshot Hasil Selection

Proses preprocessing mencakup langkah-langkah dasar yang bertujuan untuk merancang strategi yang sesuai dengan dataset, guna mengatasi missing value (nilai yang tidak ada) dan noise tanpa merusak integritas data. Strategi yang dapat digunakan meliputi penggantian nilai yang bermasalah dengan nilai rata-rata dari kolom yang bersangkutan atau penghapusan satu record secara keseluruhan.

Pada fase ini, dilakukan seleksi data yang diperlukan untuk teknik data mining yang digunakan. Proses transformasi dan pemilihan data ini juga mempengaruhi kualitas hasil data mining yang dihasilkan, karena beberapa karakteristik dari teknik-teknik data mining tertentu bergantung pada tahap ini.

Tabel 1. Transformasi Data

Atribut	Keterangan	Tipe
Bekerja	Ya, Tidak	Discrete
Domisili Tinggal	Karawang, Luar Karawang	Discrete

Status Sekolah	SMA, SMK, MA	<i>Discrete</i>
IP Semester 1	IP 0.00 s.d 4.00	<i>Numeric</i>
IP Semester 2	IP 0.00 s.d 4.00	<i>Numeric</i>
IP Semester 3	IP 0.00 s.d 4.00	<i>Numeric</i>
IP Semester 4	IP 0.00 s.d 4.00	<i>Numeric</i>
IP Semester 5	IP 0.00 s.d 4.00	<i>Numeric</i>
IP Semester 6	IP 0.00 s.d 4.00	<i>Numeric</i>
SKS Semester 1	1 = SKS 0 s.d 12, 2 = SKS 13 s.d 15, 3 = SKS 16 s.d 18, 4 = SKS 19 s.d 21, 5 = SKS 22 s.d 24	<i>Numeric</i>
SKS Semester 2	1 = SKS 0 s.d 12, 2 = SKS 13 s.d 15, 3 = SKS 16 s.d 18, 4 = SKS 19 s.d 21, 5 = SKS 22 s.d 24	<i>Numeric</i>
SKS Semester 3	1 = SKS 0 s.d 12, 2 = SKS 13 s.d 15, 3 = SKS 16 s.d 18, 4 = SKS 19 s.d 21, 5 = SKS 22 s.d 24	<i>Numeric</i>
SKS Semester 4	1 = SKS 0 s.d 12, 2 = SKS 13 s.d 15, 3 = SKS 16 s.d 18, 4 = SKS 19 s.d 21, 5 = SKS 22 s.d 24	<i>Numeric</i>
SKS Semester 5	1 = SKS 0 s.d 12, 2 = SKS 13 s.d 15, 3 = SKS 16 s.d 18, 4 = SKS 19 s.d 21, 5 = SKS 22 s.d 24	<i>Numeric</i>
SKS Semester 6	1 = SKS 0 s.d 12, 2 = SKS 13 s.d 15, 3 = SKS 16 s.d 18, 4 = SKS 19 s.d 21, 5 = SKS 22 s.d 24	<i>Numeric</i>
Masa Studi	Cepat, Tepat, Telat	<i>Discrete</i>

Proses klasifikasi akan dilakukan dengan aplikasi yang dibuat dalam penelitian ini. Kode sumber untuk algoritma CART akan disusun dan diterapkan sebagai bagian dari penelitian. Algoritma CART berfungsi untuk mengonversi data menjadi pohon keputusan (decision tree) dan aturan-aturan keputusan (rule), serta menyederhanakan aturan-aturan tersebut melalui proses pruning.

Selanjutnya dilakukan analisis interpretatif terhadap pohon keputusan yang dihasilkan untuk mengidentifikasi atribut-atribut yang paling berpengaruh terhadap klasifikasi masa studi mahasiswa. Hasil analisis tersebut digunakan sebagai dasar penyusunan rekomendasi bagi STTI Sony Sugema dalam merancang strategi pembinaan akademik, sistem peringatan dini (*early warning system*), serta kebijakan peningkatan layanan akademik bagi mahasiswa yang berpotensi mengalami keterlambatan studi sehingga dapat meningkatkan persentase kelulusan tepat waktu.

3. Hasil dan Pembahasan

Pemodelan Data

Data yang digunakan dalam penelitian ini merupakan data yang telah melalui tahapan pemilihan (*data selection*), pra-pemrosesan (*preprocessing*), dan transformasi sehingga siap digunakan dalam proses klasifikasi. Total dataset yang dianalisis berjumlah 142 record mahasiswa yang diperoleh dari basis data akademik STTI Sony Sugema dan disimpan dalam format Microsoft Excel. Dataset tersebut memuat sejumlah atribut yang dijadikan sebagai variabel prediktor, yaitu status pekerjaan, domisili atau tempat tinggal, status pendidikan/asal sekolah, serta nilai Indeks Prestasi (IP) dan jumlah Satuan Kredit Semester (SKS) yang ditempuh pada semester 1 hingga semester 6. Adapun variabel target dalam penelitian ini adalah masa studi mahasiswa, yang dikelompokkan ke dalam tiga kategori, yaitu "cepat" untuk mahasiswa yang menyelesaikan studi dalam waktu kurang dari 8 semester sebanyak 6 record, "tepat" untuk mahasiswa yang lulus tepat dalam waktu 8 semester sebanyak 128 record, dan "telat" untuk mahasiswa yang menyelesaikan studi lebih dari 8 semester sebanyak 8 record. Dataset tersebut selanjutnya digunakan sebagai dasar dalam proses pembangunan model klasifikasi menggunakan algoritma Classification and Regression Tree (CART) untuk mengidentifikasi pola dan faktor-faktor yang memengaruhi masa studi mahasiswa.

Tabel 2. Sampel Dataset

Bekerja	Domisili	Status Sekolah	IP Smt 1	IP Smt 6	SKS Smt 1	SKS Smt 6	Masa Studi
Tidak	Karawang	MA	3.6	3.78	3	4	Tepat
Tidak	Luar Karawang	SMA	3.57	3.61	4	4	Tepat
Tidak	Karawang	SMK	3.75	3.55	4	4	Cepat

DOI: <https://doi.org/10.69693/ijmst.v4i2.12433>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tidak Ya	Karawang Luar	SMK MA	3.4 3.71	..	3.63 3.66	4 4	..	5 5	Tepat Tepat
Tidak	Karawang	SMK	3.66	..	3.47	3	..	3	Telat
Tidak	Karawang	SMK	3.55	..	3.58	3	..	4	Tepat
Ya	Luar	SMA	3.42	..	3.76	3	..	2	Telat
Tidak	Karawang	SMK	3.42	..	3.62	4	..	3	Tepat
Tidak	Karawang	SMK	3.54	..	3.51	3	..	3	Tepat
Ya	Karawang	SMK	3.43	..	3.67	3	..	3	Telat

Proses Manual Klasifikasi CART

Berikut adalah sampel data latih dan data uji yang memanfaatkan data akademik mahasiswa, yang akan diterapkan dalam perhitungan manual menggunakan metode algoritma decision tree CART:

Tabel 3. Data latih

Bekerja	Domisili Tinggal	Status Sekolah	IP Semester 1	Masa Studi
Tidak	Luar Karawang	SMA	3.57	Tepat
Ya	Karawang	SMK	3.53	Tepat
Tidak	Karawang	SMK	3.4	Telat
Tidak	Karawang	MA	3.6	Tepat
Tidak	Karawang	SMK	3.75	Cepat
Tidak	Karawang	SMK	3.63	Tepat
Ya	Luar Karawang	SMA	3.42	Telat
Tidak	Karawang	SMK	3.57	Tepat
Tidak	Karawang	SMK	3.55	Tepat
Ya	Luar Karawang	MA	3.44	Telat

Proses pengujian model dilakukan menggunakan satu data uji dengan karakteristik status pekerjaan tidak bekerja, domisili Karawang, asal sekolah SMK, dan Indeks Prestasi (IP) Semester 1 sebesar 3,50. Tujuan pengujian ini adalah untuk menentukan kategori masa studi mahasiswa, yaitu cepat, tepat, atau telat, berdasarkan pohon keputusan yang dibangun menggunakan algoritma Classification and Regression Tree (CART). Sebelum proses pembentukan pohon keputusan dilakukan, terlebih dahulu dihitung nilai Gini Index pada simpul induk (*parent node*) sebagai ukuran ketidakhomogenan data. Perhitungan dilakukan terhadap 10 data training menggunakan rumus $Gini\ Index = 1 - \sum_{j=1}^c (P_j)^2$, dengan c menyatakan jumlah kelas, yaitu tiga kelas yang terdiri atas cepat, tepat, dan telat, sedangkan P_j merupakan proporsi masing-masing kelas terhadap jumlah keseluruhan data. Berdasarkan data training pada Tabel 4.2 diketahui terdapat 1 data pada kelas cepat, 6 data pada kelas tepat, dan 3 data pada kelas telat, sehingga proporsi masing-masing kelas adalah 1/10, 6/10, dan 3/10. Dengan demikian, nilai Gini Index pada simpul induk diperoleh melalui perhitungan $1 - ((1/10)^2 + (6/10)^2 + (3/10)^2) = 1 - (0,01 + 0,36 + 0,09) = 1 - 0,46 = 0,54$. Nilai Gini Index sebesar 0,54 menunjukkan bahwa data pada simpul induk masih memiliki tingkat heterogenitas yang cukup tinggi sehingga diperlukan proses pemisahan (*splitting*) berdasarkan atribut yang memiliki nilai penurunan impurity terbesar. Selanjutnya, algoritma CART akan mengevaluasi setiap atribut, seperti status pekerjaan, domisili, asal sekolah, serta nilai IP Semester 1, untuk menentukan atribut terbaik yang dijadikan sebagai simpul akar (*root node*) dalam pembentukan pohon keputusan. Hasil dari proses ini akan menghasilkan aturan klasifikasi (*decision rules*) yang digunakan untuk menentukan apakah data uji dengan karakteristik tersebut termasuk ke dalam kategori masa studi cepat, tepat, atau telat.

Hitung nilai gini index untuk masing-masing sub-atribut

Tabel 4. Gini index Bekerja (ya) tiap class

Bekerja (Ya)	CEPAT	TEPAT	TELAT	TOTAL	Gini Index	Gini Gain
TRUE (LEFT NODE)	0	1	2	3	0.444	0.0925

FALSE (RIGHT NODE)	1	5	1	7	0.449
--------------------	---	---	---	---	-------

Nilai yang disajikan pada Tabel 4 diperoleh melalui proses perhitungan Gini Index pada setiap simpul hasil pemisahan (*split*) berdasarkan atribut status pekerjaan. Tahap pertama adalah menghitung nilai Gini Index pada left node, yaitu kelompok data dengan kategori pekerjaan "Ya". Berdasarkan data training, terdapat 3 data pada kelompok ini yang terdiri atas 0 mahasiswa dengan masa studi cepat, 1 mahasiswa dengan masa studi tepat, dan 2 mahasiswa dengan masa studi telat. Menggunakan rumus $Gini\ Index = 1 - \sum(P_j^2)$, diperoleh nilai Gini Index = $1 - ((0/3)^2 + (1/3)^2 + (2/3)^2) = 1 - (0 + 0,111 + 0,444) = 0,444$. Selanjutnya dilakukan perhitungan pada right node, yaitu kelompok data dengan kategori pekerjaan bukan "Ya" atau tidak bekerja, yang berjumlah 7 data. Komposisi kelas pada kelompok ini terdiri atas 1 mahasiswa dengan masa studi cepat, 5 mahasiswa dengan masa studi tepat, dan 1 mahasiswa dengan masa studi telat. Berdasarkan distribusi tersebut diperoleh nilai Gini Index = $1 - ((1/7)^2 + (5/7)^2 + (1/7)^2) = 1 - (0,020 + 0,510 + 0,020) = 0,449$. Setelah nilai Gini Index pada parent node sebesar 0,54, left node sebesar 0,444, dan right node sebesar 0,449 diperoleh, langkah berikutnya adalah menghitung Gini Gain untuk mengetahui kualitas atribut status pekerjaan dalam memisahkan data. Perhitungan dilakukan menggunakan rumus $Gini\ Gain = I(\text{parent}) - (N_{\text{left}}/N \times I_{\text{left}}) - (N_{\text{right}}/N \times I_{\text{right}})$ sehingga diperoleh hasil Gini Gain = $0,54 - (3/10 \times 0,444) - (7/10 \times 0,449) = 0,54 - 0,1332 - 0,3143 = 0,0925$. Nilai Gini Gain sebesar 0,0925 menunjukkan bahwa atribut status pekerjaan mampu mengurangi tingkat ketidakmurnian data sebesar 0,0925 setelah dilakukan pemisahan. Dalam algoritma *Classification and Regression Tree* (CART), semakin besar nilai Gini Gain yang dihasilkan oleh suatu atribut, semakin baik atribut tersebut dalam membedakan kelas-kelas target. Oleh karena itu, nilai ini selanjutnya dibandingkan dengan nilai Gini Gain dari atribut lainnya, seperti domisili, asal sekolah, serta Indeks Prestasi (IP) Semester 1 hingga Semester 6, untuk menentukan atribut terbaik yang akan dipilih sebagai simpul pada pohon keputusan. Hasil perbandingan tersebut menjadi dasar dalam proses pembentukan struktur pohon keputusan yang optimal untuk mengklasifikasikan masa studi mahasiswa ke dalam kategori cepat, tepat, atau telat.

Tabel 5. Gini index Karawang tiap class

Domisili Tinggal (Karawang)	CEPAT	TEPAT	TELAT	TOTAL	Gini Index	Gini Gain
TRUE (LEFT NODE)	1	5	1	7	0.449	0.0925
FALSE (RIGHT NODE)	0	1	2	3	0.444	

Nilai yang disajikan pada Tabel 5 diperoleh melalui proses perhitungan Gini Index pada setiap simpul hasil pemisahan berdasarkan atribut domisili, khususnya pada subatribut Karawang. Tahap pertama dilakukan dengan menghitung Gini Index pada left node, yaitu kelompok data mahasiswa yang berdomisili di Karawang, yang terdiri atas 7 data dengan distribusi kelas 1 mahasiswa memiliki masa studi cepat, 5 mahasiswa memiliki masa studi tepat, dan 1 mahasiswa memiliki masa studi telat. Berdasarkan distribusi tersebut, nilai Gini Index dihitung menggunakan rumus $Gini\ Index = 1 - \sum(P_j^2)$ sehingga diperoleh hasil $1 - ((1/7)^2 + (5/7)^2 + (1/7)^2) = 0,449$. Selanjutnya dilakukan perhitungan Gini Index pada right node, yaitu kelompok data mahasiswa yang berdomisili selain Karawang, yang terdiri atas 3 data dengan distribusi kelas 0 mahasiswa pada kategori cepat, 1 mahasiswa pada kategori tepat, dan 2 mahasiswa pada kategori telat, sehingga diperoleh nilai Gini Index sebesar 0,444. Setelah nilai Gini Index pada parent node sebesar 0,54, left node sebesar 0,449, dan right node sebesar 0,444 diketahui, langkah berikutnya adalah menghitung Gini Gain untuk mengukur kemampuan subatribut Karawang dalam memisahkan data. Perhitungan dilakukan menggunakan rumus $Gini\ Gain = I(\text{parent}) - (N_{\text{left}}/N \times I_{\text{left}}) - (N_{\text{right}}/N \times I_{\text{right}})$, sehingga diperoleh hasil Gini Gain = $0,54 - (7/10 \times 0,449) - (3/10 \times 0,444) = 0,0925$. Nilai Gini Gain sebesar 0,0925 menunjukkan bahwa pemisahan data berdasarkan subatribut Karawang mampu mengurangi tingkat ketidakmurnian data sebesar 0,0925. Nilai ini kemudian dibandingkan dengan nilai Gini Gain dari subatribut maupun atribut lainnya dalam algoritma *Classification and Regression Tree* (CART) untuk menentukan atribut yang paling optimal sebagai simpul pada pohon keputusan. Semakin besar nilai Gini Gain yang dihasilkan, semakin baik kemampuan atribut tersebut dalam membedakan kelas masa studi mahasiswa, sehingga proses pembentukan pohon keputusan dapat menghasilkan model klasifikasi yang lebih akurat dan mudah diinterpretasikan.

Tabel 6. Gini index SMK tiap class

Status Sekolah (SMK)	CEPAT	TEPAT	TELAT	TOTAL	Gini Index	Gini Gain
						0.04
TRUE (<i>LEFT NODE</i>)	1	4	1	6	0.5	
FALSE (<i>RIGHT NODE</i>)	0	2	2	4	0.5	

Nilai yang disajikan pada Tabel 6 diperoleh melalui proses perhitungan Gini Index berdasarkan atribut asal sekolah, khususnya pada subatribut SMK. Tahap pertama dilakukan dengan menghitung Gini Index pada left node, yaitu kelompok mahasiswa yang berasal dari SMK, yang terdiri atas 6 data dengan distribusi kelas 1 mahasiswa memiliki masa studi cepat, 4 mahasiswa memiliki masa studi tepat, dan 1 mahasiswa memiliki masa studi telat. Berdasarkan distribusi tersebut, nilai Gini Index dihitung menggunakan rumus $Gini\ Index = 1 - \sum (P_j^2)$ sehingga diperoleh hasil $1 - ((1/6)^2 + (4/6)^2 + (1/6)^2) = 0,50$. Selanjutnya dilakukan perhitungan Gini Index pada right node, yaitu kelompok mahasiswa yang berasal dari selain SMK, yang berjumlah 4 data dengan komposisi kelas 0 mahasiswa pada kategori cepat, 2 mahasiswa pada kategori tepat, dan 2 mahasiswa pada kategori telat. Berdasarkan komposisi tersebut diperoleh nilai $Gini\ Index = 1 - ((0/4)^2 + (2/4)^2 + (2/4)^2) = 0,50$. Setelah nilai Gini Index pada parent node sebesar 0,54, left node sebesar 0,50, dan right node sebesar 0,50 diketahui, langkah berikutnya adalah menghitung Gini Gain untuk mengetahui kemampuan subatribut SMK dalam memisahkan data. Perhitungan dilakukan menggunakan rumus $Gini\ Gain = I(\text{parent}) - (N_{\text{left}}/N \times I_{\text{left}}) - (N_{\text{right}}/N \times I_{\text{right}})$ sehingga diperoleh hasil $Gini\ Gain = 0,54 - (6/10 \times 0,50) - (4/10 \times 0,50) = 0,04$. Nilai Gini Gain sebesar 0,04 menunjukkan bahwa pemisahan data berdasarkan subatribut SMK hanya mampu mengurangi tingkat ketidakmurnian data dalam jumlah yang relatif kecil dibandingkan atribut lainnya. Dengan demikian, kemampuan atribut asal sekolah pada subatribut SMK dalam membedakan kategori masa studi mahasiswa tergolong lebih rendah sehingga kemungkinan tidak dipilih sebagai simpul utama dalam pembentukan pohon keputusan algoritma Classification and Regression Tree (CART). Proses ini menjadi bagian penting dalam menentukan atribut yang paling optimal untuk menghasilkan model klasifikasi masa studi mahasiswa yang memiliki tingkat akurasi dan kemampuan prediksi yang lebih baik.

Pembahasan

Berdasarkan hasil perhitungan algoritma *Classification and Regression Tree* (CART), proses klasifikasi masa studi mahasiswa diawali dengan menghitung nilai Gini Index pada simpul induk (parent node) yang menghasilkan nilai sebesar 0,54. Nilai tersebut menunjukkan bahwa distribusi data pada node awal masih bersifat heterogen sehingga diperlukan proses pemisahan (splitting) untuk memperoleh kelompok data yang lebih homogen. Dalam algoritma CART, Gini Index digunakan sebagai ukuran impurity atau ketidakmurnian suatu node, di mana semakin kecil nilai Gini Index maka semakin baik kualitas pemisahan data yang dihasilkan. Dengan demikian, tahapan evaluasi setiap atribut menjadi langkah yang sangat penting dalam menentukan atribut terbaik sebagai akar pohon keputusan. Pendekatan ini sejalan dengan penelitian Setiawan dan Wibowo (2022) yang menyatakan bahwa penggunaan Gini Index pada algoritma CART mampu menghasilkan proses seleksi atribut yang objektif sehingga meningkatkan interpretabilitas model klasifikasi pada data akademik mahasiswa. Selain itu, Pratama, et al., (2023) menjelaskan bahwa proses pengukuran impurity melalui Gini Index menjadi dasar utama dalam menghasilkan decision tree yang sederhana namun memiliki kemampuan prediksi yang baik.

Hasil perhitungan pada atribut status pekerjaan menunjukkan nilai Gini Gain sebesar 0,0925, yang berarti atribut tersebut mampu menurunkan tingkat impurity sebesar 0,0925 setelah dilakukan pemisahan data. Walaupun nilai tersebut belum tergolong tinggi, atribut status pekerjaan telah memperlihatkan kemampuan dalam membedakan kategori masa studi mahasiswa. Secara konseptual, kondisi ini mengindikasikan bahwa mahasiswa yang bekerja memiliki kecenderungan pola penyelesaian studi yang berbeda dibandingkan mahasiswa yang tidak bekerja. Faktor pekerjaan dapat memengaruhi alokasi waktu belajar, intensitas mengikuti perkuliahan, maupun penyelesaian tugas akademik sehingga berdampak terhadap lama studi mahasiswa. Temuan ini didukung oleh penelitian Sari dan Kurniawati (2021) yang menyimpulkan bahwa aktivitas bekerja selama masa kuliah menjadi salah satu variabel yang berkontribusi terhadap keberhasilan model klasifikasi kelulusan mahasiswa menggunakan algoritma pohon keputusan. Senada dengan itu, Rahman, et al., (2023) menjelaskan bahwa variabel non-akademik seperti status pekerjaan dapat meningkatkan kemampuan model prediksi ketika dikombinasikan dengan atribut akademik.

Pada atribut domisili, diperoleh nilai Gini Gain sebesar 0,0925, yaitu sama dengan atribut status pekerjaan. Hasil tersebut menunjukkan bahwa domisili mahasiswa memiliki kemampuan yang relatif sama dalam memisahkan data berdasarkan kategori masa studi. Mahasiswa yang berdomisili di Karawang memiliki distribusi kelas yang berbeda dibandingkan mahasiswa dari luar Karawang sehingga menghasilkan penurunan impurity yang cukup baik. Secara substantif, kondisi ini dapat dipengaruhi oleh faktor jarak tempuh menuju kampus, biaya transportasi, maupun

kemudahan akses terhadap kegiatan akademik. Mahasiswa yang tinggal lebih dekat dengan kampus umumnya memiliki peluang lebih besar untuk mengikuti proses pembelajaran secara optimal sehingga dapat menyelesaikan studi tepat waktu. Hasil penelitian ini sejalan dengan Nugraha dan Saputra (2022) yang menemukan bahwa faktor domisili memberikan kontribusi terhadap prediksi keberhasilan studi mahasiswa karena berkaitan dengan intensitas kehadiran dan efektivitas aktivitas belajar. Penelitian Utami, et al., (2024) juga mengemukakan bahwa atribut lokasi tempat tinggal merupakan salah satu variabel yang cukup berpengaruh dalam model klasifikasi data pendidikan menggunakan metode CART.

Berbeda dengan dua atribut sebelumnya, atribut asal sekolah (SMK) menghasilkan nilai Gini Gain sebesar 0,04, yang menunjukkan kemampuan pemisahan data relatif rendah. Nilai tersebut mengindikasikan bahwa distribusi mahasiswa lulusan SMK maupun non-SMK masih memiliki tingkat heterogenitas yang hampir sama sehingga atribut ini kurang efektif dijadikan simpul utama pada pembentukan pohon keputusan. Dengan kata lain, latar belakang pendidikan sebelum memasuki perguruan tinggi belum mampu membedakan secara signifikan kategori masa studi mahasiswa pada dataset yang digunakan. Hasil tersebut menunjukkan bahwa keberhasilan mahasiswa menyelesaikan studi tidak hanya ditentukan oleh asal sekolah, tetapi lebih dipengaruhi oleh faktor-faktor akademik maupun karakteristik individu selama menjalani perkuliahan. Temuan ini sesuai dengan penelitian Kusuma dan Lestari (2021) yang menyatakan bahwa variabel asal sekolah memiliki kontribusi yang relatif kecil dibandingkan variabel prestasi akademik dalam model klasifikasi kelulusan mahasiswa. Penelitian Ananda dan Fitriani (2023) juga menyebutkan bahwa atribut akademik pada masa perkuliahan lebih dominan dibandingkan karakteristik sebelum masuk perguruan tinggi.

Secara keseluruhan, hasil perhitungan Gini Gain memperlihatkan bahwa atribut status pekerjaan dan domisili memiliki kemampuan pemisahan data yang lebih baik dibandingkan atribut asal sekolah. Dalam algoritma CART, atribut dengan nilai Gini Gain tertinggi akan dipilih sebagai simpul pada pohon keputusan karena mampu menghasilkan node yang lebih homogen. Oleh sebab itu, proses evaluasi seluruh atribut secara bertahap menjadi mekanisme penting untuk memperoleh struktur pohon keputusan yang optimal. Model yang dihasilkan tidak hanya memiliki tingkat akurasi yang baik, tetapi juga mudah dipahami karena aturan keputusan (decision rules) dapat ditelusuri secara sistematis. Hal ini sejalan dengan penelitian Hidayat dan Kurniawan (2022) yang menyatakan bahwa algoritma CART memiliki keunggulan dalam menghasilkan model klasifikasi yang sederhana, mudah diinterpretasikan, serta efektif digunakan pada data pendidikan. Penelitian Fauzi dan Ramadhani (2024) juga menunjukkan bahwa pemanfaatan CART pada data akademik mampu membantu institusi pendidikan dalam mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan studi sejak dini.

Hasil penelitian ini memberikan implikasi praktis bagi perguruan tinggi dalam mengembangkan sistem peringatan dini (early warning system) terhadap mahasiswa yang berisiko mengalami keterlambatan kelulusan. Informasi mengenai atribut-atribut yang memiliki kontribusi terbesar dalam proses klasifikasi dapat dimanfaatkan sebagai dasar penyusunan kebijakan akademik, seperti pemberian layanan bimbingan belajar, konseling akademik, maupun pendampingan bagi mahasiswa yang bekerja atau berasal dari luar daerah. Dengan demikian, implementasi algoritma CART tidak hanya berfungsi sebagai metode klasifikasi, tetapi juga sebagai alat pendukung pengambilan keputusan (decision support system) dalam meningkatkan kualitas layanan akademik. Temuan ini memperkuat hasil penelitian Wahyuni, et al., (2023) yang menyatakan bahwa data mining berbasis CART sangat potensial digunakan sebagai sistem prediksi keberhasilan studi mahasiswa, serta penelitian Putra dan Firmansyah (2024) yang menyimpulkan bahwa penerapan algoritma klasifikasi pada data akademik dapat membantu perguruan tinggi dalam meningkatkan efektivitas evaluasi dan pengambilan keputusan akademik.

4. Kesimpulan

Berdasarkan hasil perhitungan algoritma Classification and Regression Tree (CART), dapat disimpulkan bahwa proses klasifikasi masa studi mahasiswa dilakukan secara efektif melalui evaluasi nilai Gini Index dan Gini Gain untuk setiap atribut prediktor. Hasil penelitian menunjukkan bahwa atribut status pekerjaan dan domisili memiliki nilai Gini Gain tertinggi, yaitu sebesar 0,0925, sehingga keduanya memiliki kemampuan yang lebih baik dalam memisahkan data dibandingkan atribut asal sekolah (SMK) yang hanya memperoleh nilai Gini Gain sebesar 0,04. Temuan ini menunjukkan bahwa faktor pekerjaan dan domisili lebih berpengaruh dalam menentukan kategori masa studi mahasiswa, sedangkan asal sekolah memiliki kontribusi yang relatif kecil terhadap proses klasifikasi. Dengan demikian, algoritma CART mampu menghasilkan model klasifikasi yang sederhana, mudah diinterpretasikan, serta dapat dimanfaatkan sebagai dasar pengembangan sistem pendukung keputusan dan *early warning system* bagi perguruan tinggi untuk mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan studi sehingga tindakan pembinaan akademik dapat dilakukan secara lebih dini dan tepat sasaran.

Daftar Pustaka

- Ananda, R., & Fitriani, D. (2023). Analisis pengaruh atribut akademik terhadap prediksi kelulusan mahasiswa menggunakan algoritma decision tree. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(2), 315–324.
- Budiman, I., Muliadi, M., & Ramadina, R. (2015). Penerapan fungsi data mining klasifikasi untuk prediksi masa studi mahasiswa tepat waktu

DOI: <https://doi.org/10.69693/ijmst.v4i2.12433>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

-
- pada sistem informasi akademik perguruan tinggi. *JUPITER: Jurnal Penelitian Ilmu dan Teknologi Komputer*, 7(1), 39–50. <https://doi.org/10.5281/zenodo.3432716>
- Fauzi, M., & Ramadhani, R. (2024). Implementasi algoritma Classification and Regression Tree (CART) untuk prediksi masa studi mahasiswa. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 8(1), 45–55.
- Hidayat, A., & Kurniawan, R. (2022). Penerapan algoritma CART pada klasifikasi data pendidikan tinggi. *Jurnal Nasional Teknologi dan Sistem Informasi*, 8(3), 184–192.
- Jananto, A., Sulastri, S., Wahyudi, E. N., & Sunardi. (2021). Data induk mahasiswa sebagai prediktor ketepatan waktu lulus menggunakan algoritma CART klasifikasi data mining. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 10(1), 108–115. <https://doi.org/10.32736/sisfokom.v10i1.991>
- Kusuma, D., & Lestari, N. (2021). Analisis faktor-faktor yang memengaruhi kelulusan mahasiswa menggunakan metode data mining. *Jurnal Informatika*, 15(2), 95–104.
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi data mining untuk klasifikasi masa studi mahasiswa menggunakan algoritma *K-Nearest Neighbor*. *Simetris: Jurnal Teknik Mesin, Industri, Elektro dan Ilmu Komputer*, 10(2), 421–432. <https://doi.org/10.24176/simet.v10i2.2882>
- Nugraha, A., & Saputra, D. (2022). Prediksi masa studi mahasiswa menggunakan algoritma CART berbasis data akademik. *Jurnal Teknologi Informasi dan Komunikasi*, 13(1), 56–66.
- Prabawati, N. I., Widodo, & Ajie, H. (2019). Kinerja algoritma *Classification and Regression Tree* (CART) dalam mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta. *PINTER: Jurnal Pendidikan Teknik Informatika dan Komputer*, 3(2), 86–94. <https://doi.org/10.21009/pinter.3.2.9>
- Pratama, A., Nugroho, S., & Kurniawan, B. (2023). Analisis performa algoritma CART pada klasifikasi data akademik mahasiswa. *Jurnal Media Informatika Budidarma*, 7(3), 1328–1337.
- Putra, R., & Firmansyah, A. (2024). Implementasi data mining untuk mendukung pengambilan keputusan akademik menggunakan algoritma CART. *Jurnal Ilmiah Informatika*, 12(1), 88–98.
- Rafika, C. S., Reynanda, R. M., & Sari, A. P. (2024). Penerapan decision tree CART untuk klasifikasi risiko gagal studi mahasiswa. *Jurnal Mahasiswa Teknik Informatika*, 4(2). <https://doi.org/10.35473/jamastika.v4i2.4145>
- Rahman, F., Hidayat, T., & Yusuf, M. (2023). Pemanfaatan atribut akademik dan non-akademik untuk prediksi kelulusan mahasiswa menggunakan decision tree. *Jurnal Sistem Informasi*, 19(2), 112–123.
- Sari, E., & Kurniawati, D. (2021). Klasifikasi masa studi mahasiswa menggunakan algoritma decision tree CART. *Jurnal Informatika Upgris*, 7(2), 134–142.
- Setiawan, B., & Wibowo, A. (2022). Implementasi algoritma CART untuk analisis data akademik mahasiswa. *Jurnal Teknik Informatika dan Sistem Informasi*, 9(1), 56–67.
- Utami, N., Prasetyo, H., & Widodo, E. (2024). Perbandingan metode klasifikasi CART pada data pendidikan tinggi. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8(2), 701–710.
- Wahyuni, S., Suryadi, A., & Maulana, F. (2023). Sistem prediksi kelulusan mahasiswa berbasis Classification and Regression Tree. *Jurnal Informatika Polinema*, 9(4), 321–330.