

Klasifikasi Profil Mahasiswa Pada Pola Nilai Akpam Dengan Metode *Decision Tree Cart*

Dihin Muriyatmoko¹, Aziz Musthafa², Salman Alfarisi Ikrima³

Teknik Informatika

Universitas Darussalam Gontor

Ponorogo, Indonesia

e-mail: ¹dihin@unida.gontor.ac.id, ²aziz@unida.gontor.ac.id, ³salmanikrma31@mhs.unida.gontor.ac.id

Correspondence : e-mail: dihin@unida.gontor.ac.id

Diajukan: 21 Agustus 2024; Direvisi: 24 Agustus 2024; Diterima: 25 Agustus 2024

Abstrak

Universitas Darussalam (UNIDA) Gontor merupakan perguruan tinggi ber sistem pesantren yang berada di bawah naungan Pondok Modern Darussalam Gontor. Dengan memadukan sistem pesantren dengan berbagai aktivitas yang lengkap, santri akan tetap melakukan beberapa aktivitas yang ada. Salah satu alat untuk memantau kegiatan tersebut adalah Angka Kumulatif Penunjang Akademik Mahasiswa (AKPAM). Penelitian ini bertujuan untuk mengklasifikasi pola nilai AKPAM di UNIDA Gontor menggunakan algoritma Decision Tree CART. Data AKPAM dari tahun 2020-2022 dengan total 11.978 data dikumpulkan dan diproses melalui tahap pembersihan, transformasi, dan pengurangan data. Pembagian data dilakukan dengan rasio 80% untuk pelatihan dan 20% untuk pengujian. Model Decision Tree CART dibangun dengan kriteria Gini Index dan kedalaman maksimal 10. Hasil pengujian menunjukkan akurasi klasifikasi yang baik di setiap fakultas di UNIDA Gontor, yaitu fakultas Ekonomi 99,70%, Humaniora 98,87%, Kesehatan 99,13%, Saintek 97,93%, Syariah 99,22%, Tarbiyah 99,56%, dan Ushuluddin 99,33%. Struktur Decision Tree mengindikasikan bahwa variabel keilmuan, kerohanian, dan keorganisasian memiliki pengaruh paling besar terhadap kelulusan AKPAM. Penelitian ini menyediakan rekomendasi bagi mahasiswa UNIDA Gontor untuk meningkatkan partisipasi dalam kegiatan yang sesuai dengan kriteria AKPAM guna mencapai nilai kelulusan yang optimal.

Kata kunci: AKPAM, Klasifikasi, Penambangan Data, Pohon Keputusan CART, Gini Indeks.

Abstract

Universitas Darussalam (UNIDA) Gontor is a pesantren system university under the auspices of Pondok Modern Darussalam Gontor. By combining the pesantren system with a complete range of activities, students will continue to carry out several existing activities. One of the tools to monitor these activities is the Student Academic Support Cumulative Score (AKPAM). This research aims to classify AKPAM score patterns at UNIDA Gontor using the CART Decision Tree algorithm. AKPAM data from 2020-2022 with a total of 11,978 data were collected and processed through the stages of data cleaning, transformation, and reduction. Data division is done with a ratio of 80% for training and 20% for testing. The CART Decision Tree model was built with Gini Index criteria and a maximum depth of 10. The test results showed good classification accuracy in each faculty at UNIDA Gontor, namely the faculties of Economics 99.70%, Humanities 98.87%, Health 99.13%, Science 97.93%, Sharia 99.22%, Tarbiyah 99.56%, and Ushuluddin 99.33%. The Decision Tree structure indicates that scientific, spiritual, and organizational variables have the greatest influence on AKPAM graduation. This research provides recommendations for UNIDA Gontor students to increase participation in activities that fit the AKPAM criteria in order to achieve optimal graduation scores.

Keywords: AKPAM, Classification, Data Mining, Decision tree CART, Gini Index.

1. Pendahuluan

Universitas Darussalam (UNIDA) Gontor merupakan perguruan tinggi ber sistem pesantren yang berada di bawah naungan Pondok Modern Darussalam Gontor. [1] UNIDA Gontor mengintegrasikan sistem pesantren dengan berbagai organisasi dan unit kegiatan kemahasiswaan, memungkinkan santri tetap aktif dalam berbagai kegiatan di luar jam perkuliahan. Salah satu alat untuk memantau kegiatan tersebut adalah

Angka Kumulatif Penunjang Akademik Mahasiswa (AKPAM) yang dilaksanakan oleh Universitas Darussalam Gontor. AKPAM menjadi acuan kegiatan mahasiswa selama menempuh studi di UNIDA Gontor. Lulus AKPAM merupakan syarat bagi mahasiswa untuk mengikuti ujian akhir semester.

Kriteria penjurian AKPAM terbagi dalam beberapa kategori antara lain keilmuan, kerohanian, keorganisasian, kesenian dan olah raga serta perlombaan dan delegasi. Dengan kriteria di atas maka mahasiswa harus sepenuhnya memenuhi kriteria evaluasi dengan jumlah skor minimum dari total semua kategori adalah 200 poin, namun dalam hal ini masih banyak siswa yang kesulitan mencapai nilai AKPAM.

Berdasarkan data AKPAM yang dihimpun melalui sistem siacad UNIDA Gontor pada tahun ajaran 2022-2023, tercatat masih ada sekitar 700 mahasiswa yang belum memenuhi ambang batas minimal untuk memperoleh gelar AKPAM. Sedangkan pada tahun ajaran 2022-2023, siswa yang belum lulus AKPAM berjumlah 379 orang. Data di atas menjadi bukti masih terlalu sulitnya siswa meraih nilai AKPAM. Hal ini dikarenakan belum adanya model kelulusan siswa berdasarkan data siswa yang lulus AKPAM, karena model pelatihan atau pemodelan dalam pencatatan nilai AKPAM dapat dilakukan dengan menggunakan metode *data mining* salah satunya dengan metode *decision tree CART*

Data mining adalah disiplin ilmu yang fokus pada proses ekstraksi pengetahuan atau pola dari basis data.[2] Dengan memanfaatkan analisis statistik, matematika, dan kecerdasan buatan, penambangan data memungkinkan penemuan pola dan tren dalam informasi yang tidak langsung terlihat. Ini terdiri dari langkah-langkah seperti pemrosesan data, transformasi data, pemilihan fitur, dan penggunaan algoritme dan model yang tepat, termasuk teknik yang umum digunakan seperti klasifikasi.

Algoritma Clasification and Regression Trees (CART) adalah salah satu jenis metode *decision tree*. Dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen, dan Charles J. Stone, metode ini merupakan teknik klasifikasi yang menggunakan algoritma partisi rekursif biner.[3] Tahapan algoritma CART terdiri dari : Pada tahap pertama, dibuat daftar calon cabang untuk semua variabel prediktor secara keseluruhan, serta daftar calon cabang yang sedang dipertimbangkan saat ini, Pada tahap kedua, kinerja keseluruhan kandidat industri dalam daftar dari Tahap pertama dievaluasi dengan menghitung nilai *goodness-of-fit*, Pada tahap terakhir, calon cabang dipilih menjadi cabang berdasarkan skor kesesuaian (*goodness score*) tertinggi. Proses algoritma CART berakhir ketika tidak ada lagi noktah keputusan; jika tidak, algoritma kembali ke tahap kedua dan mengevaluasi ulang kandidat cabang, menghasilkan daftar kandidat cabang baru. Selama pengembangan *decision tree*, nilai *impurity*, yang dirumuskan menggunakan indeks Gini, digunakan untuk memilih atribut yang akan dipisahkan.[4] Untuk kumpulan data m kelas, indeks Gini atribut D dapat dirumuskan sebagai berikut :

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2$$

Nilai P_i adalah probabilitas suatu kelas dalam kumpulan data, dihitung dengan membagi jumlah atribut kelas C_i pada suatu record data dengan jumlah total record data.[4] Pembagian dataset menjadi dua subset, D_1 dan D_2 , dengan *Gini index* terendah dapat dirumuskan sebagai berikut:

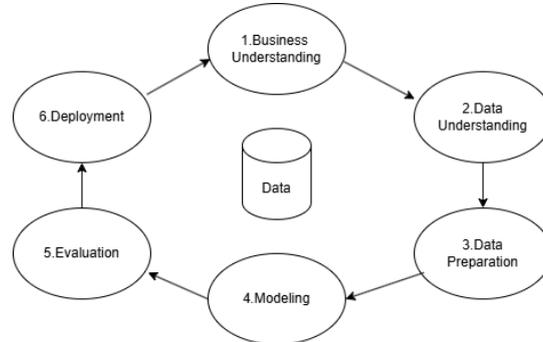
Penelitian ini menggunakan K-Fold *cross validation* untuk membagi data menjadi dua jenis, data pelatihan dan data pengujian. Teknik K-fold *cross validation* populer di kalangan peneliti karena efektif dalam mengurangi bias saat pengambilan sampel. Metode ini secara berulang membagi data menjadi data pelatihan dan pengujian, memastikan setiap data memiliki kesempatan untuk menjadi data pengujian. Nilai K menentukan jumlah pengulangan dalam pembagian data tersebut.[5]

$$Gini_A = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2)$$

Salah satu penelitian yang berkaitan dengan klasifikasi kelulusan mahasiswa menggunakan algoritma K-NN, NBC, dan SVM dalam prediksi kelulusan mahasiswa tingkat akhir. Penelitian ini bertujuan memprediksi kelulusan mahasiswa tingkat akhir. Pengujian ini membandingkan algoritma K-NN, NBC, dan SVM, penelitian melibatkan akuisisi data, *preprocessing*, *splitting data*, pembuatan model, dan uji coba. Algoritma K-NN memiliki akurasi, presisi, dan recall rata-rata lebih tinggi dibandingkan NBC dan SVM dalam prediksi kelulusan mahasiswa, dengan akurasi 87,8%, presisi 87,8%, dan *recall* 84%[6]. Perbedaan penelitian saat ini dapat dilihat dari penelitian tersebut menggunakan komparasi banyak algoritma sedangkan penelitian ini hanya menggunakan algoritma CART. Tujuan penelitian ini adalah untuk mengklasifikasi pola pada nilai akpam untuk memudahkan para mahasiswa mencapai standar nilai akpam.

2. Metode Penelitian

Penelitian ini menggunakan *Cross Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM adalah kerangka pemrosesan data populer yang membantu pengguna memahami dan mengelola proyek penambangan data secara sistematis. Model ini terdiri dari enam fase: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan terakhir penerapan.[8] Pada Gambar 1.



Gambar 1. CRISP-DM

2.1. Business Understanding

Pada tahap *Business Understanding*, tujuan utamanya adalah meningkatkan persentase kelulusan AKPAM siswa atau memberikan rekomendasi kegiatan untuk hasil yang lebih baik. Keberhasilan diukur berdasarkan ketepatan model klasifikasi yang dihasilkan. Selain itu, analisis *stakeholder* dilakukan untuk mengidentifikasi pihak-pihak terkait seperti mahasiswa, fakultas, staf, dan administrator universitas, sehingga solusi yang tepat dapat dirancang sesuai dengan kebutuhan dan harapan mereka.

2.2. Data Understanding

Data yang dikumpulkan mencakup rekapan AKPAM tiap semester dari tahun 2020 hingga 2022, dengan total 11.978 data. Data ini berasal dari seluruh fakultas di Universitas Darussalam Gontor Kampus Siman, yang terdiri dari 7 fakultas. Dengan beberapa kriteria : keilmuan, kerohanian, keorganisasian, olahraga dan kesenian, kemasyarakatan, dan pengalaman dan lomba. Rekapan data tersebut adalah data yang relevan digunakan pada penelitian, dengan sampel data yang ada pada gambar 2.

Nim	Nama	Prodi	Keilmuan	Kerohanian	Kesenian dan Keolahragaan	Keorganisasian	Kemasyarakatan	Pengalaman / Lomba / Lain-lain	Total	Keterangan
3.72016E+11	A. RIZA TAUFIQURROHMAN	Pendidikan Agama Islam	0	0	0	0	0	0	0	BELUM LULUS
3.72016E+11	ABID HIFNI MUHAMMAD	Pendidikan Agama Islam	0	0	0	0	0	0	0	BELUM LULUS
3.72016E+11	AHMAD BASMEH ROMADHON	Pendidikan Agama Islam	0	0	0	0	0	0	0	BELUM LULUS
3.72016E+11	AHMAD ZAKIAN NUR FAUZAN	Pendidikan Agama Islam	0	0	0	0	0	0	0	BELUM LULUS
3.72016E+11	AKSANI TAQWIM	Pendidikan Agama Islam	0	0	0	0	0	0	0	BELUM LULUS
3.72016E+11	ALVHI SAHRIN GUNANDAR	Pendidikan Agama Islam	0	0	0	0	0	0	0	BELUM LULUS
3.72016E+11	ARI SUSANTO	Pendidikan Agama Islam	0	0	0	0	0	0	0	BELUM LULUS
3.72016E+11	AZIZI DWI SYAIFURRIZQI	Pendidikan Agama Islam	0	0	0	0	0	0	0	BELUM LULUS
3.72016E+11	DEDI RUWANDA	Pendidikan Agama Islam	0	45	15	45	15	20	140	BELUM LULUS
3.72016E+11	DENI ANGGRAYANI	Pendidikan Agama Islam	0	0	0	0	0	0	0	BELUM LULUS
3.72016E+11	EKA NURSETIADI	Pendidikan Agama Islam	0	0	0	0	0	0	0	BELUM LULUS
3.72016E+11	ELI ELFIAN	Pendidikan Agama Islam	60	45	15	45	15	20	200	LULUS

Gambar 2. Data Kotor

Dari data tersebut, 7.335 mahasiswa dinyatakan lulus, sementara 4.643 lainnya gagal. Seperti pada tabel 1.

Tabel 1. Data Per Fakultas

	Ekonomi	Humaniora	Kesehatan	Saintek	Syariah	Tarbiyah	Ushuluddin	Total
Lulus	1016	1291	403	1120	565	1446	1494	7335
Belum lulus	981	658	59	281	455	1310	899	4643
Jumlah per fakultas	1997	1949	462	1401	1020	2756	2393	11978

2.3. Data Preparation

Dalam penelitian ini, Altair AI Studio digunakan untuk menyederhanakan proses persiapan data dan melakukan *preprocessing data*. *Preprocessing data* adalah proses yang dilakukan pada data mentah sebagai persiapan untuk operasi pemrosesan berikutnya[9], langkah pertama adalah pembersihan data (*data cleaning*), pembersihan data adalah proses menghilangkan data yang tidak lengkap dan tidak akurat dari suatu pengumpulan data.[10] Selanjutnya, transformasi data (*data transformation*) data ditransformasikan ke dalam bentuk yang sesuai untuk *data mining* melalui proses seperti normalisasi, pemilihan atribut,

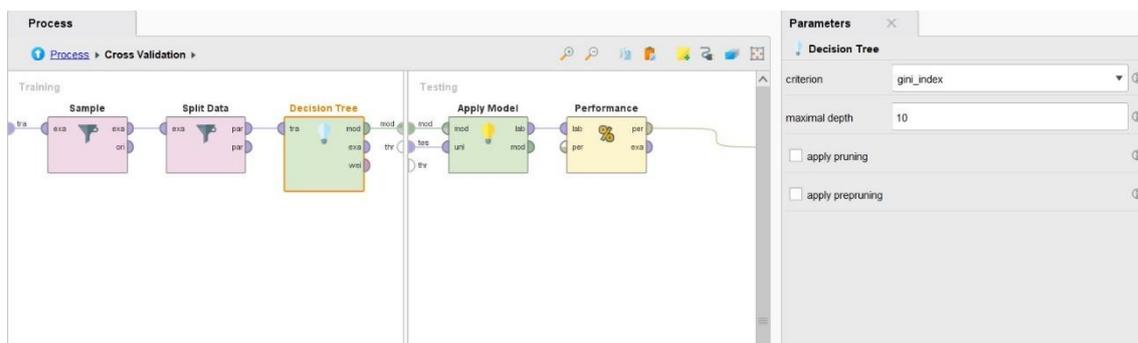
diskretisasi, dan generasi hierarki. Kemudian, dilakukan pengurangan data (*data reduction*) untuk mengeliminasi data yang tidak relevan. Selanjutnya proses penyeimbangan data (*data balancing*) untuk memastikan data yang diolah seimbang dan adil. Setelah itu, peneliti melakukan pembagian data (*split data*) data dibagi menjadi dua bagian: 80% untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*). Data pelatihan digunakan untuk melatih algoritma, dan data pengujian digunakan untuk mengevaluasi kinerja algoritma.[6]

3. Hasil dan Pembahasan

Pada tahapan ini ada pembahasan serta implementasi hasil klasifikasi keberhasilan AKPAM berdasarkan kriteria yang telah ada. Teknik *Decision Tree CART* digunakan untuk mengklasifikasikan, sementara *cross validation* dengan data *testing* digunakan untuk pengujian.

3.1. Modeling

Pemodelan *Decision Tree* dalam penelitian ini menggunakan algoritma CART. Algoritma CART adalah metodologi statistik nonparametrik yang dikembangkan untuk analisis klasifikasi variabel respon kategorikal atau kontinu[11] yang membangun pohon keputusan dari data pelatihan melalui tiga langkah: pemilihan fitur terbaik, pemisahan data, dan rekursi. Algoritma ini juga merupakan teknik klasifikasi yang menggunakan algoritma pemisahan rekursif biner (*binary recursive partitioning*)[3]. Selama tahap pengembangan *decision tree* sebagai model CART, nilai *impurity* digunakan untuk memilih atribut yang dipisahkan. Nilai *impurity* CART dapat dirumuskan dalam bentuk nilai *Gini index*[4]. *Gini index* menghitung seberapa baik masing-masing kelas muncul dalam data. Pohon keputusan yang dihasilkan memiliki *maximal depth* 10, dengan *Gini Index* sebagai kriteria utama dalam pemisahan data. Seperti pada gambar 3.



Gambar 3. Modelling *Decision Tree CART*

3.2. Evaluation

Dalam tahap ini dilakukan validasi dan evaluasi serta pengukuran keakuratan hasil yang dicapai oleh model menggunakan *confusion matrix* dengan metode *K-Fold cross validation*. Setelah melakukan serangkaian proses seperti *preprocessing data*, *split data*, pemodelan, dan validasi menggunakan *cross-validation* pada setiap fakultas, uji coba pada penelitian ini mencapai akurasi yang baik di setiap fakultasnya. Pada fakultas Ekonomi mendapat akurasi 99,70%, fakultas Humaniora mendapat akurasi 98,87%, fakultas Kesehatan mendapat akurasi 99,13%, fakultas Sainstek mendapat akurasi 97,93%, fakultas Syariah mendapat akurasi 99,22%, fakultas Tarbiyah mendapat akurasi 99,56%, fakultas Ushuluddin mendapat akurasi 99,33%.

Untuk mengukur tingkat akurasi dari algoritma klasifikasi yang digunakan yaitu menggunakan *confusion matrix*, salah satu hasil *confusion matrix* disajikan pada gambar 3. Gambar 3 merupakan hasil *confusion matrix* dari pengujian dataset AKPAM fakultas Sains dan Teknologi menggunakan algoritma CART dengan metode *K-fold cross validation*.

PerformanceVector

```

PerformanceVector:
accuracy: 97.93% +/- 1.63% (micro average: 97.93%)
ConfusionMatrix:
True:   LULUS  BELUM LULUS
LULUS: 1097   6
BELUM LULUS: 23   275
    
```

Gambar 4. Hasil *Confusion Matrix* Fakultas Sains dan Teknologi

yang menunjukkan bahwa aspek agama atau spiritual juga penting dalam memutuskan apakah seorang mahasiswa akan lulus.

3. Keorganisasian : Meskipun variabel organisasi tidak sekuat keilmuan dan kerohanian, namun tetap mempunyai pengaruh yang cukup besar terhadap hasil yang diprediksi. Terdapat beberapa pembagian di dalam *decision tree* yang bergantung pada nilai "Keorganisasian", Nilai-nilai keorganisasian menunjukkan bahwa partisipasi dalam kegiatan organisasi atau ekstrakurikuler mempengaruhi prediksi lulus atau tidak lulus.

Penelitian ini masih banyak kekurangan dan hanya menggunakan algoritma CART, untuk penelitian lebih lanjut diharapkan menambah data dari periode lebih panjang untuk meningkatkan generalisasi hasil penelitian. Selain itu, metode klasifikasi lain seperti SVM, *Neural Networks*, dan lainnya perlu dipertimbangkan untuk membandingkan kinerja model secara menyeluruh.

Daftar Pustaka

- [1] H. F. Zarkasyi, F. Mas'ud, R. A. Hidayatullah, and U. Khakim, "Value Management in Pesantren-Based University (Grounded Research on AKPAM System of UNIDA Gontor)," *QALAMUNA J. Pendidikan, Sos. dan Agama*, vol. 15, no. 2, pp. 113–1126, 2023, doi: 10.37680/qalamuna.v15i2.4018.
- [2] A. Rohman and M. Rochcham, "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa," *Neo Tek.*, vol. 5, no. 1, pp. 23–29, 2019, doi: 10.37760/neoteknika.v5i1.1379.
- [3] S. H. Sumartini, "Penggunaan Metode Classification and Regression Trees (CART) untuk Klasifikasi Rekrutasi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya," *J. Sains dan Seni ITS*, vol. 4, no. 2, pp. 211–216, 2015, [Online]. Available: <https://journal.universitasmuhammadia.ac.id/index.php/matrik/article/view/676/479>
- [4] A. Jananto, S. Sulastri, E. Nur Wahyudi, and S. Sunardi, "Data Induk Mahasiswa sebagai Prediktor Ketepatan Waktu Lulus Menggunakan Algoritma CART Klasifikasi Data Mining," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 1, pp. 71–78, 2021, doi: 10.32736/sisfokom.v10i1.991.
- [5] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [6] A. Putri *et al.*, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
- [7] D. Muriyatmoko, A. Musthafa, and M. H. Wijaya, "Klasifikasi Profil Kelulusan Nilai AKPAM Dengan Metode Decision Tree," no. April, 2024.
- [8] I. Irawan, M. R. Qisthiano, M. Syahril, and P. M. Jakak, "Optimasi Prediksi Kelulusan Tepat Waktu: Studi Perbandingan Algoritma Random Forest dan Algoritma K-NN Berbasis PSO," *J. Pengemb. Sist. Inf. dan Inform.*, vol. 4, no. 4, pp. 26–35, 2023, doi: 10.47747/jpsii.v4i4.1374.
- [9] D. Septhya *et al.*, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 15–19, 2023, doi: 10.57152/malcom.v3i1.591.
- [10] J. Amarda, N. Suarna, and O. Nurdiawan, "Klasifikasi Industri Kerajinan Berdasarkan Provinsi Menggunakan Metode Decision Tree Untuk Membandingkan Pengujian Algoritma Id3 Algoritma C45 Dan Algoritma Cart," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 790–796, 2022, doi: 10.36040/jati.v6i2.5736.
- [11] A. Nurfaizin, N. Dengen, and H. S. Pakpahan, "Analisis Evaluasi Kinerja Akademik Mahasiswa Program Studi Teknik Informatika Universitas Mulawarman menggunakan Algoritma CART," *J. Rekayasa Teknol. Inf.*, vol. 4, no. 2, p. 116, 2020, doi: 10.30872/jurti.v4i2.1256.