

ALGORITME DECISION TREE UNTUK PROSES SELEKSI CALON ANGGOTA UNIT KEGIATAN MAHASISWA CREATIVE COMPUTER CLUB

Kristivera Sibagariang^{*1}, Dhaevita Amartika¹, Nurani Anisa Putri¹, Laili Cahyani¹

¹Program Studi Pendidikan Informatika, Universitas Trunodjoyo Madura, Indonesia

*e-mail: 240631100027@student.trunodjoyo.ac.id

Abstrak

Melakukan proses seleksi yang terbuka, jujur, dan adil merupakan hal yang sangat penting dalam memilih anggota baru di organisasi kemahasiswaan. Karena prosedur penerimaan di Unit Kegiatan Mahasiswa (UKM) Creative Computer Club (Triple C) sampai saat ini masih menggunakan cara yang tradisional dan mudah dipengaruhi oleh penilaian pribadi, penelitian ini menawarkan metode klasifikasi yang menggunakan data nyata sebagai alternatifnya. Penilaian terhadap hasil kelulusan calon pengurus UKM Triple C angkatan 2024 dilakukan dengan menggunakan algoritma *Decision Tree*. Dataset yang digunakan terdiri dari 146 data pendaftar, di mana komponen Nilai Tugas, Nilai Organisasi, dan Total Nilai berperan sebagai variabel prediktor, sedangkan Status Kelulusan berfungsi sebagai variabel target. Model prediktif dibuat menggunakan algoritma *Decision Tree Classifier* dengan *Gini Index* sebagai kriteria pemisahan data. Temuan penelitian menunjukkan bahwa atribut Total Nilai berperan paling penting dalam membentuk pohon keputusan dan menghasilkan aturan klasifikasi yang mudah dipahami. Berdasarkan hasil evaluasi menggunakan *confusion matrix*, model menunjukkan performa yang sangat baik dengan tingkat akurasi rata-rata mencapai 99,33% yang diperoleh melalui metode *10-Fold Cross Validation*. Hasil tersebut menunjukkan bahwa algoritma *Decision Tree* mampu mengklasifikasikan kelulusan calon anggota secara akurat. Oleh karena itu, penggunaan algoritma *Decision Tree* layak direkomendasikan sebagai sistem pendukung keputusan yang objektif dan efisien dalam proses seleksi anggota organisasi.

Kata kunci: Data Mining; *Decision Tree*; Klasifikasi; UKM.

Abstract

Conducting an open, transparent, and objective selection process is essential in recruiting new members of student organizations. Since the recruitment process at the Creative Computer Club (Triple C) Student Activity Unit (UKM) has traditionally been carried out manually and is potentially influenced by subjective judgment, this study proposes a data-driven classification approach as an alternative. The classification of prospective UKM Triple C members for the 2024 recruitment period was performed using the Decision Tree algorithm. The dataset consisted of 146 applicant records, with Task Score, Organizational Score, and Total Score serving as predictor variables, while Admission Status was used as the target variable. A predictive model was developed using the Decision Tree Classifier algorithm with the Gini Index as the splitting criterion. The results indicate that the Total Score attribute played the most influential role in constructing the decision tree and generating classification rules that are easy to interpret. Based on confusion matrix evaluation, the model achieved excellent performance with an average accuracy of 99.33% obtained through 10-Fold Cross Validation. These findings demonstrate that the Decision Tree algorithm is effective in accurately classifying applicant admission outcomes. Therefore, the Decision Tree algorithm can be recommended as an objective and efficient decision-support tool for student organization recruitment.

Keywords: Data Mining; Decision Tree; Classification; Student Organization.

1. PENDAHULUAN

Menurut Irawan dan Ningsih [1], penerapan algoritma *Decision Tree* pada sistem pendukung keputusan dapat membantu proses seleksi menjadi lebih terstruktur, objektif, dan konsisten. Di Universitas Trunodjoyo Madura, Unit Kegiatan Mahasiswa (UKM) Creative Computer Club (Triple C) menjadi wadah bagi mahasiswa untuk mengembangkan kemampuan di bidang teknologi dan organisasi. Proses rekrutmen anggota baru melibatkan penilaian kemampuan teknis (*hard skill*) dan kemampuan organisasi (*soft skill*). Namun, proses seleksi masih dilakukan secara manual sehingga membutuhkan waktu yang lebih lama, rentan terhadap kesalahan manusia, serta berpotensi menghasilkan keputusan yang subjektif. Untuk mengatasi permasalahan tersebut, diperlukan penerapan teknik data mining menggunakan algoritma *Decision Tree*. Kokash dan Makhniz [2] menjelaskan bahwa algoritma *Decision Tree* mampu menghasilkan model klasifikasi yang mudah dipahami karena aturan keputusan yang dihasilkan dapat diinterpretasikan secara langsung oleh pengguna.

Kotu dan Deshpande [3] menjelaskan bahwa data mining merupakan proses ekstraksi informasi dan pola dari kumpulan data untuk mendukung pengambilan keputusan. Sementara itu, Sarker [4] menjelaskan

bahwa algoritma *machine learning* telah banyak diterapkan pada berbagai bidang karena mampu membantu proses prediksi dan pengambilan keputusan berbasis data. Dalam konteks klasifikasi data, Sharma dan Kumar [5] menjelaskan bahwa algoritma *Decision Tree* banyak digunakan karena mampu menyajikan model dalam bentuk pohon keputusan yang transparan dan mudah diinterpretasikan.

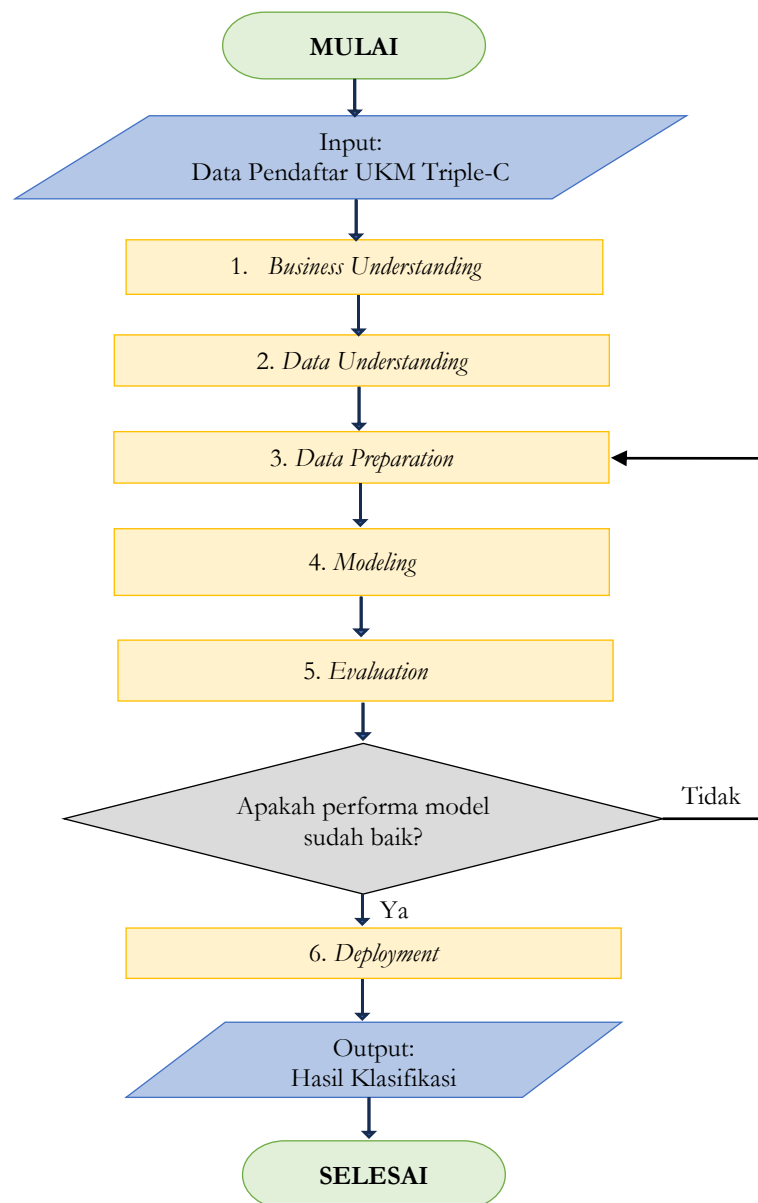
Sejumlah penelitian terdahulu membuktikan efektivitas algoritma *Decision Tree* di bidang pendidikan dan klasifikasi data. Mundok dkk. [6] menjelaskan bahwa model yang dikembangkan memperoleh tingkat akurasi sebesar 92% dalam klasifikasi kelulusan mahasiswa. Sutoyo [7] menjelaskan bahwa algoritma *Decision Tree* C4.5 menghasilkan akurasi sebesar 96,73% menggunakan metode *10-Fold Cross Validation*. Hardian dkk. [8] menunjukkan bahwa penerapan teknik seleksi fitur mampu meningkatkan nilai *recall* hingga 92,24%. Qisthiano dkk. [9] memperoleh akurasi sebesar 87,93% pada prediksi kelulusan mahasiswa berbasis data alumni. Selanjutnya, Ameliatus dan Fatah [10] menjelaskan bahwa hasil akurasi sebesar 99,00% melalui pemanfaatan variabel nilai ujian, sedangkan Fatah dan Hasanah [11] memperoleh akurasi sebesar 94,48% pada klasifikasi kelulusan berdasarkan ambang batas IPK. Untuk memperkuat analisis komparasi algoritma, Hartanti dan Pradana [12] menjelaskan batasan performa *Decision Tree* pada data numerik, sementara Ghaisani dkk. [13] mengonfirmasi keandalan metodologi CRISP-DM dalam menstrukturkan pemodelan *Decision Tree*. Selain itu, Nabarian dkk. [14] mengevaluasi efisiensi komparatif *Decision Tree* untuk prediksi kelulusan mahasiswa serta menegaskan bahwa akurasi klasifikasi yang baik menjadi faktor penting dalam mendukung pengambilan keputusan di lingkungan perguruan tinggi.

Hasil kajian literatur menunjukkan bahwa penerapan algoritma *Decision Tree* pada konteks seleksi anggota organisasi mahasiswa belum banyak dieksplorasi. Penelitian-penelitian sebelumnya umumnya memanfaatkan variabel akademik, seperti IPK, nilai perkuliahan, dan data alumni mahasiswa. Berbeda dengan penelitian tersebut, penelitian ini menggunakan data seleksi anggota organisasi kemahasiswaan yang menggabungkan aspek *hard skill* dan *soft skill* sebagai atribut dalam proses klasifikasi. Pendekatan tersebut belum banyak ditemukan pada penelitian klasifikasi berbasis *Decision Tree* di lingkungan organisasi mahasiswa. Selain itu, penelitian ini menggunakan data riil hasil seleksi anggota UKM Triple C Tahun 2024 yang belum banyak dimanfaatkan sebagai objek penelitian data mining. Dengan demikian, penelitian ini memberikan kontribusi dalam pengembangan sistem pendukung keputusan berbasis data pada lingkungan organisasi mahasiswa.

Berlandaskan kesenjangan (*gap*) tersebut, studi ini diarahkan untuk mengimplementasikan algoritma *Decision Tree* dengan panduan metodologi CRISP-DM dalam memetakan hasil seleksi calon pengurus UKM Triple C. Basis data yang dianalisis mencakup 146 entri pendaftar, dengan memposisikan atribut Nilai Tugas, Nilai Organisasi, beserta Total Nilai selaku variabel prediktor, serta Status Kelulusan sebagai variabel targetnya. Luaran dari riset ini diproyeksikan tidak sekadar memproduksi model klasifikasi berakurasi tinggi, melainkan juga menyelidik parameter penilaian yang paling dominan serta merumuskan aturan keputusan (*decision rules*) logis yang interpretable untuk dijadikan landasan sistem pendukung kebijakan pada rekrutmen internal organisasi.

2. METODE

Penelitian ini merupakan penelitian yang berkonsep data mining dengan pendekatan deskriptif dan implementatif dalam mengklasifikasikan kelulusan calon anggota UKM Creative Computer Club (Triple C). Penelitian ini menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) sebagai kerangka kerja utama karena menyediakan alur yang terstruktur dalam pelaksanaan proyek data mining, mulai dari identifikasi permasalahan hingga evaluasi hasil pemodelan. Chapman dkk. [15] menjelaskan bahwa CRISP-DM terdiri atas enam tahapan, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Dalam penelitian ini, algoritma *Decision Tree* dipilih sebagai metode klasifikasi karena mampu menghasilkan model dalam bentuk pohon keputusan yang mudah dipahami dan diinterpretasikan. Karakteristik tersebut memungkinkan proses pengambilan keputusan dapat dijelaskan secara lebih transparan melalui aturan-aturan klasifikasi yang dihasilkan. Oleh karena itu, kombinasi CRISP-DM dan *Decision Tree* dinilai sesuai untuk mendukung proses seleksi anggota UKM yang memerlukan pendekatan yang lebih sistematis, objektif, dan berbasis data.



Gambar 1. *Flowchart* Alur Pelaksanaan Penelitian Klasifikasi Kelulusan Anggota UKM.

Dataset yang digunakan merupakan data historis hasil proses seleksi calon anggota baru UKM Triple C Tahun 2024 yang diperoleh dari panitia pelaksana seleksi. Dataset terdiri dari 146 data pendaftar dengan atribut Nilai Tugas, Nilai Organisasi, dan Total Nilai sebagai variabel prediktor, serta Status Kelulusan sebagai variabel target. Tahap awal penelitian mencakup verifikasi kualitas dataset melalui pemeriksaan terhadap kelengkapan dan konsistensi data agar layak digunakan dalam analisis. Selanjutnya, proses pemodelan klasifikasi diimplementasikan menggunakan *Python* dengan dukungan pustaka *Scikit-Learn*. Algoritma *Decision Tree Classifier* digunakan untuk membangun model klasifikasi berdasarkan atribut yang telah dipilih. Tahapan penelitian dimulai dari pengumpulan data, pemahaman data, pengolahan data, pemodelan menggunakan algoritma *Decision Tree*, hingga evaluasi model menggunakan *confusion matrix*. Alur penelitian pada studi ini digambarkan melalui diagram alir yang disajikan pada Gambar 1.

Diagram tersebut memperlihatkan enam tahapan utama dalam kerangka CRISP-DM yang digunakan sebagai panduan pelaksanaan.

1. Tahap *business understanding* dilakukan analisis terhadap permasalahan seleksi anggota yang masih bersifat manual dan berpotensi subjektif
2. Tahap *data understanding* mencakup eksplorasi karakteristik dataset untuk mengidentifikasi pola awal serta kualitas data.

3. Tahap *data preparation* berfokus pada penyiapan dataset sebelum proses pemodelan dilakukan. Kegiatan pada tahap ini mencakup pengecekan kelengkapan data, verifikasi keseragaman format, serta pemilihan atribut yang memiliki keterkaitan dengan tujuan klasifikasi. Hasil proses tersebut menunjukkan bahwa Nilai Tugas, Nilai Organisasi, dan Total Nilai digunakan sebagai variabel prediktor, sedangkan Status Kelulusan ditetapkan sebagai variabel target dalam pembentukan model. Total Nilai diperoleh dari penjumlahan Nilai Tugas dan Nilai Organisasi dan digunakan sebagai salah satu variabel prediktor dalam proses klasifikasi. Total Nilai digunakan karena merupakan nilai akhir yang digunakan panitia seleksi sebagai dasar penentuan kelulusan calon anggota.
4. Pada fase *modeling*, dilakukan proses pembentukan model prediksi menggunakan algoritma *Decision Tree*. Dataset yang telah dipersiapkan dibagi ke dalam dua kelompok, yaitu 75% sebagai data pelatihan dan 25% sebagai data pengujian. Data pelatihan digunakan sebagai dasar pembentukan aturan klasifikasi yang dipelajari oleh model, sedangkan data pengujian berfungsi untuk mengukur tingkat keberhasilan model dalam memprediksi kelas pada data yang tidak termasuk dalam proses pelatihan. Pemilihan atribut pada setiap percabangan pohon keputusan mengacu pada nilai *Gini Index*, di mana atribut dengan tingkat kemurnian kelas yang lebih baik diprioritaskan sebagai node pemisah. Breiman dkk. [16] menjelaskan bahwa atribut dengan nilai Gini terendah dipilih sebagai node pemisah karena mampu menghasilkan tingkat homogenitas kelas yang lebih tinggi.

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

Pada persamaan (1), $Gini(D)$ menunjukkan nilai impurity pada dataset D , sedangkan p_i merupakan proporsi data pada kelas ke- i .

5. Pada tahap *evaluation*, performa model dianalisis menggunakan *confusion matrix*. Berdasarkan matriks tersebut dihitung beberapa indikator evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Menurut Kotu dan Deshpande [3], *accuracy* digunakan untuk mengukur tingkat ketepatan prediksi secara keseluruhan, *precision* menunjukkan proporsi prediksi yang benar pada suatu kelas, *recall* mengukur kemampuan model dalam menemukan seluruh data yang relevan, sedangkan *F1-score* merupakan ukuran gabungan antara *precision* dan *recall*. Rumus perhitungan metrik evaluasi ditunjukkan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

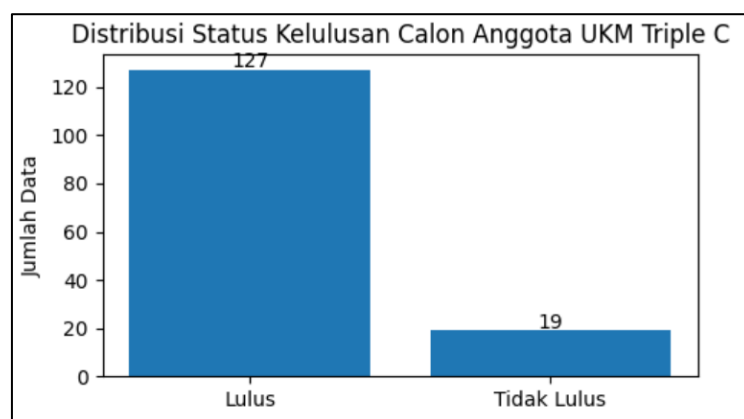
Pada *confusion matrix*, TP (*True Positive*) menunjukkan jumlah data yang diprediksi positif dan sesuai dengan kondisi aktual, TN (*True Negative*) menunjukkan jumlah data yang diprediksi negatif dan sesuai dengan kondisi aktual, FP (*False Positive*) menunjukkan jumlah data yang diprediksi positif tetapi sebenarnya negatif, sedangkan FN (*False Negative*) menunjukkan jumlah data yang diprediksi negatif tetapi sebenarnya positif.

6. Tahap *deployment* merupakan fase pemanfaatan hasil pemodelan ke dalam bentuk aturan keputusan yang dapat digunakan sebagai acuan dalam proses seleksi anggota UKM. Pada tahap ini, hasil klasifikasi yang diperoleh dari algoritma *Decision Tree* diinterpretasikan sehingga dapat mendukung proses pengambilan keputusan secara lebih terstruktur. Melalui penerapan pendekatan tersebut, penelitian tidak hanya menghasilkan model klasifikasi yang memiliki kinerja baik, tetapi juga mendukung peningkatan konsistensi, transparansi, dan objektivitas dalam proses seleksi anggota organisasi mahasiswa.

3. HASIL PENELITIAN

A. Karakteristik Dataset

Data yang digunakan pada penelitian ini berasal dari hasil proses rekrutmen calon anggota UKM Creative Computer Club (Triple C) Tahun 2024. Dataset memuat atribut Nilai Tugas yang merepresentasikan aspek kemampuan teknis, Nilai Organisasi yang menggambarkan kemampuan organisasi, serta Status Kelulusan yang digunakan sebagai kelas target. Sebelum tahap pembentukan model dilakukan, terlebih dahulu dilakukan analisis distribusi data untuk memperoleh gambaran karakteristik masing-masing kelas dan memahami kondisi dataset yang akan digunakan dalam proses klasifikasi. Dataset penelitian terdiri atas 146 data calon anggota UKM Triple C, yang terdiri dari 127 data (86,99%) berstatus Lulus dan 19 data (13,01%) berstatus Tidak Lulus. Distribusi tersebut menunjukkan adanya ketidakseimbangan kelas (*class imbalance*), di mana jumlah data pada kelas Lulus jauh lebih besar dibandingkan kelas Tidak Lulus. Meskipun demikian, distribusi tersebut mencerminkan kondisi nyata proses seleksi anggota UKM Triple C Tahun 2024 sehingga tetap digunakan dalam pembentukan model klasifikasi.

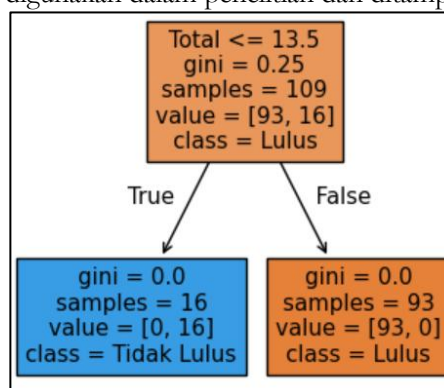


Gambar 2. Distribusi Status Kelulusan Calon Anggota UKM Triple C.

Distribusi status kelulusan juga divisualisasikan pada Gambar 2 untuk memberikan gambaran proporsi masing-masing kelas. Grafik menunjukkan bahwa kelas Lulus mendominasi dataset dengan proporsi sebesar 86,99%, sedangkan kelas Tidak Lulus sebesar 13,01%. Hasil ini menunjukkan bahwa dataset memiliki distribusi kelas yang tidak seimbang (*imbalanced dataset*), di mana jumlah data pada kelas Lulus jauh lebih besar dibandingkan kelas Tidak Lulus. Meskipun demikian, distribusi tersebut merepresentasikan kondisi seleksi anggota UKM Triple C Tahun 2024 yang sebenarnya sehingga tetap digunakan dalam proses pembangunan model klasifikasi.

B. Hasil Pemodelan *Decision Tree*

Tahap pemodelan menghasilkan sebuah pohon keputusan yang dibangun dari 109 data pelatihan menggunakan algoritma *Decision Tree*. Struktur pohon yang terbentuk menunjukkan proses pemisahan kelas berdasarkan atribut yang digunakan dalam penelitian dan ditampilkan pada Gambar 3.



Gambar 3. Pohon Keputusan Hasil Klasifikasi Kelulusan Calon Anggota UKM Triple C.

Analisis terhadap pohon keputusan pada Gambar 3 menunjukkan bahwa atribut Total Nilai memiliki kontribusi terbesar dalam proses pemisahan kelas sehingga ditempatkan sebagai *root node*. Nilai *threshold* sebesar 13,5 digunakan sebagai titik pemisah awal untuk membedakan data ke dalam kategori Lulus dan Tidak Lulus. Dengan demikian, atribut tersebut menjadi faktor utama dalam mekanisme klasifikasi yang dibentuk oleh model. Pada node akar terdapat 109 data pelatihan yang terdiri atas 93 data berstatus Lulus dan 16 data berstatus Tidak Lulus dengan nilai Gini sebesar 0,25.

Cabang sebelah kiri menunjukkan kondisi $Total \leq 13,5$ yang menghasilkan 16 data dengan status Tidak Lulus dan nilai Gini sebesar 0,00. Nilai Gini sebesar 0,00 menunjukkan bahwa seluruh data pada node tersebut berada dalam satu kelas yang sama sehingga tidak terjadi kesalahan klasifikasi. Sementara itu, cabang sebelah kanan menunjukkan kondisi $Total > 13,5$ yang menghasilkan 93 data dengan status Lulus dan nilai Gini sebesar 0,00.

Hasil tersebut menunjukkan bahwa atribut Total terpilih sebagai *root node* sehingga menjadi atribut yang memiliki kontribusi terbesar dalam proses pemisahan kelas berdasarkan nilai *Gini Index*. Aturan keputusan yang dihasilkan model dapat diinterpretasikan secara sederhana, yaitu calon anggota dengan nilai total lebih dari 13,5 diklasifikasikan sebagai Lulus, sedangkan calon anggota dengan nilai total kurang dari atau sama dengan 13,5 diklasifikasikan sebagai Tidak Lulus. Kemampuan interpretasi yang tinggi ini menjadi salah satu keunggulan algoritma *Decision Tree* dalam mendukung proses pengambilan keputusan yang transparan dan mudah dipahami.

C. Evaluasi Model

Kinerja model yang telah dibangun selanjutnya dianalisis melalui tahap evaluasi. Pengukuran performa dilakukan menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai kualitas hasil klasifikasi. Untuk memperoleh gambaran yang lebih representatif mengenai kemampuan model, dilakukan pula validasi menggunakan metode *10-Fold Cross Validation* sehingga tingkat stabilitas model dapat diamati pada beberapa pembagian dataset yang berbeda. Berdasarkan hasil pengujian, sebanyak 34 data berstatus Lulus dan 3 data berstatus Tidak Lulus berhasil diklasifikasikan dengan benar. Tidak ditemukan kesalahan klasifikasi baik pada kategori *False Positive* maupun *False Negative* sehingga model menghasilkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* sebesar 100%.

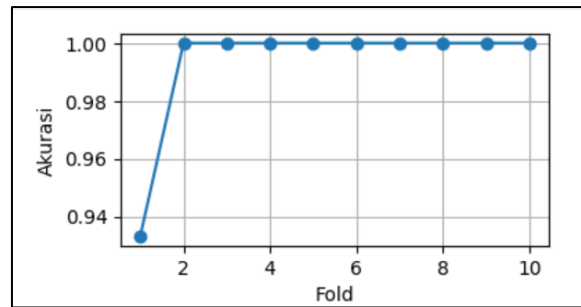
Meskipun demikian, hasil evaluasi tersebut perlu ditafsirkan secara cermat. Tingginya performa model kemungkinan dipengaruhi oleh karakteristik dataset yang memiliki pola pemisahan kelas yang relatif jelas serta jumlah data yang masih terbatas. Selain itu, penggunaan atribut Total Nilai yang memiliki keterkaitan langsung dengan proses penentuan kelulusan berpotensi memberikan pengaruh dominan dalam pembentukan model. Oleh karena itu, diperlukan pengujian pada dataset yang lebih besar dan beragam untuk memperoleh gambaran yang lebih komprehensif mengenai kemampuan generalisasi model.

Tabel 1. Hasil *10-Fold Cross Validation*.

<i>Fold</i>	<i>Accuracy</i>
1	93,33%
2	100%
3	100%
4	100%
5	100%
6	100%
7	100%
8	100%
9	100%
10	100%
Rata-rata	99,33%

Metode *10-Fold Cross Validation* diterapkan untuk mengevaluasi konsistensi kinerja model pada beberapa pembagian dataset. Hasil pengujian memperlihatkan bahwa akurasi model berada pada kisaran 93,33% hingga 100% dengan rata-rata sebesar 99,33%. Nilai tersebut menunjukkan bahwa model memiliki tingkat stabilitas yang baik, meskipun terdapat variasi performa pada beberapa *fold* yang mengindikasikan adanya pengaruh distribusi data terhadap hasil klasifikasi. Meskipun demikian, adanya satu *fold* dengan akurasi 93,33% mengindikasikan bahwa performa model masih dipengaruhi oleh variasi data yang digunakan pada proses pelatihan dan pengujian. Oleh karena itu, hasil *cross validation*

memberikan gambaran yang lebih representatif dibandingkan pengujian menggunakan satu kali pembagian data (*train-test split*).



Gambar 4. Grafik Akurasi 10-Fold Cross Validation.

Berdasarkan Gambar 4, nilai akurasi pada setiap *fold* menunjukkan tingkat konsistensi yang tinggi. Sebagian besar *fold* menghasilkan akurasi sebesar 100%, sedangkan hanya satu *fold* yang memperoleh akurasi sebesar 93,33%. Hasil tersebut menunjukkan bahwa model *Decision Tree* mampu mempertahankan kinerja klasifikasi yang konsisten pada berbagai skenario pembagian data. Rentang akurasi yang relatif sempit mengindikasikan bahwa perubahan komposisi data pelatihan dan data pengujian tidak memberikan pengaruh yang signifikan terhadap performa model. Temuan ini memperkuat hasil evaluasi sebelumnya bahwa model memiliki kemampuan klasifikasi yang baik pada dataset seleksi anggota UKM Triple C.

4. DISKUSI

Hasil penelitian menunjukkan bahwa model klasifikasi menggunakan algoritma *Decision Tree* berbasis CRISP-DM mampu menghasilkan performa yang baik dalam mengklasifikasikan data kelulusan calon anggota UKM Triple C. Berdasarkan hasil evaluasi, diperoleh nilai *accuracy*, *precision*, *recall*, dan *F1-score* masing-masing sebesar 100% pada data uji. Selain itu, hasil pengujian menggunakan metode *10-Fold Cross Validation* menunjukkan rata-rata akurasi sebesar 99,33%, yang mengindikasikan bahwa model memiliki tingkat stabilitas yang baik dalam melakukan klasifikasi. Tingginya nilai akurasi yang diperoleh sejalan dengan penelitian sebelumnya. Kokash dan Makhniet [2] menyatakan bahwa algoritma *Decision Tree* memiliki performa yang baik dalam proses klasifikasi, sedangkan Hardian dkk. [8] menunjukkan bahwa algoritma tersebut mampu menghasilkan tingkat *recall* yang tinggi pada proses klasifikasi data. Qisthiano dkk. [9] juga membuktikan bahwa *Decision Tree* efektif digunakan untuk prediksi kelulusan mahasiswa. Hartanti dan Pradana [12] menjelaskan karakteristik performa *Decision Tree* pada data numerik, sementara Ghaisani dkk. [13] mengonfirmasi bahwa pendekatan CRISP-DM dapat mendukung proses pemodelan *Decision Tree* secara sistematis.

Selain itu, penerapan metode CRISP-DM dalam penelitian ini memberikan kontribusi terhadap kualitas model yang dihasilkan. Chapman dkk. [15] menjelaskan bahwa CRISP-DM menyediakan tahapan sistematis dalam proses data mining, mulai dari pemahaman bisnis hingga evaluasi model. Temuan tersebut juga didukung oleh Sarker [4] yang menegaskan pentingnya pendekatan sistematis dalam pengembangan model *machine learning*. Namun demikian, hasil evaluasi yang tinggi perlu dianalisis secara kritis. Meskipun model menunjukkan hasil evaluasi yang baik, ukuran dataset yang relatif terbatas menjadi salah satu keterbatasan penelitian sehingga diperlukan pengujian pada dataset yang lebih besar untuk menguji konsistensi dan kemampuan generalisasi model. Selain itu, performa model kemungkinan dipengaruhi oleh penggunaan atribut Total Nilai yang merupakan hasil agregasi dari Nilai Tugas dan Nilai Organisasi sehingga memiliki keterkaitan langsung dengan penentuan status kelulusan. Kondisi tersebut menyebabkan proses pemisahan kelas menjadi lebih jelas sehingga berpotensi meningkatkan nilai akurasi model. Hal ini sejalan dengan teori yang dikemukakan oleh Kotu dan Deshpande [3] serta Sarker [4], yang menyatakan bahwa model yang dibangun menggunakan dataset berukuran terbatas berpotensi memiliki kemampuan generalisasi yang lebih rendah terhadap data baru.

Tingginya nilai *accuracy*, *precision*, *recall*, dan *F1-score* sebesar 100% tidak selalu menunjukkan bahwa model memiliki kemampuan generalisasi yang sempurna. Hasil tersebut dapat dipengaruhi oleh karakteristik dataset yang relatif kecil serta adanya atribut Total Nilai yang memiliki keterkaitan langsung dengan status kelulusan. Oleh karena itu, meskipun model menunjukkan performa yang baik pada dataset penelitian, diperlukan

pengujian lebih lanjut menggunakan data yang lebih besar dan beragam untuk memastikan kestabilan model terhadap data baru.

Meskipun dataset memiliki distribusi kelas yang tidak seimbang, model tetap menunjukkan kinerja yang baik berdasarkan seluruh metrik evaluasi yang digunakan. Namun demikian, penelitian selanjutnya dapat mempertimbangkan teknik penyeimbangan data seperti SMOTE untuk menguji kestabilan model pada distribusi kelas yang berbeda. Menurut Kotu dan Deshpande [3] performa model klasifikasi dipengaruhi oleh karakteristik data dan atribut yang digunakan dalam proses pembelajaran model. Perbedaan hasil penelitian ini kemungkinan disebabkan oleh karakteristik dataset yang memiliki pola pemisahan kelas yang lebih jelas serta penggunaan atribut yang relevan dalam proses klasifikasi.

Penelitian ini juga memiliki keunggulan dalam penerapan metode data mining pada konteks seleksi anggota organisasi mahasiswa berbasis teknologi. Nabarian dkk. [14] lebih berfokus pada prediksi kelulusan mahasiswa di lingkungan pendidikan formal, sedangkan penelitian ini mengaplikasikan pendekatan tersebut pada proses seleksi anggota organisasi kemahasiswaan. Temuan ini menunjukkan bahwa algoritma *Decision Tree* dapat diterapkan pada berbagai konteks pengambilan keputusan berbasis data. Dengan demikian, model yang dihasilkan berpotensi dimanfaatkan sebagai sistem pendukung keputusan dalam proses seleksi anggota UKM yang lebih objektif, transparan, dan berbasis data. Namun, untuk meningkatkan validitas dan generalisasi model, disarankan agar penelitian selanjutnya menggunakan dataset yang lebih besar, menambahkan atribut yang lebih variatif, serta membandingkan dengan algoritma lain seperti *Naïve Bayes* atau *Random Forest*.

5. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *Decision Tree* dengan pendekatan CRISP-DM untuk mendukung proses klasifikasi kelulusan calon anggota UKM Creative Computer Club (Triple C). Berdasarkan proses pemodelan yang dilakukan, sistem mampu membentuk aturan keputusan yang jelas sehingga dapat membantu proses seleksi secara lebih terstruktur dan berbasis data. Variabel Total Nilai teridentifikasi sebagai atribut yang paling berpengaruh dalam pembentukan pohon keputusan, yang menunjukkan bahwa hasil akhir seleksi dipengaruhi oleh kombinasi penilaian kemampuan teknis dan organisasi. Hasil evaluasi memperlihatkan bahwa model memiliki tingkat kinerja yang tinggi pada data penelitian, sehingga berpotensi dimanfaatkan sebagai alternatif pendukung pengambilan keputusan dalam proses rekrutmen anggota UKM. Penerapan pendekatan ini memberikan manfaat berupa peningkatan objektivitas, efisiensi, dan konsistensi dibandingkan proses seleksi yang dilakukan secara manual. Meskipun demikian, penelitian ini masih menggunakan jumlah data yang terbatas dan berasal dari satu periode rekrutmen. Oleh karena itu, penelitian berikutnya dapat memperluas cakupan data, menambahkan variabel penilaian lain yang relevan, serta melakukan perbandingan dengan metode klasifikasi berbeda untuk memperoleh gambaran performa yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] A. Irawan and D. Handayani Untari Ningsih, "Implementasi Algoritma Decision Tree dalam Sistem Pendukung Keputusan Seleksi Calon Karyawan Pada Pt G-Textile Indonesia Menggunakan Framework Django Python," *Jurnal Mahasiswa Teknik Informatika*, vol. 10, no. 1, pp. 1302–1308, Feb. 2026.
- [2] N. Kokash and L. Makhniet, "Using Decision Trees for Interpretable Supervised Clustering," *SN Comput. Sci.*, vol. 5, no. 2, Feb. 2024, doi: 10.1007/s42979-023-02590-7.
- [3] V. Kotu and B. Deshpande, *Data Science: Concepts and Practice*, 2nd ed. Cambridge, MA, USA: Morgan Kaufmann, 2019.
- [4] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021.
- [5] S. Sharma and A. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," *International Journal of Scientific & Technology Research*, vol. 9, no. 4, pp. 1234–1239, 2020.
- [6] G. A. Mundok, Amiruddin, and Z. Y. Lamasigi, "Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Metode Decision Tree," *Jurnal Ilmiah Ilmu Komputer*, vol. 3, no. 1, pp. 31–36, 2024.
- [7] I. Sutoyo, "Implementasi Algoritma Decision Tree untuk Klasifikasi Data Peserta Didik," *PILAR Nusa Mandiri*, vol. 14, no. 2, pp. 217–224, Sep. 2018, [Online]. Available: www.bsi.ac.id
- [8] R. Wahyu Hardian, Jasmir, and Sharipuddin, "Peningkatan Algoritma Decision Tree dalam Mengklasifikasi Tingkat Kelulusan Mahasiswa Universitas Jambi Dengan Seleksi Fitur Chi-Square,"

- Jurnal Manajemen Teknologi dan Sistem Informasi (JMS)*, vol. 5, no. 2, pp. 1105–1114, Sep. 2025, doi: 10.33998/jms.v5i2.
- [9] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, “Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa,” *G-Tech: Jurnal Teknologi Terapan*, vol. 7, no. 1, pp. 21–28, Jan. 2023, doi: 10.33379/gtech.v7i1.1850.
- [10] Q. Ameliatus and Z. Fatah, “Klasifikasi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Menggunakan Aplikasi Rapidminer,” *Jurnal Riset Teknik Komputer*, vol. 1, pp. 58–64, Dec. 2024, doi: 10.69714/em8qpw54.
- [11] Z. Fatah and M. Hasanah, “Penerapan Decision Tree pada Klasifikasi Kelulusan Mahasiswa,” *JAMASTIKA*, vol. 4, Oct. 2025.
- [12] D. Hartanti and A. I. Pradana, “Komparasi Metode Decission Tree, Logistic Regression, SVM, dan ANN Dalam Klasifikasi Kualitas Air,” *SMARTICS Journal*, vol. 9, no. 1, pp. 1–6, Apr. 2023, doi: 10.21067/smartics.v9i1.8113.
- [13] B. A. Ghaisani, G. A. Damayanti, R. M. A. Rani, N. Fitri, and W. S. Dharmawan, “Pemodelan Klasifikasi Kategori Harga Rumah Menggunakan Algoritma Decision Tree dengan Pendekatan CRISP-DM,” *Jurnal Informatika UPGRIS*, vol. 11, no. 2, pp. 23–29, 2025.
- [14] T. Nabarian, I. Wahyuni, and S. El Farisi, “Penerapan Kerangka CRISP-DM dalam Evaluasi Performa Logistic Regression dan Decision Tree untuk Prediksi Kelulusan Mahasiswa,” *Journal Computer Science*, vol. 5, no. 1, pp. 46–55, Apr. 2026.
- [15] P. Chapman *et al.*, “CRISP-DM 1.0: Step-by-Step Data Mining Guide,” 2000.
- [16] Leo. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. New York: Chapman & Hall, 1984.