

Penerapan Metode *Decision Tree* dalam Penentuan Jurusan Siswa

Yuli Nurul Aini¹, Ahmad Faqih², Gifthera Dwilestari³

^{1,2,3} STMIK IKMI Cirebon, Jl. Perjuangan No 10B Majasem Kec. Kesambi Kota. Cirebon 45135, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 09 Desember 2024

Revisi Akhir: 13 Februari 2025

Diterbitkan Online: 15 Maret 2025

KATA KUNCI

Jurusan

Klasifikasi

Decision Tree C4.5

KORESPONDENSI

E-mail: ggdwilestari@gmail.com

ABSTRACT

Choosing a major for vocational students is an important step in determining their future career. Inappropriate decisions in choosing a major can have a negative impact on the learning process and career prospects of students. This research aims to apply the Decision Tree C4.5 algorithm in determining student majors based on historical data from SMK Plus Qurrota A'yun. The main attributes used in this model include gender, practical score, report card score, and rank. The data analysis process is carried out using KDD (Knowledge Discovery in Databases) stages which include preprocessing (normalization and labeling), data transformation with SMOTE to balance the data, and dataset division (80% for training and 20% for testing). The results of the C4.5 Decision Tree model show an accuracy rate of 86.59%, with report card and practical grades as the most significant attributes in determining student majors. Evaluation of the model showed variation in performance between majors, with high precision in TSM (92.22%) and weak recall in TKJ (73.91%). Overall, the algorithm proved to be effective for providing major recommendations based on historical data, although optimization is needed for majors with lower precision and recall values.

1. PENDAHULUAN

Pemilihan jurusan akademik di tingkat Sekolah Menengah Kejuruan (SMK) merupakan komponen penting dalam mendorong pengembangan potensi dan bakat siswa. Proses penjurusan ini bertujuan agar siswa dapat fokus dalam mengembangkan kemampuan diri dan minat yang dimilikinya. Namun, pemilihan jurusan yang tidak tepat dapat berdampak buruk pada proses pembelajaran dan masa depan siswa, mengingat adanya ketidakseimbangan antara minat siswa dan kemampuan yang dimilikinya. Hal ini sering kali menghambat proses belajar dan dapat menurunkan motivasi akademik siswa[1]. Oleh karena itu, penting untuk memastikan bahwa penentuan jurusan akademik dilakukan dengan tepat untuk mendukung perkembangan siswa secara maksimal.

Dalam beberapa tahun terakhir, pemanfaatan metodologi penambangan data dalam bidang pendidikan telah menjadi fokus penelitian banyak ilmuwan. Penambangan data memungkinkan pengolahan dan analisis kumpulan data yang besar, sehingga menghasilkan informasi yang lebih akurat dan relevan untuk membantu pengambilan keputusan. Salah satu algoritma yang sering digunakan dalam penambangan data adalah Pohon Keputusan (*Decision Tree*), terutama algoritma C4.5, yang mampu mengidentifikasi pola dan tren dalam data berdasarkan

atribut yang ada. Algoritma ini telah terbukti efektif dalam menganalisis data dan menghasilkan keputusan yang mudah dipahami. Dalam penelitian sebelumnya, algoritma C4.5 digunakan untuk meningkatkan akurasi prediksi dalam berbagai bidang, termasuk penentuan jurusan siswa [2]. Dengan penerapan metode ini, diharapkan sekolah dapat memberikan rekomendasi jurusan yang lebih tepat bagi siswa mereka.

Penelitian-penelitian sebelumnya juga menunjukkan bahwa variabel seperti kinerja akademik siswa, minat, dan bakat memainkan peran penting dalam kesuksesan siswa di jurusan yang dipilih. Meskipun berbagai inisiatif telah dilakukan untuk mengintegrasikan teknologi dalam pendidikan di Indonesia, penggunaan teknik penambangan data di sekolah menengah kejuruan (SMK) masih relatif terbatas. Oleh karena itu, penelitian ini bertujuan untuk menerapkan model berbasis penambangan data menggunakan algoritma C4.5 untuk menentukan jurusan siswa, yang nantinya akan dievaluasi dalam konteks SMK Plus Qurrota A'yun. Dengan menggunakan data historis yang tersedia, penelitian ini diharapkan dapat memberikan kontribusi penting bagi pengembangan sistem penentuan jurusan yang lebih efektif.

2. TINJAUAN PUSTAKA

2.1 Penjurusan Siswa

Menteri Pendidikan Indonesia telah menetapkan wajib belajar 12 tahun namun hanya pada saat Sekolah Menengah Atas (SMA) saja peserta didik baru bisa menentukan jurusan yang diminatinya[3]. Tahap pemilihan jurusan bagi calon siswa SMK adalah langkah awal dari penentuan karier selanjutnya. Siswa lebih condong memilih jurusan dipengaruhi oleh teman dan juga banyak dari pilihan orang tua. Pemilihan Jurusan yang tepat tentunya dapat meningkatkan prestasi serta memberi rasa nyaman kepada siswa dalam belajar sehingga dalam kegiatan belajar dapat berjalan dengan lancar dan meminimalisasi kesulitan demi mewujudkan peningkatan prestasi belajar peserta didik [4].

2.2 Data Mining

Data mining merupakan aktivitas yang melibatkan pengumpulan data dan penggunaan data historis untuk menemukan keteraturan, pola atau hubungan dalam data besar kemudian mengekstrak data tersebut menjadi informasi-informasi yang nantinya dapat digunakan. Tujuan utama data mining adalah digunakan untuk mencari pengetahuan yang ada dalam basis data yang berukuran besar[5]. Sebagai suatu rangkaian proses, data Mining dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge discovery in database*.

2.3 Klasifikasi

Klasifikasi merupakan salah satu dari metode data mining dimana klasifikasi ini merupakan jenis analisis data yang membantu dalam melakukan prediksi sesuai label kelas sample yang diklasifikasikan. Klasifikasi merupakan proses analisa data yang menghasilkan model-model yang nantinya digambarkan melalui kelas-kelas yang ada didalam data, metode ini memiliki persyaratan yakni atribut datanya harus numerik atau nominal dan label data nominal[6]. Dalam hal ini, *Machine Learning* dapat digunakan untuk mengembangkan model klasifikasi yang dapat memprediksi tingkat akurasi.

2.4 Algoritma *Decision Tree C4.5*

Decision tree adalah teknik klasifikasi yang diakui secara luas dan sangat efektif. Fakta-fakta besar direpresentasikan sebagai aturan dengan pendekatan *Decision Tree*, yang membuat aturan tersebut mudah dimengerti oleh orang-orang. Adapun dalam prosedurnya pemilihan atribut sebagai cabang pada algoritma C4.5 menggunakan hasil perhitungan dari *gain ratio*[7].

Pohon keputusan sendiri memiliki 3 jenis bagian node, yaitu sebagai berikut :

1. *Root Node*
Root node merupakan node teratas. Pada node ini, tidak ada input atau simpul akar tidak memiliki input cabang yang masuk dan memiliki cabang lebih dari satu.
2. *Node Internal*
Node Internal node atau simpul internal hanya memiliki 1 cabang yang masuk dan dapat memiliki lebih dari 1 cabang yang keluar.
3. *Leaf Node*
Leaf node atau simpul daun merupakan simpul akhir yang hanya memiliki 1 cabang yang masuk, tidak memiliki

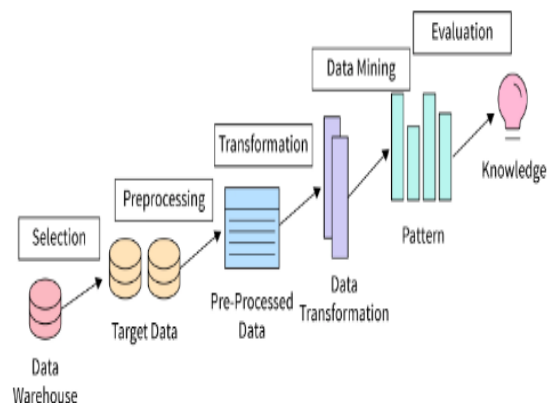
cabang sama sekali, dan menandakan bahwa simpul tersebut merupakan label kelas.

2.5 Penelitian Terdahulu

Dalam penelitian sebelumnya [8], mengaplikasikan algoritma C4.5 untuk klasifikasi jurusan di MAN 1 Inhil dengan menggunakan data nilai akademik dan minat siswa, yang menghasilkan akurasi 87,22%. Penelitian ini menyoroti pentingnya pemilihan atribut yang relevan, seperti nilai akademik dan minat siswa, dalam meningkatkan performa model klasifikasi. Dalam penelitian lainya untuk klasifikasi atau prediksi, termasuk dalam konteks pendidikan. Pada penelitian yang dilakukan oleh[9], yang menggunakan *Decision Tree C4.5* untuk seleksi penjurusan di SMK Plus Al-Hilal Arjawinangun dengan akurasi 98.02%.

3. METODOLOGI

Knowledge Discover in Database (KDD) adalah teknik data mining yang digunakan untuk mengekstrak informasi atau pola dari data yang telah dipilih sebelumnya.



Gambar 1. Tahapan *Knowledge Discovery In Database*
Proses KDD yang diterapkan dalam penelitian ini meliputi lima tahapan sebagai berikut:

- a. *Selection*
Pemilihan data merupakan tahap awal yang bertujuan untuk menentukan data yang relevan dari dataset besar yang diperoleh dari SMK Plus Qurrota A'yun. Dalam penelitian ini, data yang diambil yaitu nilai rapor, nilai praktik, peringkat, jurusan, nama siswa dan jenis kelamin.
- b. *Preprocessing*
Pada tahap ini, dilakukan proses pembersihan dan normalisasi data untuk mengurangi kemungkinan kesalahan yang dapat mempengaruhi akurasi model. Praproses data meliputi penanganan data hilang (missing values), penghilangan outliers, serta normalisasi data yang bertujuan untuk meningkatkan kualitas data.
- c. *Transformation*
Setelah data di praproses, dilakukan transformasi data untuk menyederhanakan variabel agar lebih mudah dipahami oleh algoritma Decision Tree. Proses ini melibatkan konversi nilai rapor menjadi kategori

tertentu, seperti “rendah,” “sedang,” dan “tinggi,” yang akan mempermudah analisis dan pembuatan aturan keputusan.

d. Data Mining

Tahapan ini merupakan proses utama dalam data mining untuk membangun model klasifikasi menggunakan algoritma *Decision Tree C4.5*. Algoritma ini dipilih karena kemampuannya dalam menghasilkan aturan keputusan yang jelas dan interpretatif. Metode C4.5 bekerja dengan membangun pohon keputusan dari dataset terstruktur, di mana setiap cabang dalam pohon mewakili atribut yang digunakan untuk memprediksi keputusan atau kelas. C4.5 bekerja dengan memilih atribut yang memiliki information gain tertinggi pada setiap tahap, sehingga model yang dihasilkan memiliki akurasi yang optimal dalam mengklasifikasikan data siswa.

e. Evaluasi

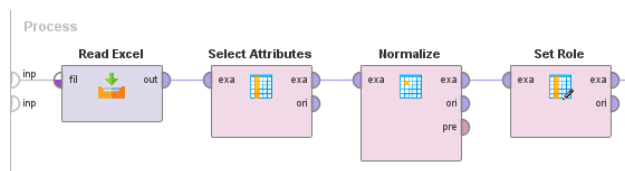
Tahap akhir adalah interpretasi dari hasil model yang dibangun. Pada tahap ini, dilakukan evaluasi terhadap performa model dengan menggunakan *confusion matrix*, *accuracy*, *precision*, dan *recall* untuk mengukur tingkat akurasi model yang dibangun. Penilaian ini penting untuk menilai apakah model yang dihasilkan mampu memprediksi jurusan siswa secara tepat.

4. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data siswa Smk Plus Qurrota A’yun. Data yang terkumpul sebanyak 375 siswa kelas XI dari enam jurusan yang ada di Smk Plus Qurrota A’yun, data ini berisi nama, jenis kelamin, nilai semua mata pelajaran, nilai praktik, nilai raport dan peringkat. Dari data tersebut yang akan digunakan sebagai atribut yaitu jenis kelamin, nilai praktik, nilai raport dan peringkat Sedangkan jurusan sebagai label. Gambaran lengkap bisa dilihat pada tabel berikut ini.

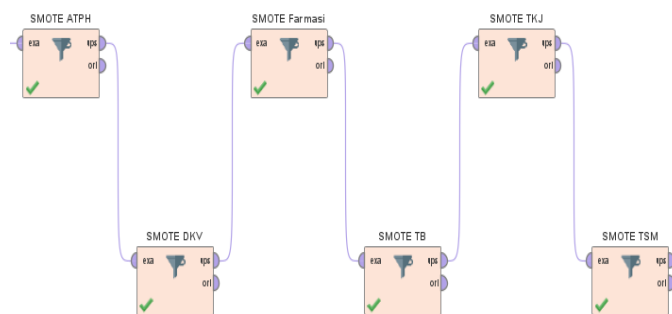
Nama	Jenis kelamin	Jurusan	Nilai Praktik	Nilai Raport	Peringkat
Pathan Maulana	Laki-Laki	ATPH	81	78,3	16
Hilman Syah Maulana	Laki-Laki	ATPH	85	82,1	14
Adila Suci Fitria	Perempuan	DKV	90	83,3	12
Adimas Geraldi Subagdja	Laki-Laki	DKV	91	81,3	22
Alfa Alghozali	Laki-Laki	Farmasi	85	82	19
Andini Siti Haildha	Perempuan	Farmasi	86	87	6
Ika Purnamasari	Perempuan	TB	90	85,7	13
Juli Khoerul Padilah	Perempuan	TB	86	84,8	24
Citra Alia	Perempuan	TKJ	92	86,1	3
Desti Khalifa	Perempuan	TKJ	90	86,8	1
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
Hasbi Ubaydillah	Laki-Laki	TSM	87	83,6	3
Hildan Ali Lukman	Laki-Laki	TSM	86	83,1	4

Gambar 1. DataSet



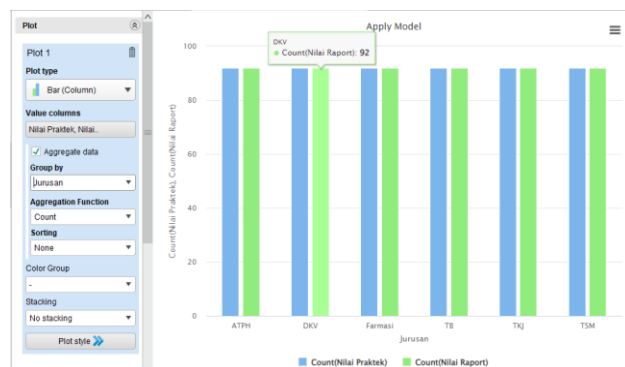
Gambar 2. Proses Pengolahan Data

Operator *Read Excel* digunakan membaca dataset penentuan jurusan siswa yang disimpan dalam file *Microsoft Excel*, pada tahap ini melakukan *import data*. Tahap selanjutnya adalah menentukan atribut-atribut yang akan digunakan pada tahap analisis data. Untuk memilih atribut yang akan digunakan pada perangkat lunak *RapidMiner*, menggunakan operator *select attributes*. Selanjutnya melakukan normalisasi terhadap atribut dengan tipe data numerical, kecuali atribut yang awalnya nominal. Atribut-atribut yang dilakukan normalisasi adalah Nilai Raport dan Nilai praktik. Proses pelabelan untuk menentukan target prediksi menggunakan operator *Set Role* penting untuk menentukan target prediksi untuk model *Decision tree*. Label untuk dataset penentuan jurusan siswa ini adalah atribut jurusan yang terdiri dari ATPH, DKV, Farmasi, TKJ, TB, TSM.



Gambar 3. Penggunaan Operator Smote

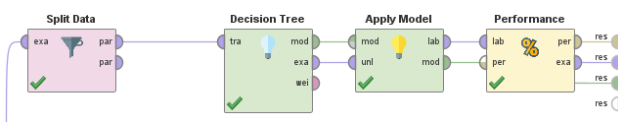
Operator ini menangani ketidakseimbangan kelas dengan menambah sampel sintesis pada kelas minoritas menggunakan metode SMOTE, sehingga model dapat belajar lebih baik dan tidak bias terhadap kelas mayoritas.



Gambar 3. Hasil Smote

Grafik batang menampilkan jumlah data pada setiap kelas untuk dua atribut, yaitu Nilai Raport dan Nilai Praktek, yang telah diolah menggunakan SMOTE. Sebelum SMOTE, kemungkinan terdapat ketidakseimbangan jumlah data antar jurusan, misalnya beberapa jurusan memiliki data yang jauh lebih sedikit dibandingkan lainnya. Setelah SMOTE diterapkan, jumlah data pada setiap jurusan seperti ATPH, DKV, Farmasi, TB, TKJ, dan TSM menjadi merata. membuktikan bahwa metode SMOTE telah berhasil untuk menyelesaikan masalah ketidakseimbangan dalam

data ini, menjadikan data menjadi seimbang antar tiap jurusan. Setiap kategori jurusan masing-masing berjumlah 92 baris data.



Gambar 4. Data Mining dan Evaluasi Model

Pengimplementasian algoritma *Decision Tree* membutuhkan beberapa operator di dalam *RapidMiner*. Proporsi pembagian dataset dengan metode *split data* yang akan dicoba dalam penelitian ini adalah 80:20, dimana 80% dataset sebagai data latih, dan 20% menjadi data uji.

Dalam konteks penentuan jurusan siswa, operator ini memanfaatkan algoritma pohon keputusan untuk menganalisis hubungan antara nilai raport dan nilai praktik sebagai atribut prediktor dan jurusan siswa sebagai target. Model ini membagi data secara iteratif berdasarkan atribut yang paling signifikan dalam menentukan target jurusan siswa, menghasilkan struktur pohon dengan aturan-aturan keputusan di setiap cabangnya.

Decision Tree digunakan untuk membangun model *Decision Tree C4.5* berdasarkan data pelatihan (*tra*) dan parameter tertentu. Data pelatihan ini berisi atribut-atribut yang relevan untuk penentuan jurusan siswa. Model yang dihasilkan (*mod*) kemudian diteruskan ke tahap selanjutnya. *Apply Model* menerapkan model yang telah dibangun pada data uji (unlabeled data). Proses ini menghasilkan prediksi (*lab*) untuk setiap data uji berdasarkan struktur pohon keputusan yang telah dibuat. *Performance Output* berupa persentase akurasi menunjukkan seberapa baik model *Decision Tree C4.5* dalam memprediksi jurusan siswa dengan benar

accuracy: 86.59%

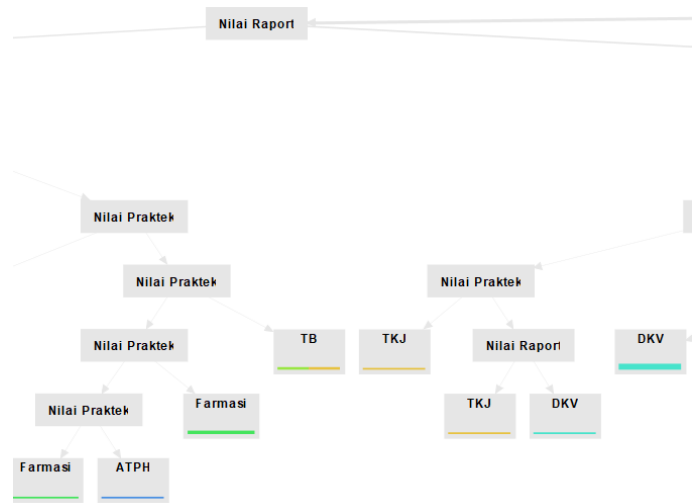
	true ATPH	true DKV	true Farmasi	true TB	true TKJ	true TSM	class preci...
pred. ATPH	88	2	3	2	1	2	89.80%
pred. DKV	1	88	0	4	6	4	85.44%
pred. Farmasi	0	0	78	5	2	0	91.76%
pred. TB	0	0	6	73	12	0	80.22%
pred. TKJ	2	1	4	7	88	3	80.00%
pred. TSM	1	1	1	1	3	83	92.22%
class recall	95.65%	95.65%	84.78%	79.35%	73.91%	90.22%	

Gambar 5. Confusion Matrix

Berdasarkan tabel confusion matrix tersebut, model memiliki akurasi keseluruhan sebesar 86.59%, yang menunjukkan bahwa model berhasil mengklasifikasikan data dengan baik secara umum. Presisi mengukur keakuratan prediksi model untuk setiap kelas, dihitung sebagai rasio prediksi benar terhadap total prediksi untuk kelas tersebut. Recall mengukur sejauh mana model mampu menangkap semua instance dari kelas tertentu, dihitung sebagai rasio prediksi benar terhadap total instance sebenarnya dari kelas tersebut.

Kelas TSM menunjukkan performa terbaik dengan presisi tertinggi (92.22%) dan recall (94.12%). Hal ini mengindikasikan bahwa model mampu memprediksi kelas ini dengan tingkat akurasi yang sangat baik. Kelas ATPH juga menunjukkan performa yang tinggi dengan recall sebesar 95.65%, yang berarti hampir semua instance kelas ini berhasil diidentifikasi dengan benar. Kelas TKJ memiliki performa terendah dengan presisi (76.09%) dan recall (80.00%), yang mengindikasikan bahwa

model mengalami kesulitan dalam mengklasifikasikan data pada kelas ini. Hal ini kemungkinan disebabkan oleh distribusi data yang kurang representatif meskipun SMOTE telah diterapkan. Performa kelas TB juga perlu diperhatikan, dengan recall (84.52%) yang sedikit lebih rendah dibandingkan kelas lainnya.



Gambar 6. Visualisasi Pohon Keputusan

Pohon keputusan ini dimulai dengan nilai raport sebagai *root node*. Nilai raport digunakan untuk membagi data siswa ke dalam dua kategori misalnya, apakah nilai raport mereka tinggi atau rendah. Berdasarkan kategori ini, data akan dibagi menjadi dua cabang yaitu satu untuk siswa dengan nilai raport tinggi, dan satu lagi untuk siswa dengan nilai raport rendah.

Melalui struktur pohon keputusan yang dihasilkan, atribut yang paling signifikan dalam mempengaruhi penentuan jurusan siswa adalah nilai rapor dan nilai praktik. Pohon keputusan membagi data secara iteratif berdasarkan atribut ini, menunjukkan bahwa siswa dengan nilai rapor dan nilai praktik yang tinggi cenderung diarahkan ke jurusan ATPH dan TSM sesuai pola historis pada data pelatihan. Analisis ini menunjukkan bahwa selain faktor akademik, indikator seperti peringkat dan jenis kelamin juga memiliki pengaruh dalam proses penentuan jurusan meskipun tidak sekuat atribut nilai. Proses ini membantu menganalisis faktor utama yang dapat digunakan untuk memberikan rekomendasi penjurusan secara lebih terarah.

5. KESIMPULAN DAN SARAN

Berdasarkan uji yang telah dilakukan, data penjurusan siswa baru dapat dilakukan klasifikasi menggunakan *algoritma C4.5* melalui *RapidMiner*. Hasil klasifikasi menggunakan pohon keputusan, sistem memberikan gambaran yang komprehensif mengenai variabel-variabel yang mempengaruhi pilihan yang dibuat saat memilih program studi bagi setiap siswa berdasarkan nilai raport dan nilai praktik. Dari pemodelan klasifikasi menggunakan *algoritma C4.5* tingkat akurasi mencapai 86.59%, nilai presisi sebesar 89.80% dan nilai recall 95.65%. Adapun saran untuk penelitian selanjutnya yaitu dengan menambahkan jumlah data, menggunakan metode klasifikasi lainnya, menambah atribut atau mengganti atribut yang sesuai dengan kebutuhan penelitian.

UCAPAN TERIMA KASIH

Pada kesempatan kali ini penulis menyampaikan ucapan terima kasih kepada:

- Seluruh civitas akademika STMIK IKMI Cirebon yang telah memberikan kesempatan untuk menempuh pendidikan di kampus ini.
- Dosen Pembimbing yang telah memberikan arahan serta bimbingan.
- Pihak Smk Plus Qurrota A'yun yang telah mengizinkan penulis melakukan observasi.

DAFTAR PUSTAKA

- [1] R. Kuniyasi and A. Fatmawati, "Penerapan Data Mining dengan Algoritma C4.5 untuk Penentuan Jurusan Sekolah Menengah Atas," *J. INSYPRO (Information Syst. Process.*, vol. 6, no. 2, pp. 3–9, 2021, doi: 10.24252/insypro.v6i2.7912.
- [2] N. Manullang, R. W. Sembiring, I. Gunawan, I. Perlina, and Irawan, "Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5," *J. Ilmu Komput. dan Teknol.*, vol. 2, no. 2, pp. 1–5, 2021, [Online]. Available: <http://creativecommons.org/licenses/by/4.0/%0Ahttp://ejournal.uhb.ac.id/index.php/IKOMTI>
- [3] Y. S. Eirlangga and A. E. Syaputra, "Klasifikasi Penjurusan pada Sekolah Menengah Atas (SMA) dengan Metode Algoritma C4.5," *J. Inf. dan Teknol.*, vol. 4, no. 3, pp. 154–159, 2022, doi: 10.37034/jidt.v4i3.235.
- [4] N. A. Sinaga, Ramadani, K. Dalimunthe, M. S. A. A. Lubis, and R. Rosenlly, "Komparasi Metode Decision Tree, KNN, dan SVM Untuk Menentukan Jurusan Di SMK Novendra," *J. Sist. Komput. dan Inform.*, vol. 3, no. 2, pp. 94–100, 2021, doi: 10.30865/json.v3i2.3598.
- [5] M. Zainuri, M. H. Fahmi, and R. A. Hamdhana, "Komparasi Metode Klasifikasi Algoritma C5.0 dan Naïve Bayes untuk Menentukan Jurusan Siswa," *J. Sist. Inf. dan Inform.*, vol. 1, no. 1, pp. 1–8, 2022, doi: 10.33379/jusifor.v1i1.1277.
- [6] M. Y. Putra and D. I. Putri, "Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI Mardi," *J. Tekno Kompak*, vol. 16, no. 2, pp. 176–187, 2022, doi: 10.33365/jtk.v16i2.2002.
- [7] A. Baktiar, B. Soedijono, and A. H. Muhammad, "Menentukan Jurusan Perguruan Tinggi Untuk Peserta Didik SMK Berdasar Nilai Raport Serta Minat Dan Bakat Dengan Metode Decision Tree Algoritma C4.5 (Studi Kasus SMK Negeri 1 Donorojo Pacitan)," *J. PILAR Teknol. J. Ilm. Ilmu Ilmu Tek.*, vol. 8, no. 2, pp. 47–54, 2023, doi: 10.33319/piltek.v8i2.169.
- [8] S. Monalisa and F. Hadi, "Algoritma C4.5 dalam Penentuan Jurusan Siswa Baru," *Ultim. J. Tek. Inform.*, vol. 12, no. 2, pp. 108–113, 2020, doi: 10.31937/ti.v12i2.1838.
- [9] M. Rogib, N. Rahaningsih, and R. danar Dana, "Penerapan algoritma C4.5 untuk seleksi penjurusan siswa baru pada Sekolah Menengah Kejuruan (Studi Kasus: Smk Plus Al-Hilal Arjawinangun)," *J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 861–866, 2024.

BIODATA PENULIS



Yuli Nurul Aini

adalah seorang mahasiswa STMIK IKMI Cirebon dengan Program Studi Teknik Informatika.



Ahmad Faqih

adalah seorang Dosen dari Program studi Teknik Informatika, STMIK IKMI Cirebon.



Gifthera Dwilestari

adalah seorang Dosen dari program studi Sistem Informasi, STMIK IKMI Cirebon.