

Perbandingan Algoritma *Decision Tree*, *ID3*, dan *Random Forest* dalam Klasifikasi Faktor-Faktor yang Mempengaruhi Karier Mahasiswa Ilmu Komputer

Dedy Hartama ¹, Nanda Amalya ^{2*}

^{1,2*} Program Studi Teknik Informatika, STIKOM Tunas Bangsa, Kota Pematangsiantar, Provinsi Sumatera Utara, Indonesia.

Email: dedyhartama@amiktunasbangsa.ac.id ¹, nandaamalya2323@gmail.com ^{2*}

Histori Artikel:

Dikirim 11 Oktober 2024; *Diterima dalam bentuk revisi* 11 November 2024; *Diterima* 20 November 2024; *Diterbitkan* 10 Januari 2025. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma klasifikasi, yaitu Decision Tree, ID3, dan Random Forest, dalam mengidentifikasi faktor-faktor yang mempengaruhi karier mahasiswa ilmu komputer. Algoritma-algoritma ini diaplikasikan pada dataset yang mencakup berbagai atribut mahasiswa, seperti IPK, keterampilan pemrograman, dan proyek yang diselesaikan. Hasil menunjukkan bahwa Random Forest memberikan hasil prediksi yang lebih akurat dan stabil dibandingkan Decision Tree dan ID3, terutama dalam mengurangi risiko overfitting. Mahasiswa dengan keterampilan tinggi di Python dan SQL serta yang berfokus pada pengembangan perangkat lunak cenderung memilih karier di bidang Software Engineering. Sedangkan mereka yang terlibat dalam proyek berbasis AI/ML cenderung memilih Data Science. Kesimpulan dari penelitian ini memberikan wawasan yang berharga bagi institusi pendidikan untuk merancang strategi pengembangan karier yang lebih efektif bagi mahasiswa.

Kata Kunci: Decision Tree; ID3; Random Forest; Klasifikasi; Karier Mahasiswa.

Abstract

This study aims to compare the performance of three classification algorithms, namely Decision Tree, ID3, and Random Forest, in identifying factors that influence the careers of computer Science students. These algorithms are applied to a dataset that includes various student attributes, such as GPA, programming skills, and completed projects. The results show that Random Forest provides more accurate and stable prediction results than Decision Tree and ID3, especially in reducing the risk of overfitting. Students with high skills in Python and SQL and who focus on software development tend to choose a career in Software Engineering. While those involved in AI/ML-based projects tend to choose Data Science. The conclusions of this study provide valuable insights for educational institutions to design more effective career development strategies for students.

Keyword: Decision Tree; ID3; Random Forest; Classification; Student Career.

1. Pendahuluan

Perkembangan teknologi digital yang pesat meningkatkan kebutuhan industri akan lulusan ilmu komputer yang kompeten dan memiliki keterampilan teknis yang unggul. Das *et al.* (2021) mengemukakan bahwa lulusan yang menguasai keterampilan pemrograman dengan baik dan memiliki pengalaman kerja lebih banyak diminati oleh perusahaan teknologi dibandingkan mereka yang hanya mengandalkan kemampuan akademik. Keberhasilan karier mahasiswa tidak hanya dipengaruhi oleh prestasi akademik, tetapi juga oleh berbagai faktor lain seperti keterampilan praktis, pengalaman kerja, kemampuan interpersonal, dan keterlibatan dalam kegiatan ekstrakurikuler (Goh *et al.*, 2020). Identifikasi faktor-faktor yang berpengaruh signifikan terhadap keberhasilan karier mahasiswa menjadi penting, baik bagi perguruan tinggi maupun mahasiswa itu sendiri, untuk mempersiapkan lulusan yang lebih siap menghadapi tantangan dunia kerja. Proses pengelompokan dan analisis faktor-faktor yang memengaruhi karier mahasiswa sering kali dilakukan dengan menggunakan teknik klasifikasi dalam *machine learning*. Petkovic *et al.* (2020) menyatakan bahwa *machine learning* dapat memprediksi hasil karier berdasarkan berbagai atribut yang terkait dengan pendidikan dan keterampilan teknis. Beberapa algoritma yang umum digunakan dalam klasifikasi data adalah *Decision Tree*, *ID3*, dan *Random Forest*. Algoritma-algoritma ini dapat memetakan variabel independen seperti nilai akademik, pengalaman kerja, dan keterampilan lainnya untuk memprediksi keberhasilan karier mahasiswa (Smith, 2019; Kotsiantis, 2019).

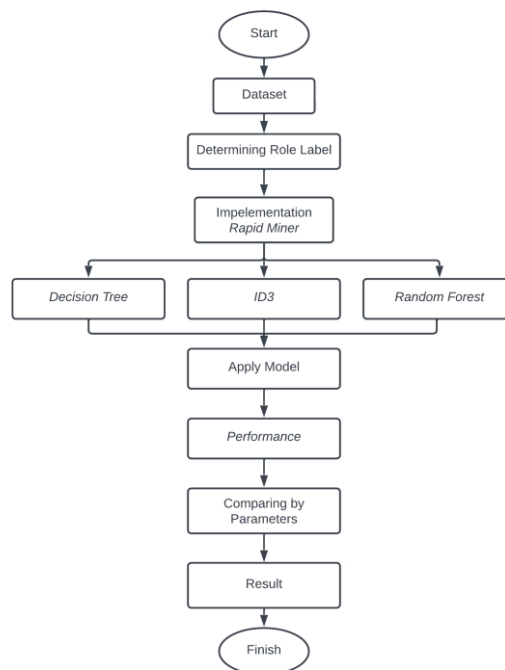
Setiap algoritma memiliki keunggulan dan kelemahan dalam mengklasifikasikan data. *Decision Tree* dikenal mudah dipahami dan diterapkan, namun cenderung rentan terhadap overfitting pada dataset yang besar (Wang *et al.*, 2021). Penelitian Kotsiantis *et al.* (2019) menunjukkan bahwa *Decision Tree* memberikan hasil prediksi yang baik pada dataset kecil, tetapi kinerjanya sering menurun pada data yang lebih kompleks. Algoritma *ID3*, yang merupakan variasi dari *Decision Tree*, menggunakan konsep *information gain* untuk memilih atribut terbaik pada setiap node. Namun, algoritma ini sering kali kurang akurat pada dataset besar (Liaw dan Wiener, 2020). Sebaliknya, *Random Forest*, yang menggabungkan banyak pohon keputusan, dapat menghasilkan klasifikasi yang lebih akurat dengan meminimalkan risiko overfitting, meskipun membutuhkan waktu komputasi yang lebih lama (Kumar *et al.*, 2021). Penelitian ini bertujuan untuk membandingkan kinerja ketiga algoritma tersebut dalam mengklasifikasikan faktor-faktor yang mempengaruhi karier mahasiswa ilmu komputer. Penelitian oleh Kumar *et al.* (2021) juga telah membahas kinerja algoritma ini dalam klasifikasi data pendidikan, di mana *Random Forest* menunjukkan hasil yang lebih stabil dibandingkan algoritma lainnya. Oleh karena itu, penelitian ini juga akan menganalisis faktor-faktor yang paling dominan dalam menentukan kesuksesan karier mahasiswa berdasarkan hasil klasifikasi dari ketiga algoritma tersebut. Hasil penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik bagi institusi pendidikan dan mahasiswa dalam merancang strategi pengembangan karier yang efektif.

2. Metode Penelitian

2.1 Rancangan Penelitian

Proses penelitian ini dimulai dengan persiapan *dataset* yang berisi atribut-atribut terkait karier mahasiswa. Setelah itu, dilakukan penentuan peran label untuk membedakan variabel target dan variabel independen. *Dataset* kemudian diimplementasikan menggunakan perangkat lunak *RapidMiner*, di mana *dataset* yang sama digunakan untuk melatih tiga algoritma, yaitu *Decision Tree*, *ID3*, dan *Random Forest* secara paralel. Model yang telah dilatih diterapkan pada *dataset* uji untuk memprediksi hasil, kemudian kinerjanya dievaluasi menggunakan metrik performa seperti akurasi, *precision*, *recall*, dan *F1-score*. Kinerja ketiga algoritma dibandingkan berdasarkan parameter performa tersebut untuk mengetahui algoritma mana yang paling efektif dalam mengklasifikasikan faktor-faktor yang mempengaruhi karier mahasiswa. Akhirnya, hasil penelitian menunjukkan algoritma dengan kinerja

terbaik dan memberikan pemahaman mengenai faktor-faktor dominan yang mempengaruhi jalur karier mahasiswa.



Gambar 1. Rancangan Penelitian

2.2 Dataset

Dataset yang digunakan mencakup berbagai atribut yang berkaitan dengan profil akademik dan kemampuan teknis mahasiswa ilmu komputer. Setiap mahasiswa diidentifikasi dengan StudentID, nama, jenis kelamin, umur, IPK, serta bidang yang diminati, baik Pemrograman maupun Non-Pemrograman. Dataset juga mencakup kemampuan teknis mahasiswa dalam bahasa pemrograman seperti Python, SQL, dan Java yang dikategorikan sebagai "Mahir", "Sedang", atau "Tidak Mahir". Selain itu, proyek yang diselesaikan seperti Software Development, Data Science, atau AI/ML juga dicatat sebagai faktor penentu dalam jalur karier mahasiswa. Variabel Karier Masa Depan berfungsi sebagai target klasifikasi untuk memprediksi karier yang ingin dicapai mahasiswa, seperti Software Engineering, Data Science, atau Web Development. Keseluruhan atribut ini digunakan dalam model klasifikasi berbasis machine learning untuk memprediksi karier mahasiswa berdasarkan profil teknis dan akademis mereka.

Tabel 1. Data Karier Mahasiswa Ilmu Komputer

Student ID	Nama	Jenis Kelamin	Umur	IPK	Bidang Yang Diminati	Python	SQL	Java	Proyek Yang Diselesaikan	Karier Masa Depan
1	Person 1	Laki-Laki	Muda	Sedang	Pemrograman	Mahir	Mahir	Tidak Mahir	AI/ML	Software Engineering
2	Person 2	Perempuan	Muda	Sedang	Pemrograman	Sedang	Mahir	Tidak Mahir	Data Science	Data Science
3	Person 3	Laki-Laki	Dewasa	Baik	Pemrograman	Mahir	Mahir	Sedang	Software Development	Software Engineering
4	Person 4	Perempuan	Muda	Baik	Pemrograman	Tidak Mahir	Mahir	Mahir	Software Development	Web Development
5	Person 5	Laki-Laki	Dewasa	Sedang	Non Pemrograman	Sedang	Tidak Mahir	Mahir	Software Development	Software Engineering
6	Person 6	Perempuan	Dewasa	Baik	Non Pemrograman	Mahir	Sedang	Tidak Mahir	Software Development	Software Engineering
7	Person 7	Laki-Laki	Muda	Baik	Non Pemrograman	Sedang	Mahir	Tidak Mahir	Software Development	Software Engineering

8	Person 8	Perempuan	Muda	Baik	Non Pemrograman	Tidak Mahir	Mahir	Sedang	Software Development	Software Engineering
9	Person 9	Laki-Laki	Dewasa	Sedang	Non Pemrograman	Sedang	Tidak Mahir	Mahir	Software Development	Software Engineering
10	Person 10	Perempuan	Dewasa	Sedang	Non Pemrograman	Tidak Mahir	Sedang	Mahir	Software Development	Software Engineering
...	...	...	...	...	...	...	...	...	...	...
180	Person 180	Laki-Laki	Dewasa	Baik	Non Pemrograman	Mahir	Mahir	Tidak Mahir	Software Development	Software Engineering

2.3 Machine Learning dan Klasifikasi

Machine learning adalah salah satu bidang penting dalam kecerdasan buatan (*artificial intelligence*), yang memungkinkan sistem untuk belajar dari data dan melakukan prediksi atau pengambilan keputusan tanpa harus diprogram secara eksplisit. Klasifikasi adalah teknik yang paling umum dalam *machine learning*, yang digunakan untuk mengelompokkan data ke dalam kelas-kelas yang telah ditentukan. Menurut Das *et al.* (Das *et al.*, 2021), klasifikasi berbasis *machine learning* sangat efektif untuk memecahkan masalah dalam berbagai domain, termasuk pendidikan (Amiruddin *et al.*, 2019). Algoritma seperti *Decision Tree*, *ID3*, dan *Random Forest* digunakan secara luas dalam penelitian untuk memprediksi berbagai hal berdasarkan atribut data (Quinlan, 2021).

2.4 Algoritma Decision Tree

Decision Tree adalah algoritma yang membentuk pohon keputusan berdasarkan data input. Setiap node dalam pohon mewakili suatu atribut, sementara cabang-cabang mewakili hasil yang mungkin terjadi. Algoritma ini bekerja dengan membagi dataset berdasarkan atribut yang paling relevan untuk menghasilkan klasifikasi. Menurut Petkovic *et al.* (Petkovic *et al.*, 2020), algoritma *Decision Tree* sangat intuitif dan mudah diinterpretasikan, sehingga banyak digunakan dalam dunia akademis maupun industri. Namun, kelemahannya adalah mudah terjadi *overfitting* ketika pohon menjadi terlalu kompleks (Liaw and Wiener, 2020), (Kumar *et al.*, 2021).

2.5 Algoritma ID3

Algoritma *ID3* (*Iterative Dichotomiser 3*) adalah varian dari *Decision Tree* yang diperkenalkan oleh Quinlan, yang menggunakan *information gain* untuk menentukan atribut terbaik pada setiap node. Atribut dengan *information gain* tertinggi dipilih untuk membagi dataset di setiap tahap pembentukan pohon. Penelitian yang dilakukan oleh Kotsiantis *et al.* (Kotsiantis, 2019) menunjukkan bahwa algoritma *ID3* bekerja sangat baik pada dataset kecil, tetapi memiliki keterbatasan pada dataset besar karena rentan terhadap ketidakakuratan (Breiman, 2020).

2.6 Algoritma Random Forest

Random Forest adalah pengembangan dari *Decision Tree*, yang menggunakan banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko *overfitting*. Setiap pohon dalam *Random Forest* dibentuk dengan subset data yang berbeda, dan prediksi akhir diperoleh dengan melakukan voting dari hasil prediksi setiap pohon (Boie, 2020). Menurut Liaw *et al.* (Liaw and Wiener, 2021), *Random Forest* adalah salah satu algoritma klasifikasi terbaik dalam hal stabilitas dan akurasi prediksi, namun membutuhkan waktu komputasi yang lebih lama dibandingkan dengan algoritma lain (Delgado *et al.*, 2019).

2.7 Faktor-Faktor Yang Mempengaruhi Karier Mahasiswa

Faktor-faktor seperti prestasi akademik, keterampilan teknis, pengalaman kerja, dan keterlibatan dalam kegiatan ekstrakurikuler diketahui memiliki pengaruh besar terhadap peluang karier mahasiswa. Menurut penelitian yang dilakukan oleh Goh *et al.* (Goh *et al.*, 2020), mahasiswa dengan keterampilan teknis yang baik, seperti pemrograman dan pengolahan data, memiliki peluang lebih besar untuk berhasil dalam karier mereka. Selain itu, penelitian yang dilakukan oleh Howard *et al.* (Howard *et al.*, 2020) menunjukkan bahwa keterlibatan dalam kegiatan organisasi dan magang selama masa kuliah juga sangat berpengaruh terhadap kesuksesan karier.

2.8 Faktor-Faktor Yang Mempengaruhi Karier Mahasiswa

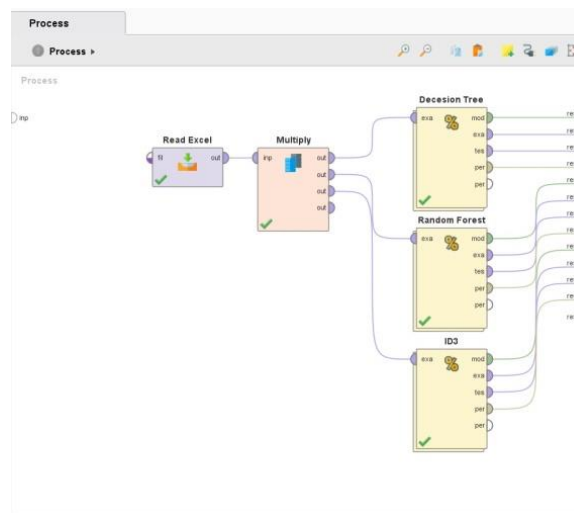
Beberapa penelitian telah membandingkan algoritma klasifikasi dalam konteks pendidikan. Menurut Kumar *et al.* (Kumar *et al.*, 2021), *Random Forest* sering kali lebih akurat dibandingkan dengan *Decision Tree* dan *ID3* dalam masalah klasifikasi yang melibatkan dataset besar dan kompleks (Nicole, 2020). Namun, *ID3* lebih efisien dari segi waktu komputasi pada dataset kecil. Penelitian lain oleh Young (Young, 2021) mengungkapkan bahwa *Decision Tree* dapat memberikan hasil yang cukup akurat, tetapi lebih rentan terhadap overfitting jika dataset yang digunakan terlalu besar (Zhang and Du, 2020).

3. Hasil dan Pembahasan

3.1 Hasil

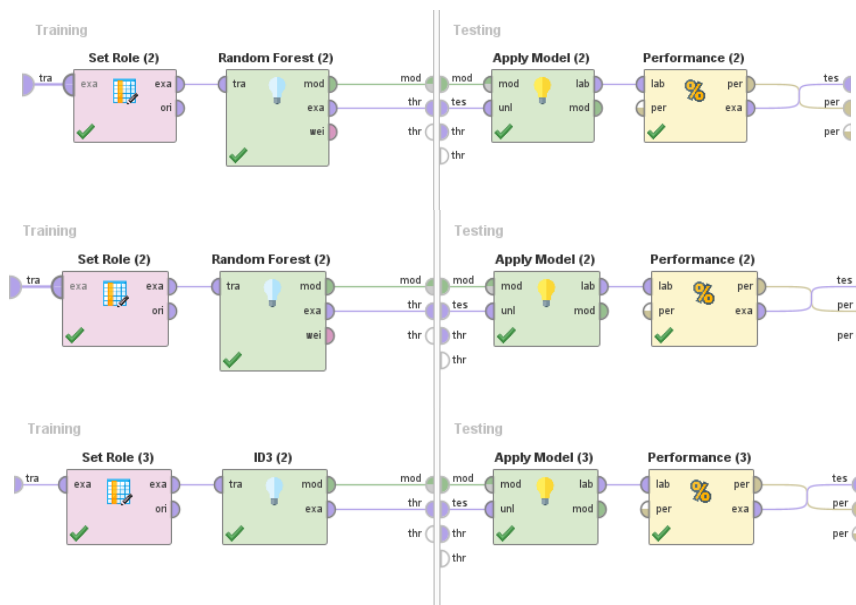
3.1.1 Design Rapidminer

Proses klasifikasi yang dijalankan pada *RapidMiner*, baik dari desain luar maupun dalam, secara keseluruhan melibatkan beberapa langkah kunci yang bekerja secara sinergis untuk memprediksi karier mahasiswa berdasarkan faktor-faktor yang tersedia dalam dataset.



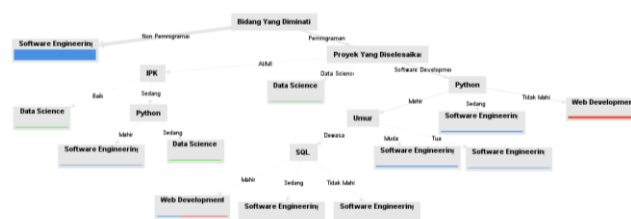
Gambar 2. Design Proses Perhitungan *Rapidminer*

Proses perhitungan di *RapidMiner* melibatkan beberapa langkah utama untuk melakukan klasifikasi menggunakan algoritma *Decision Tree*, *Random Forest*, dan *ID3*. Tahap pertama adalah mengimpor dataset menggunakan komponen *Read Excel*, yang berfungsi sebagai titik awal untuk memasukkan data ke dalam alur pemrosesan. Setelah itu, data diduplikasi menggunakan komponen *Multiply* agar dapat digunakan oleh beberapa algoritma sekaligus. Dataset yang sama kemudian dimasukkan ke dalam tiga algoritma pembelajaran mesin berbeda: *Decision Tree* untuk membangun pohon keputusan berdasarkan variabel yang ada, *Random Forest* yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi, dan *ID3* yang menggunakan information gain untuk memilih atribut paling relevan pada setiap tahap. Setelah model dilatih menggunakan masing-masing algoritma, proses evaluasi dilakukan melalui beberapa output, termasuk *exa* (evaluasi model berdasarkan data pelatihan), *mod* (modifikasi hasil prediksi), *tes* (pengujian model dengan data baru), dan *per* (evaluasi performa menggunakan metrik seperti precision, recall, dan F1-score). Dengan desain ini, setiap algoritma dijalankan secara paralel, memungkinkan perbandingan langsung untuk menentukan algoritma mana yang paling efektif dalam mengklasifikasikan faktor-faktor yang mempengaruhi karier mahasiswa.



Gambar 3. Design Proses Perhitungan Rapidminer Training Testing

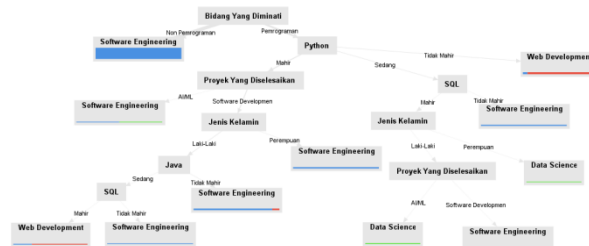
Pada tahap *Training*, data diatur menggunakan komponen *Set Role* untuk menetapkan variabel mana yang menjadi label (variabel dependen) dan fitur (variabel independen). Setelah itu, data diproses oleh tiga algoritma pembelajaran mesin, yaitu *Random Forest*, *Decision Tree*, dan *ID3*. Setiap algoritma melatih model berdasarkan dataset yang ada, di mana *Random Forest* menggunakan beberapa pohon keputusan untuk mengurangi risiko *overfitting*, *Decision Tree* membangun satu pohon keputusan berdasarkan atribut yang ada, dan *ID3* membentuk pohon dengan memilih atribut berdasarkan *information gain*. Pada tahap *Testing*, model yang telah dilatih diterapkan pada data uji menggunakan komponen *Apply Model*, dan hasil prediksi dari model tersebut dianalisis lebih lanjut. Kinerja setiap model dievaluasi menggunakan komponen *Performance*, yang mengukur akurasi, presisi, recall, dan F1-score dari prediksi yang dihasilkan. Proses ini mencakup alur lengkap dari membaca data, melatih model, hingga mengevaluasi hasil prediksi, dan memungkinkan perbandingan langsung kinerja dari ketiga algoritma dalam tugas klasifikasi faktor yang mempengaruhi karier mahasiswa.



Gambar 4. Hasil Algoritma Random Forest

Algoritma *Random Forest* menunjukkan hasil yang lebih akurat dan stabil dalam memprediksi faktor-faktor yang mempengaruhi karier mahasiswa. Hasil visualisasi menunjukkan bahwa *Random Forest* menggunakan berbagai atribut seperti kemampuan Python, SQL, serta proyek yang telah diselesaikan sebagai penentu utama dalam menentukan arah karier mahasiswa. Algoritma ini juga berhasil meminimalkan risiko *overfitting* dengan cara menggabungkan beberapa pohon keputusan yang berbeda, sehingga prediksi tidak terlalu bergantung pada satu model saja. Mahasiswa yang memiliki kemampuan lebih tinggi di Python dan SQL serta berfokus pada proyek *Software Development*

cenderung memilih karier di bidang *Software Engineering*. Sementara itu, mereka yang terlibat dalam proyek berbasis AI/ML lebih banyak mengarah ke *Data Science*.



Gambar 5. Hasil Algoritma *Decision Tree*

Algoritma *Decision Tree* memberikan struktur yang lebih sederhana dalam klasifikasi. Pada setiap node, algoritma ini memilih atribut dengan *information gain* tertinggi, yang memungkinkan pemetaan keputusan lebih langsung. Namun, hasil visualisasi menunjukkan bahwa algoritma ini lebih rentan terhadap *overfitting*, khususnya pada dataset yang lebih besar. Meskipun demikian, *Decision Tree* tetap berhasil mengidentifikasi atribut penting seperti kemampuan pemrograman Python dan SQL sebagai faktor dominan dalam menentukan karier mahasiswa. Mahasiswa yang memiliki keterampilan pemrograman yang kuat dan fokus pada pengembangan perangkat lunak umumnya diprediksi akan memilih jalur karier di bidang *Software Engineering*.



Gambar 6. Hasil Algoritma *ID3*

Algoritma *ID3*, yang merupakan varian dari *Decision Tree*, juga memperlihatkan hasil yang cukup baik dalam mengklasifikasi data. *ID3* memilih atribut berdasarkan *information gain*, di mana atribut dengan nilai tertinggi menjadi penentu utama di setiap tahap pengambilan keputusan. Hasil dari *ID3* mengkonfirmasi bahwa mahasiswa yang mahir dalam Python dan terlibat dalam proyek *Software Development* kemungkinan besar akan memilih karier di *Software Engineering*, sementara mereka yang terlibat dalam proyek berbasis AI/ML akan lebih tertarik pada *Data Science*. Namun, seperti halnya *Decision Tree*, *ID3* juga berpotensi kurang akurat pada dataset yang lebih kompleks karena model yang dihasilkan sering kali terlalu spesifik pada data pelatihan. Secara keseluruhan, hasil dari ketiga algoritma ini menunjukkan bahwa *Random Forest* memberikan hasil klasifikasi yang lebih akurat dan generalis jika dibandingkan dengan *Decision Tree* dan *ID3*. Prediksi dari *Random Forest* lebih stabil dan tidak rentan terhadap *overfitting*, sehingga menghasilkan kinerja yang lebih baik dalam klasifikasi faktor-faktor yang mempengaruhi karier mahasiswa ilmu komputer. Faktor-faktor seperti kemampuan teknis dalam Python, SQL, dan proyek yang diselesaikan terbukti menjadi penentu utama dalam pemilihan karier di bidang *Software Engineering* dan *Data Science*.

3.2 Pembahasan

Penelitian ini mengevaluasi tiga algoritma pembelajaran mesin, yaitu *Random Forest*, *Decision Tree*, dan *ID3*, dalam mengklasifikasikan faktor-faktor yang memengaruhi karier mahasiswa ilmu komputer. Berdasarkan tahapan klasifikasi menggunakan *RapidMiner*, setiap algoritma diuji secara paralel menggunakan *dataset* yang sama. Proses ini mencakup pengolahan data dengan komponen *Multiply* untuk menduplikasi data ke dalam tiga model yang berbeda. Pendekatan ini sejalan dengan panduan dari Kumar *et al.* (2021) yang menunjukkan pentingnya perbandingan kinerja algoritma untuk mengidentifikasi model yang paling sesuai berdasarkan metrik akurasi, *presisi*, *recall*, dan *F1-score*. *Random Forest* menunjukkan performa terbaik dalam penelitian ini. Algoritma ini menggunakan banyak pohon keputusan untuk menghasilkan prediksi yang lebih akurat dan stabil. Keunggulan ini tercermin dalam kemampuan algoritma untuk meminimalkan *overfitting* dan menghasilkan prediksi yang lebih umum pada *dataset* yang kompleks. Penemuan ini didukung oleh Breiman (2020) dan Liaw dan Wiener (2020), yang menyatakan bahwa *Random Forest* mampu menghasilkan model klasifikasi dengan tingkat akurasi yang tinggi pada berbagai jenis *dataset*. Selain itu, atribut seperti keterampilan *Python*, *SQL*, dan keterlibatan dalam proyek berbasis *AI/ML* atau pengembangan perangkat lunak muncul sebagai prediktor utama jalur karier mahasiswa, sesuai dengan temuan Goh *et al.* (2020) mengenai peran keterampilan teknis dalam menentukan keberhasilan karier. *Decision Tree* menghasilkan struktur klasifikasi yang sederhana dengan menggunakan nilai *information gain* untuk memilih atribut terbaik. Algoritma ini efektif untuk *dataset* kecil dan memberikan hasil yang mudah dipahami, sebagaimana diuraikan oleh Petkovic *et al.* (2020). Namun, algoritma ini rentan terhadap *overfitting* pada *dataset* besar, seperti yang dilaporkan oleh Wang *et al.* (2021). Dalam penelitian ini, *Decision Tree* berhasil mengidentifikasi atribut penting seperti *Python* dan *SQL* sebagai variabel dominan yang memengaruhi prediksi karier mahasiswa.

Algoritma *ID3*, yang merupakan pengembangan dari *Decision Tree*, juga menunjukkan efisiensi dalam memilih atribut berdasarkan nilai *information gain*. Quinlan (2021) menjelaskan bahwa algoritma ini dirancang untuk mengoptimalkan pemilihan atribut, namun kurang mampu menangani *dataset* yang kompleks karena model yang dihasilkan sering kali terlalu spesifik terhadap data pelatihan. Pada penelitian ini, *ID3* menunjukkan hasil yang cukup baik pada *dataset* yang lebih kecil, namun kurang kompetitif dibandingkan *Random Forest* untuk data yang lebih besar atau heterogen. Hasil penelitian menunjukkan bahwa *Random Forest* adalah algoritma yang paling optimal untuk klasifikasi faktor-faktor yang memengaruhi karier mahasiswa ilmu komputer. Faktor teknis seperti keterampilan dalam *Python* dan *SQL* serta keterlibatan dalam proyek pengembangan perangkat lunak menjadi indikator utama. Studi ini sejalan dengan temuan Zhang dan Du (2020) yang menegaskan bahwa algoritma seperti *Random Forest* unggul dalam menangkap pola kompleks dalam data pendidikan. Institusi pendidikan dapat menggunakan hasil ini untuk merancang program pelatihan teknis yang lebih terarah guna meningkatkan kesiapan mahasiswa dalam menghadapi tuntutan industri teknologi informasi.

4. Kesimpulan

Penelitian ini membandingkan kinerja tiga algoritma pembelajaran mesin, yakni *Decision Tree*, *ID3*, dan *Random Forest*, dalam mengklasifikasikan faktor-faktor yang mempengaruhi karier mahasiswa ilmu komputer. Dari hasil evaluasi, *Random Forest* terbukti memberikan akurasi prediksi yang lebih baik dibandingkan dengan *Decision Tree* dan *ID3*, terutama dalam mencegah *overfitting* pada *dataset* yang lebih besar dan kompleks. Faktor-faktor seperti keterampilan teknis di *Python* dan *SQL*, serta keterlibatan dalam proyek pengembangan perangkat lunak, terbukti menjadi penentu utama dalam jalur karier mahasiswa. Mahasiswa dengan fokus pada proyek *AI/ML* cenderung memilih karier di *Data Science*, sementara yang berfokus pada pengembangan perangkat lunak memilih jalur *Software Engineering*. Kesimpulan ini dapat digunakan oleh institusi pendidikan untuk merancang kurikulum dan strategi pengembangan karier yang lebih tepat guna bagi mahasiswa.

5. Daftar Pustaka

- Agarwal, S., Pandey, G. N., & Tiwari, M. D. (2012). Data mining in education: data classification and decision tree approach. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 2(2), 140.
- Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, 9(5), 272.
- Azam, Z., Islam, M. M., & Huda, M. N. (2023). Comparative analysis of intrusion detection systems and machine learning based model analysis through decision tree. *IEEE Access*.
- Chaudhuri, A. K., Ray, A., Banerjee, D. K., & Das, A. (2021). A multi-stage approach combining feature selection with machine learning techniques for higher prediction reliability and accuracy in cervical cancer diagnosis. *International Journal of Intelligent Systems and Applications*, 10(5), 46.
- Gupta, U. G., & Houtz, L. E. (2000). High school students' perceptions of information technology skills and careers. *Journal of Industrial Technology*, 16(4), 2-8.
- Jha, K., Likhitha, D., Chandana, M. S., Reddy, M. R. P., & Bhargavi, M. (2024, July). Career Prediction Using Machine Learning. In *2024 8th International Conference on Inventive Systems and Control (ICISC)* (pp. 118-122). IEEE.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3-24.
- Lewis, C. P. (2004). *The relation between extracurricular activities with academic and social competencies in school-age children: A meta-analysis*. Texas A&M University.
- Liaw, A., & Wiener, M. (2002). *Classification and regression by randomForest*. *R News* 2 (3): 18–22.
- Mikhalkina, E. V., Skachkova, L. S., & Gerasimova, O. Y. (2019). Academic or non-academic career: What choice do graduates of federal universities make. *Terra Economicus*, 17(4), 148-173.
- Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing*, 5(2), 221.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1, 81-106.
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, 20(1), 3-29. DOI: <https://doi.org/10.1177/1536867X20909688>.
- VidyaShreeram, N., & Muthukumaravel, A. (2021, June). Student career prediction using machine learning approaches. In *First International Conference on Computing, Communication and Control System* (p. 444).
- Yang, Z., Yang, J., Rice, K., Hung, J. L., & Du, X. (2020). Using convolutional neural network to recognize learning images for early warning of at-risk students. *IEEE Transactions on Learning Technologies*, 13(3), 617-630.