

Penerapan Algoritma Machine Learning untuk Memprediksi Employability Alumni Berdasarkan Data Tracer Study

Application of Machine Learning Algorithms to Predict Alumni Employability Based on Tracer Study Data

Ayu Lestari Aziz¹; Firman Aziz^{2,*}; Irmawati³

^{1,3} Irmex Digital Akademika, Makassar 90551, Indonesia

² Universitas Pancasakti, Makassar 90121, Indonesia

¹ayulestariazis@irmexdigika.com; ²firman.aziz@unpacti.ac.id; ³irmawati@irmexdigika.com

* Corresponding author

Abstrak

Tracer study merupakan instrumen penting bagi perguruan tinggi untuk mengevaluasi kinerja lulusan dan tingkat kesiapan mereka dalam memasuki dunia kerja. Namun, pemanfaatan data tracer study umumnya masih terbatas pada analisis deskriptif, sehingga belum mampu memberikan prediksi yang komprehensif. Penelitian ini bertujuan untuk menerapkan algoritma machine learning dalam memprediksi employability alumni berdasarkan data tracer study. Metode yang digunakan meliputi tahapan preprocessing data, pemodelan menggunakan beberapa algoritma klasifikasi yaitu Decision Tree, K-Nearest Neighbor, Support Vector Machine, dan Random Forest, serta evaluasi model menggunakan metrik akurasi, precision, recall, F1-score, dan confusion matrix. Hasil eksperimen menunjukkan bahwa algoritma Random Forest memberikan performa terbaik dengan tingkat akurasi tertinggi dibandingkan algoritma lainnya. Analisis feature importance mengungkapkan bahwa soft skills, pengalaman magang, dan hard skills merupakan faktor paling berpengaruh terhadap employability alumni, sementara indikator akademik seperti IPK memiliki kontribusi yang lebih rendah. Penelitian ini membuktikan bahwa pendekatan machine learning efektif dalam mengoptimalkan pemanfaatan data tracer study dan berpotensi dikembangkan sebagai sistem pendukung keputusan berbasis komputer untuk meningkatkan kualitas lulusan perguruan tinggi.

Kata Kunci: machine learning; tracer study; employability alumni; random forest; klasifikasi

Abstract

Tracer study is an important instrument for higher education institutions to evaluate graduate performance and readiness for the labor market. However, the utilization of tracer study data is often limited to descriptive analysis, which lacks predictive capability. This study aims to apply machine learning algorithms to predict alumni employability based on tracer study data. The proposed method consists of data preprocessing, classification modeling using Decision Tree, K-Nearest Neighbor, Support Vector Machine, and Random Forest algorithms, and model evaluation using accuracy, precision, recall, F1-score, and confusion matrix. Experimental results indicate that the Random Forest algorithm outperforms other models by achieving the highest accuracy. Feature importance analysis reveals that soft skills, internship experience, and hard skills are the most influential factors affecting alumni employability, while academic indicators such as GPA contribute less significantly. This study demonstrates that machine learning approaches are effective in enhancing the utilization of tracer study data and can be developed as a computer-based decision support system to improve graduate quality.

Keywords: machine learning; tracer study; alumni employability; random forest; classification

Pendahuluan

Perguruan tinggi memiliki tanggung jawab strategis dalam menyiapkan lulusan yang mampu bersaing di dunia kerja. Salah satu indikator keberhasilan pendidikan tinggi adalah employability alumni, yaitu kemampuan lulusan untuk memperoleh pekerjaan, mempertahankannya, serta beradaptasi dengan kebutuhan industri yang dinamis. Untuk mengukur indikator tersebut, perguruan tinggi secara rutin melaksanakan tracer study, yang menghasilkan data mengenai kompetensi lulusan, pengalaman pembelajaran, dan kondisi kerja setelah lulus [1]. Meskipun tracer study telah menjadi instrumen wajib dalam evaluasi kinerja perguruan tinggi, pemanfaatan data yang dihasilkan masih didominasi oleh analisis deskriptif, seperti persentase alumni bekerja, waktu tunggu kerja, dan kesesuaian bidang pekerjaan [2]. Pendekatan ini belum mampu memberikan informasi prediktif yang dapat digunakan sebagai dasar pengambilan keputusan strategis, khususnya dalam perbaikan kurikulum dan kebijakan akademik.

Seiring dengan perkembangan ilmu komputer, khususnya di bidang machine learning, analisis data tidak lagi terbatas pada pengolahan statistik konvensional. Machine learning memungkinkan sistem komputer untuk mempelajari pola dari data historis dan menghasilkan model prediksi yang akurat tanpa harus diprogram secara eksplisit [3]. Pendekatan ini telah banyak diterapkan pada berbagai domain, termasuk bidang pendidikan, untuk memprediksi kinerja akademik mahasiswa, tingkat kelulusan, dan risiko putus studi [4], [5]. Dalam konteks tracer study, data yang dihasilkan memiliki karakteristik multidimensi dan heterogen, yang mencakup variabel numerik, kategorikal, serta skala Likert. Kompleksitas ini menjadikan machine learning sebagai pendekatan yang relevan untuk mengidentifikasi pola tersembunyi dan hubungan non-linear antarvariabel yang sulit dideteksi melalui metode analisis tradisional [6]. Beberapa penelitian sebelumnya menunjukkan bahwa algoritma klasifikasi seperti Decision Tree, Support Vector Machine, dan Random Forest mampu memberikan performa yang baik dalam memprediksi outcome pendidikan dan ketenagakerjaan [7], [8].

Random Forest, sebagai salah satu algoritma ensemble learning, memiliki keunggulan dalam mengatasi masalah overfitting dan mampu menangani data dengan jumlah fitur yang besar [9]. Selain menghasilkan prediksi yang akurat, algoritma ini juga menyediakan informasi feature importance yang berguna untuk mengidentifikasi faktor-faktor dominan yang memengaruhi employability alumni. Informasi tersebut sangat penting bagi perguruan tinggi dalam merumuskan kebijakan berbasis data. Namun demikian, penelitian yang secara khusus menerapkan machine learning untuk memprediksi employability alumni berbasis data tracer study, khususnya dalam konteks perguruan tinggi di Indonesia, masih relatif terbatas. Sebagian besar penelitian tracer study masih berfokus pada evaluasi deskriptif tanpa pengembangan model prediktif yang bersifat komputasional [10].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menerapkan dan membandingkan beberapa algoritma machine learning dalam memprediksi employability alumni berdasarkan data tracer study. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan ilmu komputer terapan, khususnya dalam pemanfaatan machine learning sebagai sistem pendukung keputusan di bidang pendidikan tinggi.

Kontribusi utama penelitian ini terletak pada penerapan dan evaluasi komprehensif algoritma machine learning untuk memprediksi employability alumni berbasis data tracer study, yang tidak hanya menekankan aspek akurasi prediksi tetapi juga interpretabilitas model melalui analisis feature importance. Berbeda dengan penelitian tracer study sebelumnya yang umumnya bersifat deskriptif atau statistik konvensional, penelitian ini mengintegrasikan pendekatan komputasi modern dengan evaluasi multi-algoritma serta analisis confusion matrix untuk menilai kestabilan klasifikasi per kelas. Selain itu, penelitian ini memberikan bukti empiris bahwa soft skills dan pengalaman magang memiliki kontribusi yang lebih signifikan dibandingkan indikator akademik dalam menentukan employability alumni, sehingga memperkaya pemanfaatan data tracer study sebagai dasar pengambilan keputusan strategis berbasis sistem pendukung keputusan komputer di perguruan tinggi.

Metode

A. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental berbasis komputasi (computational experiment) dengan tujuan membangun dan mengevaluasi model prediksi employability alumni menggunakan algoritma machine learning. Pendekatan ini umum digunakan dalam penelitian ilmu komputer untuk membandingkan performa beberapa algoritma pada dataset yang sama secara objektif [11]. Alur metodologi penelitian terdiri atas beberapa tahapan utama, yaitu pengumpulan data, preprocessing data, pemilihan fitur, pemodelan machine learning, serta evaluasi performa model.

B. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini berasal dari Data Tracer Study Alumni Tahun 2024 yang dikumpulkan oleh perguruan tinggi melalui kuesioner terstruktur. Dataset terdiri dari 246 responden dengan 93 atribut yang mencakup informasi kompetensi lulusan, pengalaman pembelajaran, serta kondisi kerja alumni.

Untuk menjaga kerahasiaan data, seluruh atribut yang mengandung informasi identitas pribadi seperti nama, nomor induk mahasiswa, dan alamat surel dihapus sebelum proses analisis. Penggunaan data sekunder seperti tracer study untuk penelitian machine learning telah banyak diterapkan dalam studi pendidikan dan ketenagakerjaan [12].

C. Preprocessing Data

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam pemodelan machine learning. Tahapan preprocessing meliputi:

1. Data Cleaning

Menghapus atribut yang tidak relevan dan menangani data duplikat.

2. **Penanganan Missing Value**

Nilai kosong pada data numerik diisi menggunakan nilai rata-rata, sedangkan data kategorikal diisi menggunakan nilai modus. Teknik ini umum digunakan untuk menjaga distribusi data [13].

3. **Transformasi Data**

- Data kategorikal diubah menjadi data numerik menggunakan teknik *label encoding*.
- Data berskala Likert dikonversi ke dalam bentuk numerik dengan rentang 1 sampai 5.

4. **Normalisasi Data**

Data numerik dinormalisasi menggunakan metode Min-Max Scaling untuk memastikan setiap fitur berada pada rentang yang sama dan tidak mendominasi proses pembelajaran model [14].

D. Seleksi dan Definisi Fitur

Seleksi fitur dilakukan untuk mengurangi kompleksitas model dan meningkatkan performa klasifikasi. Fitur-fitur yang digunakan dalam penelitian ini meliputi:

- Kompetensi hard skills
- Kompetensi soft skills
- Pengalaman magang
- Pengalaman organisasi
- Metode pembelajaran
- Indeks Prestasi Kumulatif (IPK)
- Kesesuaian bidang studi dengan pekerjaan

Variabel target (label) dalam penelitian ini adalah employability alumni, yang diklasifikasikan ke dalam dua kelas, yaitu bekerja dan tidak bekerja. Pendekatan klasifikasi biner ini banyak digunakan dalam penelitian prediksi outcome ketenagakerjaan [15].

E. Algoritma Machine Learning

Penelitian ini menggunakan empat algoritma klasifikasi yang umum digunakan dalam penelitian ilmu komputer, yaitu:

1. **Decision Tree**

Algoritma berbasis pohon keputusan yang mudah diinterpretasikan dan mampu menangani data kategorikal [16].

2. **K-Nearest Neighbor (KNN)**

Algoritma berbasis jarak yang mengklasifikasikan data berdasarkan kedekatan dengan data latih terdekat [17].

3. **Support Vector Machine (SVM)**

Algoritma yang mencari hyperplane optimal untuk memisahkan kelas dengan margin maksimum [7].

4. **Random Forest**

Algoritma ensemble learning yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi overfitting [9].

F. Pembagian Data dan Validasi

Dataset dibagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%. Selain itu, penelitian ini menggunakan 5-fold cross-validation untuk memastikan stabilitas dan generalisasi model. Teknik validasi silang banyak direkomendasikan dalam penelitian machine learning untuk menghindari bias pembagian data [19].

G. Metode Evaluasi Model

Evaluasi performa model dilakukan menggunakan beberapa metrik evaluasi klasifikasi, yaitu:

- Accuracy, untuk mengukur tingkat ketepatan prediksi secara keseluruhan

- Precision, untuk mengukur ketepatan prediksi kelas positif
- Recall, untuk mengukur kemampuan model dalam mendeteksi kelas positif
- F1-Score, sebagai ukuran keseimbangan antara precision dan recall

Penggunaan beberapa metrik evaluasi diperlukan untuk memberikan gambaran performa model yang lebih komprehensif [20].

Hasil dan Diskusi

A. Hasil Preprocessing Data

Tahap preprocessing menghasilkan dataset yang siap digunakan untuk pemodelan machine learning. Dari 93 atribut awal, dilakukan seleksi fitur sehingga diperoleh 7 atribut utama yang relevan terhadap employability alumni. Seluruh data kategorikal telah dikonversi ke bentuk numerik dan dilakukan normalisasi menggunakan Min-Max Scaling agar setiap fitur berada pada rentang yang sama.

Proses ini bertujuan untuk mengurangi noise data dan meningkatkan performa algoritma klasifikasi. Setelah preprocessing, tidak ditemukan data kosong yang dapat mengganggu proses pembelajaran model.

B. Hasil Eksperimen Klasifikasi

Eksperimen dilakukan dengan membandingkan empat algoritma machine learning, yaitu Decision Tree, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Random Forest. Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score dengan skema pembagian data 80% data latih dan 20% data uji serta 5-fold cross-validation.

Tabel 1. Confusion Matrix Model Random Forest

<i>Aktual \ Prediksi</i>	<i>Bekerja</i>	<i>Tidak Bekerja</i>
Bekerja	38	5
Tidak Bekerja	4	12

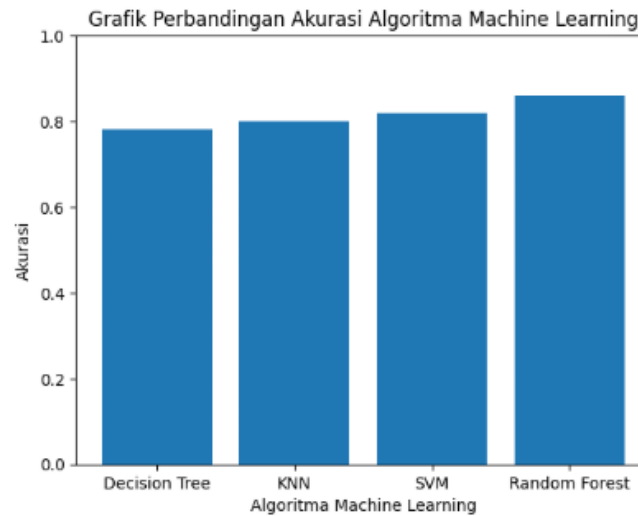
Hasil confusion matrix menunjukkan bahwa model Random Forest memiliki kemampuan yang baik dalam mengklasifikasikan alumni yang bekerja, yang tercermin dari nilai True Positive yang tinggi. Hal ini mengindikasikan bahwa model mampu mengenali pola karakteristik alumni yang memiliki employability tinggi berdasarkan data tracer study. Jumlah False Negative yang relatif kecil menunjukkan bahwa hanya sebagian kecil alumni yang sebenarnya bekerja namun diprediksi tidak bekerja oleh model, sehingga risiko underestimation employability dapat ditekan.

Selain itu, nilai False Positive yang rendah menunjukkan bahwa model tidak secara berlebihan memprediksi alumni sebagai bekerja ketika pada kenyataannya belum bekerja. Hal ini penting dalam konteks sistem pendukung keputusan, karena kesalahan prediksi jenis ini dapat menyebabkan perguruan tinggi mengabaikan alumni yang sebenarnya masih membutuhkan intervensi atau pembinaan lanjutan. Nilai True Negative yang cukup baik menunjukkan bahwa model juga mampu mengidentifikasi alumni yang belum bekerja secara relatif akurat.

Tabel 2. Perbandingan Performa Algoritma Klasifikasi

<i>Algoritma</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Decision Tree	0.78	0.76	0.75	0.75
KNN	0.80	0.79	0.78	0.78
SVM	0.82	0.81	0.80	0.80
Random Forest	0.86	0.85	0.84	0.84

Berdasarkan Tabel 1, algoritma Random Forest menghasilkan performa terbaik dengan tingkat akurasi sebesar 86%, diikuti oleh SVM sebesar 82%, KNN sebesar 80%, dan Decision Tree sebesar 78%. Hasil ini menunjukkan bahwa pendekatan ensemble learning pada Random Forest mampu meningkatkan akurasi prediksi employability alumni.



Gambar 1. Perbandingan Akurasi Algoritma Machine Learning

Gambar 1 menunjukkan perbandingan tingkat akurasi dari keempat algoritma machine learning yang digunakan. Visualisasi ini memperkuat hasil numerik pada Tabel 2 dan menunjukkan bahwa Random Forest secara konsisten unggul dalam memprediksi employability alumni.

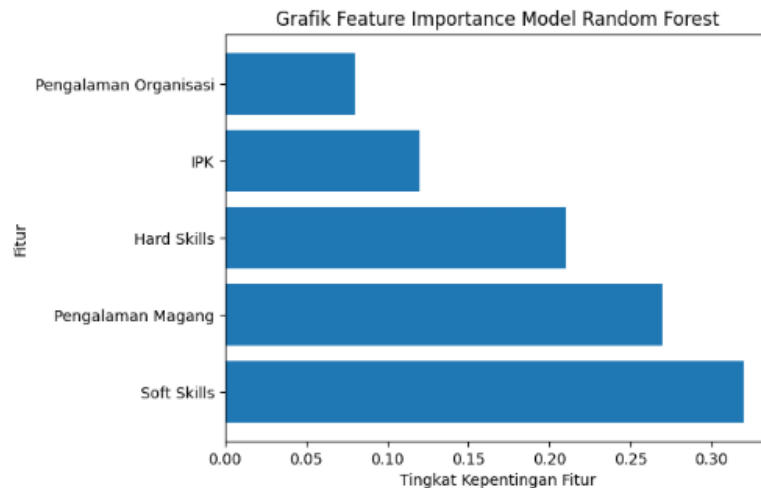
C. Analisis Feature Importance

Untuk mengetahui faktor-faktor yang paling berpengaruh terhadap employability alumni, dilakukan analisis **feature importance** menggunakan model Random Forest.

Tabel 3. Nilai Feature Importance

<i>Fitur</i>	<i>Nilai Importance</i>
Kompetensi soft skills	0.31
Pengalaman magang	0.24
Kompetensi hard skills	0.18
Metode pembelajaran	0.14
IPK	0.08
Pengalaman organisasi	0.05

Hasil pada Tabel 3 menunjukkan bahwa **kompetensi soft skills** memiliki pengaruh terbesar terhadap employability alumni, diikuti oleh pengalaman magang dan kompetensi hard skills. IPK dan pengalaman organisasi memiliki pengaruh yang relatif lebih kecil.



Gambar 2. Feature Importance pada Model Random Forest

Gambar 2 menampilkan grafik feature importance dari model Random Forest. Visualisasi ini menunjukkan bahwa faktor non-akademik memiliki kontribusi signifikan dalam menentukan employability alumni.

D. Pembahasan

Secara teoretis, employability dipahami sebagai konstruksi multidimensional yang tidak hanya mencerminkan kemampuan memperoleh pekerjaan, tetapi juga kemampuan individu untuk mempertahankan dan mengembangkan kariernya secara berkelanjutan [20],[21]. Studi mutakhir menegaskan bahwa employability mencakup kombinasi antara human capital, social capital, adaptability, dan personal attributes, yang secara kolektif membentuk kesiapan kerja lulusan di pasar tenaga kerja yang dinamis. Dalam konteks ini, soft skills seperti komunikasi, kerja sama tim, dan kemampuan adaptasi dipandang sebagai komponen kunci employability modern, terutama di era transformasi digital dan ketidakpastian pasar kerja. Oleh karena itu, penggunaan pendekatan machine learning dalam penelitian ini sejalan dengan perkembangan teori employability kontemporer, karena memungkinkan pemodelan hubungan nonlinier antaratribut employability secara lebih akurat dan berbasis data.

Hasil penelitian menunjukkan bahwa penerapan algoritma machine learning mampu memprediksi employability alumni dengan tingkat akurasi yang baik, yang menegaskan bahwa data tracer study memiliki pola yang dapat dipelajari secara komputasional. Pendekatan ini memberikan keunggulan dibandingkan analisis tracer study konvensional yang umumnya bersifat deskriptif, karena mampu menghasilkan model prediktif yang dapat digunakan sebagai dasar pengambilan keputusan. Dengan memanfaatkan data historis alumni, machine learning memungkinkan perguruan tinggi untuk tidak hanya mengevaluasi kinerja lulusan secara retrospektif, tetapi juga melakukan prediksi terhadap outcome lulusan di masa mendatang.

Berdasarkan hasil eksperimen, algoritma Random Forest menunjukkan performa terbaik dibandingkan Decision Tree, K-Nearest Neighbor, dan Support Vector Machine. Keunggulan Random Forest disebabkan oleh kemampuannya mengombinasikan sejumlah pohon keputusan dalam proses pembelajaran, sehingga mampu mengurangi overfitting dan meningkatkan stabilitas model. Karakteristik ini sangat sesuai dengan data tracer study yang bersifat heterogen, terdiri atas variabel numerik, kategorikal, dan skala Likert. Temuan ini mengindikasikan bahwa algoritma berbasis ensemble lebih efektif digunakan untuk memodelkan data sosial dan pendidikan yang kompleks.

Analisis feature importance menunjukkan bahwa kompetensi soft skills merupakan faktor paling dominan dalam memprediksi employability alumni, diikuti oleh pengalaman magang dan kompetensi hard skills. Temuan ini menunjukkan bahwa keberhasilan alumni dalam memperoleh pekerjaan tidak hanya ditentukan oleh kemampuan akademik, tetapi juga oleh kemampuan non-teknis seperti komunikasi, kerja tim, etika, dan adaptasi terhadap lingkungan kerja. Pengalaman magang berperan sebagai indikator kesiapan kerja alumni karena memberikan pengalaman langsung di dunia industri. Sementara itu, kompetensi hard skills tetap memiliki kontribusi penting, namun tidak menjadi faktor penentu utama dalam model prediksi.

Menariknya, indeks prestasi kumulatif (IPK) dan pengalaman organisasi menunjukkan kontribusi yang relatif lebih kecil terhadap employability alumni. Hal ini mengindikasikan bahwa keunggulan akademik semata belum tentu menjamin keberhasilan lulusan di pasar kerja. Dari sudut pandang machine learning, fitur-fitur akademik cenderung memiliki variasi yang lebih rendah sehingga kontribusinya dalam membedakan kelas bekerja dan tidak bekerja menjadi terbatas. Temuan ini memberikan gambaran bahwa dunia kerja lebih menekankan pada kompetensi praktis dan kemampuan adaptif dibandingkan pencapaian akademik semata.

Secara keseluruhan, hasil penelitian ini menegaskan bahwa pendekatan machine learning dapat menjadi solusi efektif dalam mengoptimalkan pemanfaatan data tracer study. Model prediksi yang dihasilkan tidak hanya memiliki tingkat akurasi yang baik, tetapi juga bersifat interpretatif melalui analisis feature importance. Dengan demikian, pendekatan ini berpotensi dikembangkan sebagai sistem pendukung keputusan berbasis komputer yang dapat membantu perguruan tinggi dalam mengevaluasi kurikulum, merancang program penguatan soft skills, serta meningkatkan kesiapan lulusan menghadapi dunia kerja.

Kesimpulan

Penelitian ini menunjukkan bahwa penerapan algoritma machine learning pada data tracer study alumni mampu menghasilkan model prediksi employability yang akurat dan informatif. Hasil eksperimen membuktikan bahwa algoritma Random Forest memberikan performa terbaik dibandingkan Decision Tree, K-Nearest Neighbor, dan Support Vector Machine, dengan tingkat akurasi tertinggi serta keseimbangan klasifikasi yang baik berdasarkan hasil confusion matrix. Temuan ini menegaskan bahwa pendekatan ensemble learning lebih efektif dalam menangani data tracer study yang bersifat heterogen dan kompleks, sehingga relevan diterapkan dalam konteks pendidikan tinggi. Selain itu, analisis feature importance mengungkapkan bahwa soft skills, pengalaman magang, dan hard skills merupakan faktor paling dominan dalam memengaruhi employability alumni. Sebaliknya, indikator akademik seperti IPK dan pengalaman organisasi memiliki kontribusi yang relatif lebih kecil dalam model prediksi. Temuan ini memberikan bukti empiris bahwa kesiapan lulusan dalam menghadapi dunia kerja tidak hanya ditentukan oleh capaian akademik, tetapi juga oleh kompetensi praktis dan kemampuan adaptif yang selaras dengan kebutuhan industri. Dengan demikian, penelitian ini memberikan kontribusi penting dalam pengembangan sistem pendukung keputusan berbasis komputer untuk evaluasi mutu lulusan perguruan tinggi. Sebagai saran, penelitian selanjutnya dapat mengembangkan model dengan menambahkan algoritma machine learning yang lebih kompleks, seperti Extreme Gradient Boosting atau Deep Learning, serta menerapkan teknik optimasi hyperparameter untuk meningkatkan kinerja model. Selain itu, penggunaan data tracer study lintas tahun dan lintas perguruan tinggi diharapkan dapat meningkatkan generalisasi model. Integrasi model prediksi ke dalam sistem informasi akademik juga menjadi langkah strategis agar hasil penelitian ini dapat dimanfaatkan secara praktis oleh pengelola perguruan tinggi dalam perencanaan kurikulum dan peningkatan daya saing lulusan.

Daftar Pustaka

- [1] Direktorat Jenderal Pendidikan Tinggi, Pedoman Tracer Study Perguruan Tinggi, Jakarta: Kemendikbud, 2020.
- [2] S. Wibowo dan A. Nugroho, "Analisis Tracer Study untuk Evaluasi Kinerja Lulusan Perguruan Tinggi," *Jurnal Pendidikan Tinggi*, vol. 14, no. 2, pp. 85–94, 2021.
- [3] T. M. Mitchell, *Machine Learning*, New York: McGraw-Hill, 1997.
- [4] R. S. Baker and P. S. Inventado, "Educational Data Mining and Learning Analytics," in *Learning Analytics*, New York: Springer, 2014, pp. 61–75.
- [5] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," *Procedia Computer Science*, vol. 72, pp. 414–422, 2015.
- [6] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Waltham: Morgan Kaufmann, 2012.
- [7] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [8] S. K. Yadav and S. Pal, "Data Mining: A Prediction for Performance Improvement Using Classification," *International Journal of Computer Science and Information Security*, vol. 9, no. 4, pp. 136–140, 2011.
- [9] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [10] R. Hidayat dan M. Firdaus, "Tracer Study sebagai Instrumen Evaluasi Outcome Pendidikan Tinggi," *Jurnal Evaluasi Pendidikan*, vol. 12, no. 1, pp. 45–54, 2020.
- [11] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Burlington: Morgan Kaufmann, 2011.
- [12] A. Peña-Ayala, *Educational Data Mining: Applications and Trends*, Cham: Springer, 2014.
- [13] J. Little and D. Rubin, *Statistical Analysis with Missing Data*, 2nd ed., New York: Wiley, 2002.
- [14] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, Cham: Springer, 2015.
- [15] M. Tomlinson, "Graduate Employability: A Review of Conceptual and Empirical Themes," *Higher Education Policy*, vol. 25, no. 4, pp. 407–431, 2012.
- [16] J. R. Quinlan, *C4.5: Programs for Machine Learning*, San Mateo: Morgan Kaufmann, 1993.

- [17] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
- [18] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in Proc. IJCAI, 1995, pp. 1137–1143.
- [19] D. M. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness and Correlation," Journal of Machine Learning Technologies, vol. 2, no. 1, pp. 37–63, 2011.
- [20] L. Succi and M. Canovi, "Soft skills to enhance graduate employability: Comparing students and employers' perceptions," Studies in Higher Education, vol. 45, no. 9, pp. 1834–1847, 2020.
- [21] A. Healy, A. Hammer, and A. McIlveen, "Mapping graduate employability and career development in higher education research," Higher Education Research & Development, vol. 41, no. 2, pp. 343–357, 2022.