JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1513-1522
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1398

Analisis Prediksi Kesesuaian Lulusan Dengan Dunia Kerja Menggunakan Model Ensemble *Random Forest* dan *XGBoost* Berdasarkan Data Tracer Study Universitas Royal

Rudi Hermawan[1*], Muhammad Iqbal[2], Muhammad Irfan Sarif[3]

[1,2,3]Pascasarjana, Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia
Email: [1*]rudirk60@gmail.com, [2]muhammadiqbal@dosen.pancabudi.ac.id, [3]irfanberbagi@gmail.com
(* Email Corresponding Author: rudirk60@gmail.com)
Received: 28 Juni 2026 | Revision: 4 Juli 2026 | Accepted: 7 Juli 2026

Abstrak

Transformasi digital mendorong institusi pendidikan tinggi untuk mengoptimalkan pengolahan data alumni guna meningkatkan mutu layanan dan efisiensi manajemen pendidikan. Universitas Royal memiliki data *tracer study* yang melimpah, namun pengelolaannya selama ini masih didominasi oleh pendekatan deskriptif konvensional sehingga potensi prediktifnya belum termanfaatkan secara maksimal. Penelitian ini bertujuan untuk memprediksi tingkat kesesuaian lulusan dengan dunia kerja menggunakan pendekatan *ensemble learning* yang menggabungkan algoritma *Random Forest* dan *XGBoost* melalui mekanisme *soft voting classifier*. Dataset awal diolah melalui tahap pembersihan data yang meliputi eliminasi label kosong, penghapusan data duplikat (deduplikasi), dan imputasi statistik, sehingga menghasilkan sebaran kelas target yang terdiri atas 759 data "Sesuai" (83,5%) dan 150 data "Tidak Sesuai" (16,5%). Proses prapemrosesan melibatkan imputasi median untuk data numerik, imputasi modus untuk data kategorikal, transformasi fitur dengan *one-hot encoding* pada variabel jurusan dan instansi, serta normalisasi menggunakan *Min-Max Scaler*. Dataset dengan total 909 data valid ini dibagi secara proporsional menjadi 727 data latih (80%) dan 182 data uji (20%). Hasil eksperimen menunjukkan bahwa model *ensemble voting* mencatatkan performa tertinggi dibandingkan model tunggal, dengan nilai akurasi sebesar 80,77%, presisi 85,45%, *recall* 92,76%, dan *F1-score* 88,96%. Sebagai pembanding, algoritma *Random Forest* tunggal memperoleh akurasi 78,57% dan *XGBoost* memperoleh 79,67%. Melalui analisis *feature importance*, ditemukan bahwa masa tunggu kerja merupakan faktor paling dominan dengan kontribusi sebesar 33,3%, diikuti oleh kompetensi (26,7%), Indeks Prestasi Kumulatif (24,6%), instansi tempat kerja (12,8%), dan program studi (2,6%). Pendekatan *ensemble* ini terbukti efektif dalam memberikan stabilitas generalisasi yang superior untuk memproyeksikan linearitas karier alumni serta mendukung pengambilan keputusan strategis institusi.

Kata Kunci: Tracer Study, One-Hot Encoding, Random Forest, XGBoost, Ensemble Learning

Abstract

Digital transformation drives higher education institutions to optimize the utilization of alumni tracking data to improve service quality and management efficiency. Universitas Royal possesses abundant tracer study data, but its management has been predominantly conventional and descriptive, leaving its predictive potential unexploited. This study aims to predict the alignment of graduates with the workforce using an ensemble learning approach that combines Random Forest and XGBoost algorithms via a soft voting classifier mechanism. After comprehensive data cleaning (eliminating empty target labels, deduplication, and statistical imputation), the target class distribution consists of 759 "Sesuai" records (83.5%) and 150 "Tidak Sesuai" records (16.5%). Preprocessing steps involved median imputation for numerical data, mode imputation for categorical data, feature transformation using one-hot encoding on major and institution variables, and normalization using Min-Max Scaler. The dataset, containing 909 valid records, was split with an 80:20 ratio into 727 training records and 182 testing records. Experimental results show that the ensemble voting model achieved the highest performance with an accuracy of 80.77%, precision of 85.45%, recall of 92.76%, and an F1-score of 88.96%. In comparison, the standalone Random Forest obtained an accuracy of 78.57% and XGBoost obtained 79.67%. Feature Importance analysis revealed that the job waiting period is the most dominant factor with a contribution of 33.3%, followed by competence (26.7%), Cumulative Grade Point Average (GPA) (24.6%), institution/type of work (12.8%), and study program (2.6%). This ensemble approach is proven effective in providing superior generalization stability for projecting alumni career linearity and supporting institutional strategic decision-making.

Keywords*: Tracer Study, One-Hot Encoding, Random Forest, XGBoost, Ensemble Learning*

1. PENDAHULUAN

Era globalisasi dan revolusi industri 4.0 membawa dampak transformasional yang signifikan terhadap seluruh sektor kehidupan, tidak terkecuali pada sektor pendidikan tinggi. Transformasi digital telah mendorong institusi pendidikan tinggi untuk secara agresif mengadopsi teknologi informasi dan kecerdasan buatan guna meningkatkan mutu layanan, efektivitas pembelajaran, dan efisiensi manajemen pendidikan [1]. Di tengah persaingan global yang kian ketat, perguruan tinggi tidak lagi hanya dituntut untuk menyelenggarakan administrasi akademik secara konvensional, melainkan wajib bertransformasi menjadi institusi yang digerakkan oleh data (*data-driven institution*) [2]. Dalam konteks

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1513-1522
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1398

ini, analitik data dan *machine learning* membuka peluang besar untuk melakukan pemrosesan data pendidikan secara lebih mendalam, komprehensif, dan prediktif. Pemanfaatan teknologi ini dapat membantu universitas dalammengidentifikasi pola risiko, memetakan tren linearitas, serta memprediksi kesesuaian lulusan dengan dunia kerja secara akurat [2].

Kemampuan memprediksi kesiapan dan relevansi lulusan dalam menghadapi dinamika dunia kerja kini menjadi salah satu indikator keberhasilan paling krusial bagi perguruan tinggi. Hal ini sejalan dengan tanggung jawab institusi dalam menghasilkan sumber daya manusia (SDM) yang kompeten, adaptif, dan memiliki daya saing tinggi sesuai dengan kebutuhan industri yang terus berubah [3]. Berdasarkan landasan Teori Human Capital, investasi dalam pendidikan tinggi harus mampu memberikan kontribusi nyata terhadap peningkatan produktivitas ekonomi melalui ketersediaan tenaga kerja ahli. Oleh karena itu, kesenjangan antara kompetensi yang diajarkan di bangku kuliah dengan kebutuhan nyata di sektor industri (*skills mismatch*) harus diminimalisasi [3][4]. Melalui pemanfaatan analitik prediktif, institusi pendidikan dapat mendeteksi pola tersembunyi, mengidentifikasi faktor prediktor dominan, serta memetakan risiko ketidaksesuaian lulusan yang mungkin tidak akan pernah terdeteksi apabila institusi hanya mengandalkan analisis statistik deskriptif konvensional [5].

Secara empiris, kesesuaian lulusan dengan dunia kerja dapat dipahami sebagai tingkat relevansi dan linearitas antara bidang studi atau kompetensi akademik yang ditempuh oleh mahasiswa selama perkuliahan dengan karakteristik pekerjaan yang mereka peroleh setelah menyelesaikan studi. Fenomena ini umumnya diukur, dipantau, dan dievaluasi secara berkala melalui instrumen pelacakan alumni yang dikenal sebagai *tracer study* perguruan tinggi. Data yang dihimpun melalui *tracer study* menyediakan informasi yang sangat kaya dan berharga, mencakup durasi masa tunggu kerja, profil penyerapan jenis instansi atau tempat kerja, tingkat pendapatan awal, relevansi bidang pekerjaan, hingga umpan balik objektif mengenai penguasaan kompetensi *hard skill* maupun *soft skill* alumni selama bekerja [5].

Meskipun signifikansi data hasil pelacakan alumni sudah disadari secara luas, implementasi pemanfaatannya di lapangan masih menghadapi kendala mendasar. Banyak institusi pendidikan tinggi masih menggunakan analisis data *tracer study* sebatas pada level deskriptif sederhana, seperti penyajian persentase kumulatif atau diagram batang. Pendekatan konvensional ini belum mampu mengoptimalkan potensi kekayaan data pelacakan tersebut untuk kebutuhan analisis prediktif jangka panjang maupun perencanaan strategis institusi [5]. Analisis deskriptif memiliki keterbatasan karena tidak mampu mengenali pola hubungan non-linear serta interaksi yang kompleks antarvariabel laten—seperti bagaimana interaksi multidimensi antara capaian Indeks Prestasi Kumulatif (IPK), tingkat penguasaan kompetensi praktis, dan durasi masa tunggu kerja memengaruhi probabilitas linearitas karier alumni di masa depan [6][7].

Kondisi serupa saat ini terjadi di Universitas Royal. Sebagai perguruan tinggi swasta yang berkembang pesat, Universitas Royal telah melaksanakan survei *tracer study* secara berkala. Namun, volume data alumni yang melimpah tersebut belum diolah secara mendalam untuk memprediksi kesesuaian lulusan dengan kebutuhan riil dunia kerja. Padahal, fenomena ketidaksesuaian kerja (*mismatch*) yang terjadi pada alumni memiliki dampak sistemik yang besar, mulai dari penurunan nilai akreditasi program studi, degradasi reputasi universitas di mata masyarakat, hingga ketidaksiapan lulusan dalam menghadapi pasar kerja yang semakin selektif [8]. Kondisi empiris ini menegaskan urgensi kebutuhan akan sebuah pendekatan analitik modern yang berbasis data (*Educational* Data Mining) guna memetakan, memahami, dan memproyeksikan faktor-faktor kritis yang memengaruhi kesesuaian kerja lulusan di Universitas Royal [9][10].

Dalam upaya membangun model prediksi yang andal, pemilihan algoritma *machine learning* menjadi faktor penentu. Karakteristik data *tracer study* umumnya bersifat multi-fitur, melibatkan kombinasi data numerik dan kategorikal, serta memiliki kecenderungan ketidakseimbangan kelas (*imbalanced data*). Oleh karena itu, penggunaan model klasifikasi tunggal (*single classifier*) sering kali menghasilkan bias prediksi yang tinggi atau mengalami *overfitting*. Untuk mengatasi batasan tersebut, diperlukan metode *ensemble learning* yang mampu menggabungkan kekuatan prediktif algoritma dasar. Model *ensemble* seperti *Random Forest* dinilai sangat sesuai karena menggunakan pendekatan *bagging* yang tangguh menangani data berdimensi tinggi dan menghasilkan prediksi stabil [11][12][13]. Di sisi lain, *XGBoost* (Extreme Gradient *Boosting*) merupakan algoritma berbasis *boosting* sekuensial yang sangat efisien, mampu menangani *missing values* otomatis, serta memiliki tingkat akurasi klasifikasi sangat tinggi [14][15].

Efektivitas penerapan *machine learning* dalam mengukur kesesuaian lulusan telah dibuktikan oleh beberapa peneliti terdahulu [16]. Ijlal et al. (2025) memanfaatkan data akademik (IPK) dan non-akademik (*softskill*) untuk memodelkan hubungan kausalitas kesesuaian karier [17]. Setiawan dan Rahmatulloh (2025) melaksanakan studi komparatif antara algoritma *Random Forest* dan *XGBoost* untuk memprediksi risiko turnover karyawan, menyimpulkan bahwa metode *ensemble machine learning* terbukti sangat efektif [18]. Ompusunggu et al. (2023) juga memperkuat bukti empiris bahwasanya metode *ensemble* modern memiliki keunggulan performa yang jauh lebih tinggi dan konsisten dalam mendukung proses pengambilan keputusan strategis [19].

Oleh karena itu, penelitian ini hadir untuk mengisi kesenjangan penelitian tersebut dengan mengusulkan pendekatan *Ensemble Learning* yang mengombinasikan *Random Forest* dan *XGBoost* melalui mekanisme *Soft Voting Classifier* [20][21][22]. Tujuannya adalah untuk mendongkrak performa metrik akurasi, presisi, recall, dan F1-score dalam memprediksi status kesesuaian lulusan Universitas Royal dengan dunia kerja profesional.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1513-1522
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1398

2. METODOLOGI PENELITIAN

2.1 Data Penelitian

Data yang digunakan dalam penelitian ini dikategorikan secara khusus sebagai objek kuantitatif yang bersumber dari data historis sekunder rekam jejak kelulusan alumni. Populasi dalam penelitian ini tergolong sebagai populasi terbatas (*finite population*), di mana jumlah objek atau elemen data terukur secara pasti dan dapat dihitung secara matematis. Objek populasi mencakup seluruh rekaman digital hasil pengisian kuesioner resmi *tracer study* Universitas Royal untuk rentang periode kelulusan lima tahun terakhir, terhitung mulai dari tahun 2021 hingga tahun 2025. Lokasi pengumpulan objek riset ini dilakukan pada lingkup administratif Universitas Royal yang bertempat di Kabupaten Asahan, Provinsi Sumatra Utara. Batasan waktu pelaksanaan pengerjaan keseluruhan riset berlangsung selama enam bulan, yang dimulai dari bulan Oktober 2025 hingga bulan Maret 2026. Dari proses ekstraksi pangkalan data (*database*) awal, diperoleh total populasi mentah sebanyak 2.125 data entries alumni.

Tabel 1. Dataset *Tracer Study*

No	Jurusan	Jenis Instansi	Masa Tunggu (Bulan)	IPK	Kompetensi	Kesesuaian
1	Sistem Informasi	Swasta	3	3,06	3,8571429	Sesuai
2	Sistem Informasi	Lainnya	3	3,7	3	Sesuai
3	Sistem Komputer	Wiraswasta	2	3,5	3,1428571	Sesuai
4	Sistem Komputer	Swasta	3	3,16	4,0714286	Sesuai
5	Sistem Komputer	-	1	3,18	2,7142857	-
6	Sistem Komputer	-	1	3,68	4,5	-
7	Sistem Komputer	Wiraswasta	2	3,59	4,7142857	Sesuai
8	Sistem Komputer	-	1	3,66	3,5714286	-
9	Sistem Komputer	-	1	3,59	4,0714286	-
10	Sistem Komputer	Swasta	1	3,17	3,4285714	Tidak Sesuai
…	…	…	…	…	…	…
2125	Sistem Informasi	Swasta	1	3,53	3,2142857	-

2.2 Teknik Pengumpulan Data

Teknik pengambilan sampel yang diaplikasikan dalam penelitian ini adalah teknik sampling jenuh (*total sampling*), di mana seluruh anggota populasi dilibatkan secara penuh sebagai subjek analisis guna meminimalkan kesalahan penyampelan (*sampling error*). Namun, guna membangun model prediktif yang valid dan reliabel, diterapkan kriteria inklusi data valid yang ketat, yaitu mengeliminasi baris data yang tidak memiliki label pada variabel target (kosong atau bernilai tidak valid karena alumni belum bekerja atau tidak mengisi bagian instrumen kesesuaian kerja). Akuisisi data sekunder ini diperoleh melalui kolaborasi resmi dengan unit Layanan Karir dan Pusat Data dan Informasi (Pusdatin) Universitas Royal. Setelah melewati proses pemfilteran berbasis kriteria inklusi tersebut, diperoleh volume dataset final bersih sebanyak 909 data valid yang siap dieksplorasi. Karakteristik sebaran kelas target pada dataset final ini menunjukkan fenomena ketimpangan kelas (*imbalanced data distribution*), terdiri atas 759 data alumni pada kelas "Sesuai" (83,5%) dan 150 data alumni pada kelas "Tidak Sesuai" (16,5%) [21].

2.3 Modul Penelitian

Modul pelaksanaan penelitian disusun secara terstruktur mengadopsi kerangka kerja standar pengembangan model pembelajaran mesin (*standard machine learning workflow*). Alur penelitian berjalan secara sekuensial melalui tahapan sebagai berikut: (1) Tahap identifikasi masalah penyerapan kerja alumni dan perumusan kesenjangan riset; (2) Tahap akuisisi data sekunder melalui Pusdatin; (3) Tahap prapemrosesan data (data *preprocessing*) yang memuat langkah pembersihan, imputasi nilai kosong, pengodean fitur, dan normalisasi skala; (4) Tahap pembagian subset data latih dan data uji secara terstratifikasi; (5) Tahap konstruksi dan penggabungan arsitektur model *ensemble* berbasis *base learners Random Forest* dan *XGBoost*; serta (6) Tahap pengujian validasi performa klasifikasi menggunakan pengujian matriks kontingensi *Confusion Matrix*.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1513-1522
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1398

2.4 Metode Analisis Data

Prosedur metode analisis data dijalankan secara komputasional memanfaatkan bahasa pemrograman Python. Tahap prapemrosesan data diawali dengan menangani nilai kosong *(missing values*) menggunakan imputasi statistik: nilai kosong pada fitur independen numerik diisi menggunakan nilai median untuk menangani pencilan (*outlier*), sedangkan fitur kategorikal nominal diisi menggunakan nilai modus. Proses deduplikasi juga dilakukan untuk menghapus baris data yang identik. Selanjutnya, fitur kategorikal nominal program studi (kode_jurusan) dan tempat bekerja (kode_instansi) ditransformasikan menggunakan teknik *One-Hot Encoding* untuk mengonversi kategori menjadi kolom biner independen baru berbobot 0 atau 1 tanpa memicu bias hierarki ordinal [18]. Seluruh fitur numerik kemudian dinormalisasi ke rentang interval [0.0, 1.0] menggunakan metode *Min-Max Scaler* dengan formulasi matematis sebagai berikut:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Dataset dibagi menggunakan teknik penyampelan terstratifikasi (*stratified sampling*) menjadi dua subset dengan rasio proporsional 80% sebagai data latih (*training set*) sebanyak 727 rekaman dan 20% sebagai data uji (*testing set*) sebanyak 182 rekaman. Definisi operasional dari variabel penelitian dirangkum secara rinci pada Tabel 2 di bawah ini:

Tabel 2. Definisi Operasional Variabel Dataset

No	Kategori	Nama Variabel	Kode Atribut	Tipe Data	Definisi & Skala Pengukuran
1	Independen (X)	1. Jurusan	kode_jurusan	Nominal	Program studi asal lulusan (Sistem Informasi / Sistem Komputer) - *One-Hot Encoded*.
2	Independen (X)	2. Jenis Instansi	kode_instansi	Nominal	Kategori tempat bekerja (Pemerintah, BUMN, Swasta, Non-Profit,Wirausaha, Lainnya) - *One-Hot Encoded*.
3	Independen (X)	3. Masa Tunggu	Masa Tunggu	Rasio	Durasi waktu dalam bulan mulai dari lulus hingga mendapat pekerjaan pertama (Skala 0,0 - 1,0).
4	Independen (X)	4. IPK	IPK	Rasio	Indeks Prestasi Kumulatif saat lulus (Skala 0,0 - 1,0).
5	Independen (X)	5. Skor Kompetensi	Kompetensi	Rasio	Rata-rata nilai kompetensi *hardskill* dan *softskill* lulusan (Skala 0,0 - 1,0).
6	Dependen (Y)	6. Kesesuaian	Kesesuaian	Nominal	Label klasifikasi relevansi pekerjaan dengan bidang studi (1: Sesuai, 0: Tidak Sesuai).

Model komputasi dibangun dengan menerapkan arsitektur kombinasi *Ensemble Soft Voting Classifier* yang menggabungkan dua algoritma dasar (*base learners*), yaitu *Random Forest Classifier* (200 *estimators*) dan *XGBoost Classifier* (200 *estimators*, *learning rate* 0.1). Model secara gabungan menghitung rata-rata probabilitas prediksi untuk menentukan label kelas keputusan akhir \hat{y} berdasarkan fungsi maksimasi argumentasi kelas berikut:

$$\hat{y} = argmax\left(\frac{\left(P_{(RF)}(y|x) + P_{(XGB)}(y|x)\right)}{2}\right) \quad (2)$$

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1513-1522
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1398

3. HASIL DAN PEMBAHASAN

3.1 Karakteristik dan Prapemrosesan Data

Eksperimen komputasi dan analisis data dalam penelitian ini memanfaatkan lingkungan *Google Colab* dengan dukungan penuh dari pustaka analisis data dan pembelajaran mesin tingkat lanjut seperti *scikit-learn* dan *XGBoost*. Fokus utama dari hasil penelitian ini adalah melakukan pemodelan prediktif berbasis *Educational* Data Mining untuk memetakan kesesuaian lulusan Universitas Royal dengan kebutuhan dunia kerja komersial. Data pelacakan alumni (tracer study) yang dikumpulkan untuk periode kelulusan lima tahun terakhir (2021-2025) mula-mula berjumlah 2.125 entries. Namun, melalui penyaringan data berbasis kriteria inklusi, baris-baris data yang memiliki nilai angka kesesuaian tidak valid atau kosong dieliminasi secara ketat. Eliminasi label kosong ini krusial dilakukan karena model klasifikasi terbimbing (*supervised learning*) memerlukan kepastian label target yang valid agar tidak menimbulkan misleading pada saat proses pembelajaran algoritma berlangsung.

Setelah melewati tahapan penyaringan inklusi data dan pembersihan baris data yang mengalami kerusakan, volume dataset final bersih yang digunakan untuk eksperimen adalah sebanyak 909 rekaman data alumni valid. Karakteristik sebaran kelas target pada dataset final ini menunjukkan fenomena ketimpangan kelas (*imbalanced* data *distribution*) yang cukup mencolok. Berdasarkan tabulasi data, ditemukan bahwa sebanyak 759 rekaman alumni (83,5%) tergolong ke dalam kelas "Sesuai" (label 1), di mana alumni bersangkutan melaporkan bahwa bidang pekerjaan yang mereka geluti saat ini memiliki tingkat relevansi yang tinggi dengan program studi asal mereka saat menempuh pendidikan di Universitas Royal. Sebaliknya, hanya terdapat 150 rekaman alumni (16,5%) yang masuk ke dalam kategori kelas "Tidak Sesuai" (label 0), yang mengindikasikan adanya fenomena *mismatch* di mana bidang pekerjaan alumni tidak selaras dengan latar belakang keilmuan akademis mereka. Ketimpangan sebaran kelas ini menjadi tantangan komputasi tersendiri karena model cenderung mudah mengenali pola kelas mayoritas (Sesuai) namun kesulitan mengidentifikasi pola tersembunyi pada kelas minoritas (Tidak Sesuai). Oleh karena itu, evaluasi model tidak boleh hanya bersandar pada nilai akurasi semata, melainkan wajib ditopang oleh metrik *presisi*, *recall*, dan *F1-score* yang sensitif terhadap ketimpangan kelas [23].

Tahap prapemrosesan data memegang peranan yang sangat vital dalam mentransformasikan data mentah hasil kuesioner tracer study menjadi matriks fitur berkualitas tinggi yang siap diproses oleh fungsi matematis algoritma *Random Forest* dan *XGBoost*. Langkah penanganan nilai kosong (*missing values*) berhasil dieksekusi dengan baik tanpa mengurangi volume baris data final. Untuk fitur numerik 'Masa Tunggu Kerja', di mana sebaran datanya menunjukkan kemencengan (*skewed distribution*) akibat adanya nilai-nilai ekstrem pencilan (*outlier*) seperti alumni yang membutuhkan waktu hingga 24 bulan untuk mendapat pekerjaan pertama, teknik imputasi menggunakan nilai median terbukti lebih kokoh (*robust*) dibandingkan nilai rata-rata (*mean*) karena tidak terpengaruh oleh pencilan tersebut. Nilai median yang diperoleh untuk masa tunggu kerja alumni Universitas Royal adalah sebesar 3 bulan, dan nilai inilah yang digunakan untuk mengisi baris data yang kosong. Begitu pula untuk variabel IPK dan Skor Kompetensi, nilai median digunakan sebagai parameter pengisian nilai yang hilang. Sementara itu, untuk fitur kategorikal nominal 'Jenis Instansi', nilai kosong diisi dengan menggunakan teknik imputasi modus, di mana nilai yang paling sering muncul adalah kategori "Swasta", yang mencerminkan profil penyerapan tenaga kerja alumni Universitas Royal didominasi oleh sektor swasta komersial.

Langkah prapemrosesan berikutnya yang memberikan pengaruh besar terhadap peningkatan performa model adalah transformasi fitur kategorikal nominal menggunakan teknik *One-Hot Encoding*. Variabel teks nominal seperti 'kode_jurusan' (Sistem Informasi dan Sistem Komputer) serta 'kode_instansi' (Pemerintah, BUMN, Swasta, Non-Profit, Wirausaha, dan Lainnya) dipecah menjadi kolom-kolom biner baru yang bersifat independen [24]. Sebagai contoh, variabel 'kode_jurusan' ditransformasikan menjadi dua kolom biner terpisah, di mana nilai 1 akan diberikan pada kolom jurusan yang sesuai dan nilai 0 pada kolom lainnya. Langkah ini sangat krusial dalam konteks pemodelan berbasis pohon keputusan (*tree-based models*) karena mencegah algoritma mengasumsikan nilai kategori teks tersebut memiliki urutan hierarki numerik (*ordinal*) yang keliru apabila hanya menggunakan Label Encoding biasa. Selain itu, guna mengantisipasi risiko kebocoran data (data *leakage*) yang dapat merusak validitas generalisasi, pencocokan parameter *encoder* secara ketat hanya dilakukan pada subset data pelatihan, sedangkan subset data pengujian hanya melakukan transformasi berdasarkan parameter yang telah terkunci dari data pelatihan tersebut. Terakhir, seluruh fitur numerik dinormalisasi menggunakan *Min-Max Scaler* ke dalam skala seragam 0,0 hingga 1,0. Proses normalisasi ini berhasil menyamakan bobot komputasi seluruh variabel independen sehingga fungsi loss pada model tidak didominasi oleh variabel yang memiliki rentang nilai absolut yang besar.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1513-1522
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1398

3.2 Evaluasi Kinerja Algoritma Model Tunggal

Sebelum menguji performa model kombinasi *ensemble*, penelitian ini terlebih dahulu melakukan pengujian performa secara mandiri terhadap dua algoritma klasifikasi tunggal (*single classifier*), yaitu *Random Forest Classifier* dan *XGBoost Classifier*. Pengujian ini bertujuan untuk menetapkan batas dasar (*baseline performance*) guna membuktikan apakah pendekatan kolaboratif *ensemble* benar-benar mampu memberikan peningkatan kinerja prediksi yang signifikan atau tidak. Kedua model tunggal dilatih menggunakan subset data latih (727 rekaman) dan diuji kinerjanya secara objektif menggunakan subset data uji mandiri (182 rekaman) yang disisihkan secara terpisah selama proses pelatihan model berlangsung.

Berdasarkan hasil pengujian pada subset data uji, algoritma tunggal *Random Forest Classifier* mencatatkan nilai Akurasi sebesar 78,57%. Dari segi metrik klasifikasi lainnya, *Random Forest* memperoleh nilai Presisi sebesar 85,09%, tingkat Sensitivitas atau Recall sebesar 90,13%, dan capaian nilai *F1-Score* sebesar 87,54%. Karakteristik *Random Forest* yang membangun ratusan pohon keputusan secara acak melalui mekanisme *bagging* terbukti sangat andal dalam mereduksi variansi *error* dan mencegah terjadinya *overfitting* pada data pelacakan alumni Universitas Royal yang bersifat multi-fitur. Di sisi lain, algoritma tunggal *XGBoost Classifier* (*Extreme Gradient Boosting*) mencatatkan performa numerik yang sedikit lebih tinggi dibandingkan *Random Forest*, dengan perolehan nilai Akurasi sebesar 79,67%. Algoritma *boosting* sekuensial ini memperoleh nilai Presisi sebesar 85,28%, tingkat *Recall* sebesar 91,45%, dan capaian *F1-Score* sebesar 88,25%. Keunggulan *XGBoost* ini didorong oleh kemampuannya yang sangat efisien dalam mengoptimalkan fungsi loss melalui ekspansi deret *Taylor* orde kedua serta penerapan regularisasi internal yang ketat (L1 dan L2 *regularization*) untuk mengendalikan kompleksitas model di tengah ketimpangan distribusi kelas target. Meskipun kedua model tunggal ini sudah menunjukkan kinerja klasifikasi yang cukup memuaskan dengan nilai akurasi mendekati batas 80%, namun masing-masing model tunggal masih memiliki kerentanan terhadap variansi *error* spesifik akibat pengondisian hyperparameter bawaan.

3.3 Evaluasi Kinerja Arsitektur Ensemble Soft Voting

Untuk meminimalkan kelemahan individu dari masing-masing algoritma tunggal dan memaksimalkan stabilitas generalisasi prediksi, penelitian ini mengintegrasikan *Random Forest* dan *XGBoost* ke dalam satu kesatuan sistem menggunakan metode *Ensemble Soft Voting Classifier*. Model kombinasi *ensemble* ini dilatih pada data latih yang sama dan diuji menggunakan 182 data pengujian independen. Mekanisme *soft voting* bekerja dengan cara merata-ratakan probabilitas prediksi kelas yang dikeluarkan oleh *Random Forest* dan *XGBoost* untuk setiap sampel data uji, kemudian menetapkan keputusan label final berdasarkan nilai probabilitas rata-rata yang tertinggi. Hasil eksperimen komputasi membuktikan secara empiris bahwa arsitektur *Ensemble Voting* berhasil menorehkan tingkat performa tertinggi dan paling tangguh dibandingkan kedua model tunggal pembandingnya [20][21].

Model *ensemble* ini sukses mencatatkan nilai Akurasi pengujian sebesar 80,77%, yang merepresentasikan peningkatan performa yang konsisten. Keunggulan mutakhir dari taktik kolaboratif *ensemble* ini terlihat jelas pada peningkatan seluruh metrik evaluasi klasifikasinya, di mana nilai Presisi meningkat mencapai 85,45%, nilai *Recall* atau tingkat sensitivitas deteksi kelas mencapai 92,76%, dan capaian metrik *F1-Score* yang sangat kokoh sebesar 88,96% [21]. Keberhasilan arsitektur *ensemble* dalam mendongkrak nilai *F1-score* menjadi 88,96% merupakan indikator teoretis yang sangat kuat bahwa model ini memiliki ketepatan pemisahan data yang sangat seimbang dan presisi, bahkan di tengah kondisi ketimpangan distribusi data alumni (*imbalanced dataset*) yang menjadi karakteristik bawaan dari survei tracer study Universitas Royal. Kombinasi antara pendekatan bagging (*Random Forest*) yang fokus pada urutan pengurangan variansi error dan pendekatan boosting (*XGBoost*) yang fokus pada pengurangan bias error terbukti secara ilmiah saling melengkapi dan menghasilkan efek sinergi komputasi yang superior [19][21]. Ringkasan perbandingan hasil metrik performa model dirangkum pada Tabel 3 dan gambar 1 di bawah ini:

Tabel 3. Hasil Perbandingan Metrik Performa Model Terhadap Data Uji

No.	Algoritma Model	Akurasi (Accuracy)	Presisi (Precision)	Recall (Sensitivity)	F1-Score
1	*Random Forest* Classifier	78,57%	85,09%	90,13%	87,54%
2	*XGBoost* Classifier	79,67%	85,28%	91,45%	88,25%
3	**Ensemble Voting (RF + *XGBoost*)**	**80,77%**	**85,45%**	**92,76%**	**88,96%**

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1513-1522
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1398

Gambar 1. Perbandingan Akurasi Model

3.4 Analisis Matriks Kontingensi (*Confusion Matrix*)

Guna membedah kemampuan klasifikasi model *Ensemble Voting* secara lebih transparan, akurat, dan mendalam pada level data uji (182 rekaman), hasil prediksi dianalisis menggunakan matriks kontingensi *Confusion Matrix*. Pemetaan *Confusion Matrix* membagi keputusan prediksi model ke dalam empat komponen esensial, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Analisis komponen ini sangat krusial dalam konteks manajemen pendidikan tinggi untuk mengukur jenis kesalahan prediksi apa yang paling sering dilakukan oleh model dan bagaimana implikasinya terhadap pengambilan keputusan strategis.

Berdasarkan hasil pemetaan *Confusion Matrix* terhadap 182 data uji mandiri, ditemukan sebaran komponen sebagai berikut: (1) *True Positive* (TP) mencatatkan angka sebanyak 141 rekaman. Hal ini menunjukkan bahwa terdapat 141 alumni yang secara aktual kondisi kariernya adalah "Sesuai" dengan bidang studinya dan model ensemble berhasil mengklasifikasikannya secara tepat ke dalam label kelas "Sesuai". Angka TP yang tinggi ini mengonfirmasi kekuatan model dalam mengenali karakteristik dominan alumni yang memiliki linearitas karier yang baik. (2) *True Negative* (TN) mencatatkan angka sebanyak 6 rekaman. Komponen ini merepresentasikan data alumni yang secara nyata kondisi pekerjaannya adalah "Tidak Sesuai" (*mismatch*) dan model *ensemble* berhasil mengidentifikasikannya secara benar ke dalam label kelas "Tidak Sesuai". (3) *False Positive* (FP) mencatatkan angka sebanyak 24 rekaman. Komponen ini merupakan kesalahan klasifikasi tipe I (*Error Type* I), di mana data alumni sebenarnya berstatus aktual "Tidak Sesuai" namun model *ensemble* keliru memproyeksikannya sebagai "Sesuai". (4) *False Negative* (FN) mencatatkan angka sebanyak 11 rekaman. Komponen ini merepresentasikan kesalahan klasifikasi tipe II (*Error Type* II), di mana kondisi nyata pekerjaan alumni sebenarnya adalah "Sesuai" namun sistem memprediksinya ke dalam kelas "Tidak Sesuai". Berikut ini hasil Visualisasi *Confusion Matrix* dari 3 Model dapat dilihat pada gambar 2.

Gambar 2. Visualisasi Confusion Matrix 3 Model

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1513-1522
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1398

Berdasarkan nilai sebaran komponen *Confusion Matrix* di atas, pembuktian metrik pengujian evaluasi matematis dari model *Ensemble Voting* dapat dihitung dan dijabarkan secara runtut menggunakan formulasi standar sebagai berikut:

a. Akurasi (*Accuracy*) = $\frac{TP+TN}{TP+TN+FP+FN} = \frac{147}{182} = 80{,}77\ \%$

b. Presisi (*Precision*) = $\frac{TP}{TP+FP} = \frac{141}{165} = 85{,}45\ \%$

c. *Recall* (*Sensitivitas*) = $\frac{TP}{TP+FN} = \frac{141}{152} = 92{,}76\ \%$

d. *F1-Score* = 2 x $\frac{Precision\ x\ Recall}{Precision+\ Recall} = 88{,}98\ \%$

3.5 Analisis Kepentingan Fitur (*Feature Importance*)

Kelebihan mendasar dari penggunaan model berbasis pohon keputusan (*Random Forest* dan *XGBoost*) dalam penelitian ini adalah sifatnya yang dapat diinterpretasikan secara logis (*interpretable machine learning*). Model ini mampu mengurai kompleksitas algoritma yang sering kali dianggap sebagai kotak hitam (*black box*) menjadi nilai bobot parameter yang transparan bagi pemangku kebijakan institusi melalui analisis matriks *Feature Importance*. Distribusi signifikansi kontribusi dari masing-masing fitur prediktor dalam menentukan status kesesuaian kerja lulusan Universitas Royal menunjukkan pola yang sangat menarik dan berharga untuk ditindaklanjuti.

Berdasarkan hasil ekstraksi nilai *Feature Importance* dari model final ensemble, ditemukan bahwa variabel 'Masa Tunggu Kerja' menempati posisi paling dominan dengan bobot pengaruh sebesar 33,3% terhadap keputusan klasifikasi kesesuaian kerja [25]. Posisi kedua ditempati oleh fitur 'Skor Kompetensi Lulusan' dengan kontribusi pengaruh sebesar 26,7%. Selanjutnya, fitur akademik 'Indeks Prestasi Kumulatif (IPK)' menempati peringkat ketiga dengan bobot pengaruh sebesar 24,6%. Sementara itu, fitur deskriptif tempat bekerja yaitu 'Jenis Instansi' memberikan kontribusi pengaruh sebesar 12,8%, dan fitur latar belakang keilmuan 'Program Studi' atau Jurusan berada pada posisi terakhir dengan bobot pengaruh terkecil sebesar 2,6%. Rincian nilai ini dirangkum pada Tabel 4:

Tabel 4. Distribusi Bobot Nilai Kepentingan Fitur (*Feature Importance*)

Atribut Fitur Prediktor	Bobot Pengaruh (%)
Masa Tunggu Kerja	**33,30%**
Skor Kompetensi Lulusan	26,70%
Indeks Prestasi Kumulatif (IPK)	24,60%
Jenis Instansi / Tempat Kerja	12,80%
Program Studi (Jurusan)	2,60%

4. KESIMPULAN

Penelitian ini berhasil membuktikan efektivitas dan keandalan penerapan metode *Ensemble Learning* dengan mengintegrasikan algoritma *Random Forest Classifier* dan *XGBoost Classifier* melalui mekanisme *Soft Voting Classifier* dan prapemrosesan transformasi *One-Hot Encoding* dalam memprediksi tingkat kesesuaian lulusan Universitas Royal dengan dunia kerja. Melalui pendekatan *Educational* Data Mining ini, data historis hasil pelacakan alumni (*tracer study*) Universitas Royal yang mulanya hanya dimanfaatkan sebatas instrumen pelaporan administratif deskriptif konvensional, kini berhasil ditransformasikan secara mutakhir menjadi basis pemodelan sistem peringatan dini (*early warning system*) pendukung keputusan kualitas lulusan yang akurat dan ilmiah. Arsitektur model kombinasi ensemble sukses mencatatkan tingkat performa klasifikasi tertinggi dan paling tangguh pada subset data pengujian mandiri dengan perolehan nilai Akurasi sebesar 80,77%, nilai Presisi sebesar 85,45%, tingkat Sensitivitas atau Recall sebesar 92,76%, dan capaian metrik *F1-Score* yang sangat kokoh sebesar 88,96% di tengah tantangan ketimpangan distribusi sebaran kelas target asli. Analisis kepentingan fitur (*Feature Importance*) secara transparan menempatkan variabel durasi Masa Tunggu Kerja sebagai faktor prediktor paling dominan dengan kontribusi pengaruh sebesar 33,3% terhadap status relevansi pekerjaan alumni, disusul berturut-turut oleh Skor Kompetensi Lulusan sebesar 26,7%, capaian akademik Indeks Prestasi Kumulatif (IPK) sebesar 24,6%, Jenis Instansi tempat bekerja sebesar 12,8%, dan rumpun Program Studi asal sebesar 2,6%. Temuan ini memberikan konfirmasi empiris yang kuat terhadap keselarasan Teori *Human Capital*, Teori *Person-Job Fit*, dan Teori *Signaling* dalam penyerapan tenaga kerja profesional berpendidikan tinggi. Rekomendasi praktis ditujukan bagi jajaran pengelola dan pengambil keputusan di Universitas Royal untuk segera mengimplementasikan Sistem Pendukung Keputusan (SPK) berbasis dasbor *web* terintegrasi guna memberikan intervensi dini bagi mahasiswa semester akhir, serta mengakselerasi program sinkronisasi kurikulum berkala (*link and match*) yang difokuskan pada penguatan sertifikasi

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1513-1522
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1398

kompetensi industri terapan terakreditasi. Penelitian selanjutnya diharapkan dapat memperluas cakupan dimensi variabel dengan melibatkan indikator eksternal seperti rekam jejak keikutsertaan program Magang Merdeka (MBKM) serta menerapkan teknik *Hyperparameter Tuning* tingkat lanjut seperti *Grid Search* guna menyempurnakan sensitivitas model.

REFERENCES

[1] R. S. Tanjung, E. S. Tanjung, Khoyan, and A. Nasution, "Transformasi Digital dalam Administrasi dan Manajemen Pendidikan di Perguruan Tinggi," *J. Pembelajaran dan Ilmu Civ.*, vol. 11, no. 1, pp. 37–41, 2025, doi: 10.36987/civitas.v11i1.7061.

[2] V. Y. D. Wijaya and G. Brotosaputro, "Penerapan Data Mining Dalam Prediksi Kinerja Akademik Mahasiswa Menggunakan Algoritma Machine Learning," *J. Infomedia Tek. Inform. Multimedia, dan Jar.*, vol. 10, no. 2, pp. 134–142, 2025, doi: 10.30811/jim.v10i2.7643.

[3] M. R. Suryawijaya, S. Praptodiyono, and S. N. A'kaasyah, "Peran Kecerdasan Buatan dalam Mengembangkan Kompetensi dan Meningkatkan Motivasi Belajar Mahasiswa," *Inform. J. Ilmu Komput.*, vol. 21, no. 2, pp. 155–165, Aug. 2025, doi: 10.52958/iftk.v21i2.11115.

[4] M. Kumar, N. Singh, J. Wadhwa, P. Singh, G. Kumar, and A. Qtaishat, "Utilizing Random Forest and XGBoost Data Mining Algorithms for Anticipating Students' Academic Performance," *Int. J. Mod. Educ. Comput. Sci.*, vol. 16, no. 2, pp. 29–44, 2024, doi: 10.5815/ijmecs.2024.02.03.

[5] H. Septiana, S. Chabibah, T. Indarti, U. N. Surabaya, and U. N. Surabaya, "Refleksi Kesesuaian Kompetensi Lulusan Prodi S-1 Pendidikan Bahasa Dan Sastra Indonesia Di Universitas Negeri Surabaya," *J. Pendidik. dan Teknol. Pembelajaran*, vol. 07, no. 01, pp. 1–14, 2026, doi: 10.37859/eduteach.v7i1.9177.

[6] G. S. Palupi, A. A. Utami, and I. K. D. Nuryana, "Assessing Graduate Competency Fit for the Workplace: A Tracer Study Investigation in Education," *IJORER Int. J. Recent Educ. Res.*, vol. 5, no. 2, pp. 292–304, 2024, doi: 10.46245/ijorer.v5i2.438.

[7] S. Lina, M. Sitio, Y. Samudra, T. Informatika, S. Inggris, and U. Pamulang, "Comparative Analysis Of Bagging And Boosting Models In Ensemble Learning For Graduation Prediction," *J. Ilmu Pengetah. Dan Teknol. Komput.*, vol. 11, no. 3, pp. 979–986, 2026, doi: 10.33480/jitk.v11i3.7579.

[8] P. K. Tee, L. C. Wong, M. Dada, B. L. Song, and C. P. Ng, "Demand for digital skills, skill gaps and graduate employability: Evidence from employers in Malaysia," *F1000Research*, vol. 13, pp. 1–19, 2024, doi: 10.12688/f1000research.148514.1.

[9] M. Mashuri and S. Assegaff, "Sistem Informasi Alumni (Tracer Study) Berbasis Web Pada SMA Negeri 2 Bungo," *J. Manaj. Sist. Inf.*, vol. 8, no. 3, pp. 449–460, 2023, doi: 10.33998/jurnalmsi.2023.8.3.1481.

[10] Z. Sitorus, S. Pranoto, and S. Sutiono, "Comparison of Accuracy between Naïve Bayes and Decision Tree Methods for Property Tax (PBB-P2) Compliance in Tebing Tinggi City," *J. Inf. Technol. Comput. Sci. Electr. Eng.*, vol. 1, no. 2, pp. 121–128, 2024, doi: 10.61306/jitcse.v1i2.

[11] R. Ratnasari, A. J. Wahidin, and T. H. Andika, "Deteksi Dini Stunting Pada Anak Berdasarkan Indikator Antropometri dengan Menggunakan Algoritma Machine Learning," *J. Algoritm.*, vol. 21, no. 2, pp. 378–387, 2024, doi: 10.33364/algoritma/v.21-2.2122.

[12] S. Huang, "Processing and Comparison of GBoost, XGBoost, and Random Forest in Titanic Survival Prediction," *Appl. Comput. Eng.*, vol. 102, no. 1, pp. 175–182, 2024, doi: 10.54254/2755-2721/102/20241195.

[13] M. Rasyid, Z. Sitorus, R. Farta Wijaya, and M. Iqbal, "Machine Learning Analysis in Improving the Efficiency of the Student Admission Decision Making Process New At Panca Budi Medan Development University," *Bull. Eng. Sci. Technol. Ind.*, vol. 2, no. 3, pp. 216–225, 2024, doi: 10.59733/besti.v2i3.62.

[14] Rahmanul Hoque, Suman Das, Mahmudul Hoque, and Mahmudul Hoque, "Breast Cancer Classification using XGBoost," *World J. Adv. Res. Rev.*, vol. 21, no. 2, pp. 1985–1994, 2024, doi: 10.30574/wjarr.2024.21.2.0625.

[15] R. Y. Hayuningtyas, Wina Yusnaeni, and Ida Darwati, "Penerapan Algoritma XGBoost Dalam Menganalisa Keberlanjutan Pelanggan Tour dan Travel," *Evolusi J. Sains dan Manaj.*, vol. 13, no. 2, pp. 25–32, 2025, doi: 10.31294/evolusi.v13i2.9748.

[16] R. Sari and W. Kurniadi, "Penerapan Metode Machine Learning (Random Forest & XGBoost) untuk Memprediksi Prestasi Akademik Berdasarkan Faktor Internal dan Eksternal Siswa," *Bul. Sist. Inf. dan Teknol. Islam*, vol. 6, no. 4, pp. 205–213, 2025, doi: 10.33096/busiti.v6i4.3205.

[17] M. Y. L. Ijlal, A. Setiawan, and D. L. Fithri, "Prediction of Graduate Career Relevance Based on Academic and Non-Academic Aspects using Machine Learning," *SISTEMASI*, vol. 14, no. 4, p. 1952, Jul. 2025, doi: 10.32520/stmsi.v14i4.5364.

[18] M. Bayu Setiawan and A. Rahmatulloh, "Analisis Perbandingan Model Random Forest dan XGBoost dalam Memprediksi Turnover Karyawan," *Just IT J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 15, no. 2, pp. 393–400, 2025, doi: 10.24853/justit.15.2.393%20–%20400.

[19] M. K. S. Elvis Sastra Ompusunggu, Aleksander Nainggolan, "Penentuan kelayakan promosi pegawai menggunakan algoritma random forest classifier dan xgboost classifier," *J. TEKINKOM*, vol. 6, pp. 773–783, 2023, doi: 10.37600/tekinkom.v6i2.949.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1513-1522
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1398

[20] G. M. Momole, "Perbandingan Naïve Bayes dan Random Forest Dalam Klasifikasi Bahasa Daerah," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 855–863, 2022, doi: 10.35957/jatisi.v9i2.1857.

[21] T. Prayitno, N. Nasution, and S. Susandri, "Optimization and Comparative Analysis of C4 . 5 , Random Forest and XGBoost for Mapping SMK Muhammadiyah 2 Pekanbaru Alumni Career Pathways," *JAISEN J. Adv. Inf. Syst. Eng.*, vol. 1, pp. 42–48, 2026, [Online]. Available: https://journal.unilak.ac.id/index.php/jaisen/article/view/32471.

[22] Muit Sunjaya, Zulham Sitorus, Khairul, Muhammad Iqbal, and A.P.U Siahaan, "Analysis of machine learning approaches to determine online shopping ratings using naïve bayes and svm," *Int. J. Comput. Sci. Math. Eng.*, vol. 3, no. 1, pp. 7–16, 2024, doi: 10.61306/ijecom.v3i1.60.

[23] A. Saleh, N. Dharshinni, D. Perangin-Angin, F. Azmi, and M. I. Sarif, "Implementation of Recommendation Systems in Determining Learning Strategies Using the Naïve Bayes Classifier Algorithm," *Sinkron*, vol. 8, no. 1, pp. 256–267, 2023, doi: 10.33395/sinkron.v8i1.11954.

[24] R. Haque, A. Quek, C. Y. Ting, H. N. Goh, and M. R. Hasan, "Classification Techniques Using Machine Learning for Graduate Student Employability Predictions," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 14, no. 1, pp. 45–56, 2024, doi: 10.18517/ijaseit.14.1.19549.

[25] Umar, I. H. Al Ghozali, and A. R. Handoko, "Analisis Efektifitas Feature Selection dalam Pengkayaan Machine Learning untuk Deteksi Dini Risiko Putus Kuliah Mahasiswa," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 11 No. 1, no. 1, pp. 149–157, 2025, doi: 10.26418/jp.v11i1.90362.