

Deteksi Potensi Karir Mahasiswa Universitas Airlangga Berbasis Pemodelan Machine Learning

Mohammad Nathiq Ulman, Faisal Fahmi*

Sains Informasi dan Kepustakaan, Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Airlangga

Email: mohammad.nathiq.ulman-2024@fisip.unair.ac.id, faisalfahmi@fisip.unair.ac.id

Diajukan: 27-02-2026

Direvisi: 16-04-2026

Disetujui: 14-09-2026

Diterbitkan: 30-09-2026

Post-graduation transition is a critical issue in higher education because alumni may enter employment, pursue further education, or choose entrepreneurship, while institutional predictive support is often limited. Using Universitas Airlangga alumni tracer study data in tabular form, this study develops and evaluates a multiclass classification model for predicting alumni employment status and frames the task as institutional analytics rather than a purely technical benchmark. The target variable is grouped into three classes: Employed (full-time/part-time) / previously employed, Pursuing further education, and Entrepreneur / previously entrepreneur. After data cleaning, the final dataset consisted of 25,391 observations. The modeling workflow applies XGBoost with an integrated preprocessing pipeline consisting of data cleaning, numerical missing-value imputation, numerical feature standardization, and one-hot encoding for categorical variables. The data are split using an 80:20 stratified train-test scheme, while class imbalance in the training data is addressed using SMOTE. Performance is evaluated through class-wise precision, recall, F1-score, and confusion matrix analysis; model interpretability is examined using SHAP. The model achieves an overall accuracy of approximately 0.80, but class performance is uneven: Employed performs relatively well (precision 0.89; recall 0.87; F1-score 0.88), Pursuing further education is moderate (precision 0.62; recall 0.60; F1-score 0.61), and Entrepreneur remains weak (precision 0.21; recall 0.29; F1-score 0.25). These findings show that tracer study modeling must prioritize class-wise evaluation, class imbalance analysis, and interpretability, because aggregate accuracy alone can overstate practical readiness for institutional use.

Keywords: alumni employment status, alumni tracer study, class imbalance, model interpretability, multiclass classification, SHAP, XGBoost

Abstrak. Transisi pasca-kelulusan merupakan persoalan penting dalam tata kelola pendidikan tinggi karena alumni dapat memasuki pasar kerja, melanjutkan pendidikan, atau menempuh jalur kewirausahaan, sementara dukungan prediktif kelembagaan belum selalu tersedia secara memadai. Penelitian ini menggunakan data tracer study alumni Universitas Airlangga berbentuk tabular untuk mengembangkan dan mengevaluasi model klasifikasi multikelas guna memprediksi status kerja alumni, serta menempatkan persoalan tersebut sebagai analitika kelembagaan dan bukan semata uji algoritma. Variabel target dikelompokkan ke dalam tiga kelas, yaitu Bekerja (full time/part time) / pernah bekerja, Melanjutkan pendidikan, dan Wiraswasta / pernah berwirausaha. Setelah pembersihan data, dataset akhir berjumlah 25.391 observasi. Pemodelan menggunakan XGBoost dengan preprocessing pipeline yang mengintegrasikan pembersihan data, imputasi nilai hilang numerik, standarisasi fitur numerik, dan one-hot encoding untuk fitur kategorikal. Data dibagi dengan skema stratified split 80:20, sedangkan ketidakseimbangan kelas pada data latih ditangani menggunakan SMOTE. Evaluasi dilakukan melalui precision, recall, F1-score per kelas, serta matriks konfusi; interpretabilitas model dianalisis menggunakan SHAP. Model mencapai akurasi sekitar 0,80, tetapi kinerjanya timpang antar kelas: kelas Bekerja relatif kuat (precision 0,89; recall 0,87; F1-score 0,88), kelas Melanjutkan pendidikan berada pada tingkat moderat (precision 0,62; recall 0,60; F1-score 0,61), sedangkan

kelas Wiraswasta masih lemah (precision 0,21; recall 0,29; F1-score 0,25). Temuan ini menegaskan bahwa evaluasi model pada data tracer study harus menempatkan metrik per kelas, analisis ketidakseimbangan kelas, dan interpretabilitas sebagai komponen utama, karena akurasi agregat saja dapat menutupi keterbatasan substantif model.

Kata Kunci: *interpretabilitas model, ketidakseimbangan kelas, klasifikasi multikelas, SHAP, status kerja alumni, tracer study alumni, XGBoost*

1. Pendahuluan

Transisi lulusan dari pendidikan tinggi menuju bekerja, melanjutkan studi, atau menempuh jalur kewirausahaan merupakan salah satu indikator penting bagi evaluasi mutu perguruan tinggi, relevansi kurikulum, dan efektivitas layanan karier. Dalam praktik kelembagaan, *tracer study* telah lama digunakan untuk memantau luaran lulusan, tetapi pemanfaatannya masih sering berhenti pada pelaporan deskriptif mengenai distribusi status kerja, masa tunggu, dan profil responden. Studi *tracer* deskriptif tetap memiliki nilai penting sebagai dasar evaluasi luaran lulusan, tetapi orientasi tersebut belum cukup untuk menopang pengambilan keputusan kelembagaan yang menuntut pembacaan diferensial atas variasi jalur pasca-kelulusan (Albina & Sumagaysay, 2020; Ramos, 2024). Karena itu, persoalan utama saat ini bukan lagi apakah *tracer study* berguna, melainkan bagaimana data tersebut dapat dibaca ulang sebagai basis analitika yang lebih prediktif.

Perkembangan *tracer study* yang lebih mutakhir menunjukkan bahwa instrumen ini semakin sering digunakan bukan hanya untuk pelacakan administratif, tetapi juga untuk menilai relevansi program, kualitas lulusan, dan keterhubungan pendidikan dengan pasar kerja. Pada sejumlah konteks, *tracer study* telah dipakai untuk membaca *employability* lulusan, pertumbuhan profesional, serta hubungan antara kurikulum dan keberhasilan kerja (Garcia et al., 2024; Dzomeku et al., 2024; Munthali et al., 2025). Meski demikian, sebagian besar pemanfaatannya masih tetap bersifat evaluatif-deskriptif. Artinya, *tracer study* sudah kuat sebagai instrumen *monitoring* kelembagaan, tetapi belum selalu dikembangkan menjadi perangkat prediksi yang mampu membantu institusi membaca kemungkinan jalur pasca-kelulusan secara lebih sistematis.

Di sisi lain, penelitian mengenai luaran lulusan telah bergerak dari pendekatan statistik konvensional menuju *machine learning*. Pergeseran ini cukup masuk akal karena data lulusan umumnya bersifat tabular, heterogen, serta memuat campuran fitur numerik dan kategorikal yang berpotensi membentuk relasi nonlinier. Kajian sintesis menunjukkan bahwa medan *employability prediction* ditandai oleh keragaman definisi *outcome*, variasi sumber data, perbedaan rancangan eksperimen, dan ketidakseragaman metrik evaluasi (Mezhoudi et al., 2023). Dengan demikian, persoalan dalam prediksi luaran lulusan tidak dapat direduksi menjadi urusan memilih algoritma yang paling akurat, sebab definisi *outcome*, kualitas data, dan cara membaca hasil sama-sama menentukan kualitas analisis.

Sejumlah penelitian terapan telah menunjukkan bahwa *machine learning* memang dapat digunakan untuk memprediksi *employability* lulusan. Pada konteks yang berbeda-beda, penelitian terdahulu menunjukkan pentingnya penanganan ketidakseimbangan kelas, sensitivitas performa terhadap karakter data, dan pengaruh rancangan evaluasi terhadap kesimpulan akhir model (Abdulloh et al., 2022; ElSharkawy et al., 2022; Haque et al., 2024; Mansor & Othman, 2024). Akan tetapi, sebagian besar studi tersebut masih beroperasi pada horizon *outcome* yang biner atau sangat teragregasi. Akibatnya, perbedaan substantif antara jalur bekerja, studi lanjut, dan kewirausahaan sering kali belum dibaca sebagai persoalan klasifikasi yang benar-benar berbeda, melainkan masih diperlakukan sebagai variasi dari satu kategori *employability* yang lebih umum.

Keterbatasan literatur terdahulu tampak setidaknya dalam empat hal. Pertama, mayoritas studi masih memformulasikan *outcome* sebagai persoalan biner, sehingga heterogenitas jalur pasca-kelulusan belum tertangkap secara memadai. Kedua, banyak penelitian menempatkan pemilihan model terbaik berdasarkan metrik agregat sebagai tujuan utama, sementara ketimpangan performa antarkelas belum

dianalisis secara serius. Ketiga, studi yang mengintegrasikan *explainable AI* umumnya berhenti pada identifikasi fitur penting secara umum, bukan pada hubungan antara kontribusi fitur, struktur salah-klasifikasi, dan kegagalan model mengenali kelas tertentu. Keempat, jalur kewirausahaan masih relatif jarang dibahas secara mendalam, padahal justru jalur ini sering menjadi kelas yang paling sulit dipisahkan dalam konteks prediksi *employability* dan *tracer study* (Haque et al., 2025; Azis et al., 2026). Karena itu, ruang kosong penelitian bukan sekadar pada penggunaan algoritma yang berbeda, melainkan pada cara masalah penelitian itu sendiri dirumuskan.

Persoalan tersebut menjadi sangat relevan dalam konteks *tracer study* institusional. Data *tracer study* tidak hanya merekam apakah alumni bekerja atau tidak, tetapi juga menangkap beberapa jalur pasca-kelulusan yang berbeda secara substantif, termasuk melanjutkan pendidikan dan berwirausaha. Dalam situasi seperti ini, pembacaan performa model melalui satu metrik agregat berisiko menyesatkan karena model dapat tampak sangat baik terutama akibat kemampuannya mengenali kelas mayoritas. Literatur mengenai *imbalanced learning* menegaskan bahwa distribusi kelas yang timpang memengaruhi proses pembelajaran model dan dapat menyebabkan *under-detection* pada kelas minoritas apabila evaluasi tidak diarahkan pada *precision*, *recall*, *F1-score* per kelas, *macro average*, *weighted average*, dan matriks konfusi (He & Garcia, 2009; Krawczyk, 2016). Dengan kata lain, yang dipertaruhkan di sini bukan hanya kualitas algoritma, tetapi juga validitas cara kita membaca kualitas model itu sendiri.

Berdasarkan kondisi tersebut, penelitian ini mengembangkan model klasifikasi multikelas untuk memprediksi status kerja alumni Universitas Airlangga ke dalam tiga kelas, yaitu Bekerja (*full time/part time*) / pernah bekerja, Melanjutkan pendidikan, dan Wiraswasta / pernah berwirausaha. XGBoost dipilih sebagai model utama karena sesuai untuk data tabular yang memuat relasi nonlinier dan interaksi fitur yang kompleks, sedangkan SHAP digunakan untuk menjelaskan kontribusi fitur terhadap prediksi model secara konsisten pada tingkat global maupun lokal (Chen & Guestrin, 2016; Lundberg & Lee, 2017). Berbeda dari banyak penelitian terdahulu yang berfokus pada *binary employability prediction* atau pada pemilihan model dengan metrik agregat tertinggi, penelitian ini menempatkan *tracer study* sebagai persoalan analitika kelembagaan multikelas. Oleh karena itu, perhatian utama penelitian tidak berhenti pada akurasi total, tetapi diarahkan pada metrik per kelas, struktur matriks konfusi, dan pembacaan kontribusi fitur per kelas.

Kebaruan penelitian ini tidak terletak semata pada penggunaan XGBoost atau SHAP, karena kedua perangkat tersebut telah mapan dalam literatur *machine learning*. Kebaruan penelitian ini terletak pada integrasi data *tracer study* institusional, formulasi *outcome* sebagai klasifikasi multikelas jalur pasca-kelulusan alumni, evaluasi yang sensitif terhadap ketidakseimbangan kelas, serta interpretabilitas model per kelas untuk menjelaskan arah salah-klasifikasi dan keterbatasan daya pisah fitur (Abdulloh et al., 2022; Azis et al., 2026). Dalam horizon tersebut, *tracer study* tidak lagi diposisikan semata sebagai instrumen pelaporan administratif. Sebaliknya, data *tracer study* dibaca sebagai infrastruktur data kelembagaan yang dapat mendukung evaluasi dan pengembangan layanan karier secara lebih berbasis bukti.

Rumusan masalah penelitian ini mencakup tiga hal. Pertama, seberapa memadai performa model XGBoost pada tingkat per kelas dalam memprediksi status kerja alumni. Kedua, bagaimana ketidakseimbangan distribusi target memengaruhi pola kesalahan prediksi yang tercermin dalam matriks konfusi. Ketiga, fitur apa yang dominan mendorong atau menahan prediksi pada tiap kelas menurut SHAP. Sejalan dengan itu, tujuan penelitian ini adalah membangun model prediksi status kerja alumni, mengevaluasi ketimpangan performa antarkelas, dan menjelaskan perilaku model pada tingkat fitur secara interpretatif. Artikel ini disusun sebagai berikut: Bagian 2 membahas tinjauan pustaka, Bagian 3 menjelaskan metode penelitian, Bagian 4 menyajikan hasil dan pembahasan, dan Bagian 5 memuat kesimpulan serta saran pengembangan.

2. Tinjauan Pustaka

2.1 Tracer study sebagai infrastruktur analitika kelembagaan

Tracer study merupakan instrumen yang secara luas digunakan untuk memantau transisi lulusan dari perguruan tinggi menuju dunia kerja, studi lanjut, atau jalur aktivitas pasca-kampus lainnya. Dalam tata kelola pendidikan tinggi, *tracer study* memiliki fungsi evaluatif karena menyediakan informasi mengenai keterkaitan antara proses pendidikan, kompetensi lulusan, dan luaran aktual setelah kelulusan. Pada banyak konteks, *tracer study* telah digunakan untuk memetakan status kerja, relevansi kurikulum, dan karakteristik lulusan, tetapi penggunaannya masih lebih dominan untuk kebutuhan pelaporan deskriptif daripada untuk analitika prediktif yang sistematis (Albina & Sumagaysay, 2020; Garcia et al., 2024). Dengan demikian, nilai kelembagaan *tracer study* memang sudah diakui, tetapi potensinya sebagai basis pengembangan model prediktif masih belum sepenuhnya dimanfaatkan.

Dalam perspektif analitika kelembagaan, nilai strategis *tracer study* terletak pada kemampuannya menghubungkan data akademik, data institusional, dan data luaran lulusan ke dalam satu kerangka evaluasi yang lebih menyeluruh. Ketika data *tracer study* diperlakukan hanya sebagai perangkat administratif, potensi informasional yang dikandungnya tetap terfragmentasi. Sebaliknya, ketika diperlakukan sebagai basis bukti untuk analitika prediktif, *tracer study* dapat digunakan untuk memetakan pola transisi lulusan, mengidentifikasi kelompok *outcome* yang relatif sulit diprediksi, serta membuka ruang bagi intervensi kelembagaan yang lebih tepat sasaran. Arah penggunaan seperti ini tampak semakin penting dalam studi-studi *tracer* yang menekankan hubungan antara relevansi program, keberhasilan kerja, dan kualitas luaran lulusan (Ramos, 2024; Munthali et al., 2025).

2.2 Prediksi *employability* dan luaran lulusan berbasis *machine learning*

Literatur mengenai *employability prediction* menunjukkan pergeseran yang jelas dari penggunaan pendekatan statistik konvensional ke penggunaan *machine learning*. Pergeseran ini didorong oleh kebutuhan untuk menangani relasi nonlinier, interaksi fitur, serta kombinasi atribut akademik, demografis, institusional, dan sosial-ekonomi yang semakin kompleks. Di dalam medan ini, persoalan utama tidak lagi dapat direduksi menjadi sekadar pemilihan algoritma dengan akurasi tertinggi, karena definisi *outcome*, kualitas data, dan rancangan evaluasi ikut menentukan mutu kesimpulan yang dihasilkan. Kajian sintesis oleh Mezhoudi et al. (2023) menegaskan bahwa penelitian *employability prediction* ditandai oleh keragaman sumber data, tujuan analisis, dan metrik penilaian.

Pada konteks yang lebih terapan, sejumlah penelitian menunjukkan bahwa data lulusan memang dapat dimodelkan secara prediktif, tetapi hasilnya sangat dipengaruhi oleh struktur data dan desain eksperimen yang dipilih. Studi-studi terdahulu memperlihatkan bahwa model berbasis pohon cukup kompetitif pada data tabular, sementara perbedaan konteks kelembagaan dan strategi evaluasi dapat menghasilkan performa yang sangat berbeda antar penelitian (ElSharkawy et al., 2022; Haque et al., 2024; Mansor & Othman, 2024). Hal ini menunjukkan bahwa keberhasilan model tidak dapat dipahami sebagai sifat universal dari algoritma tertentu. Sebaliknya, performa model selalu harus dibaca dalam hubungan dengan jenis data, definisi *outcome*, dan cara hasil itu dievaluasi.

Perkembangan yang lebih baru juga memperlihatkan upaya memperluas medan riset ke model dan arsitektur yang lebih canggih. Beberapa studi menekankan pentingnya optimasi *ensemble*, sebagian lain menekankan konteks institusional dan regional, sementara sebagian yang lebih baru mulai memanfaatkan *embedding*, *feature selection*, atau model bahasa besar untuk meningkatkan kualitas klasifikasi. Meski demikian, sebagian besar studi tersebut tetap beroperasi dalam horizon *employability* sebagai *outcome* biner atau sangat teragregasi, bukan sebagai klasifikasi multikelas jalur pasca-kelulusan alumni yang berbeda secara substantif (Dake, 2025; Mahdin et al., 2025; Haque et al., 2025). Di sinilah terlihat bahwa perluasan teknis belum selalu diikuti oleh perluasan cara merumuskan masalah analitiknya.

2.3 Ketidakseimbangan kelas dan kebutuhan evaluasi per kelas

Salah satu persoalan metodologis paling penting dalam prediksi luaran lulusan adalah ketidakseimbangan distribusi kelas. Dalam banyak dataset *tracer study*, kelas bekerja cenderung mendominasi, sedangkan jalur seperti studi lanjut dan terutama kewirausahaan muncul sebagai kelas minoritas. Dalam situasi seperti ini, model dapat tampak sangat baik pada tingkat agregat hanya karena kemampuannya mengenali kelas mayoritas, sementara performa terhadap kelas minoritas tetap lemah. Literatur *imbalanced learning* menunjukkan bahwa kondisi semacam ini bukan sekadar gangguan teknis kecil, melainkan masalah struktural dalam proses pembelajaran model (He & Garcia, 2009; Krawczyk, 2016).

Implikasi langsungnya adalah bahwa evaluasi model tidak dapat bertumpu pada *accuracy* semata. Evaluasi yang lebih memadai harus mencakup *precision*, *recall*, dan *F1-score* per kelas, *macro average*, *weighted average*, serta matriks konfusi. Dalam konteks *tracer study*, pendekatan ini penting karena tujuan analisis bukan hanya mengidentifikasi mayoritas lulusan yang bekerja, tetapi juga membaca bagaimana model memperlakukan jalur yang lebih jarang muncul namun secara kelembagaan penting, seperti studi lanjut dan kewirausahaan. Karena itu, evaluasi per kelas dalam penelitian ini bukan tambahan opsional, melainkan konsekuensi metodologis dari sifat data yang memang timpang (Abdulloh et al., 2022).

2.4 XGBoost dan SHAP untuk klasifikasi multikelas data tabular

XGBoost merupakan salah satu algoritma *ensemble* berbasis *gradient boosting* yang dirancang untuk memberikan performa tinggi pada data tabular melalui efisiensi komputasional, regularisasi, dan kapasitas dalam menangani relasi nonlinier serta interaksi fitur yang kompleks. Pada konteks *tracer study*, keunggulan ini relevan karena fitur-fitur akademik, institusional, dan pembiayaan tidak selalu beroperasi secara linear atau independen satu sama lain. Selain itu, XGBoost kompatibel dengan representasi hasil *one-hot encoding*, sehingga sesuai untuk dataset yang memadukan fitur numerik dan kategorikal (Chen & Guestrin, 2016). Dengan kata lain, pemilihan XGBoost dalam penelitian ini bukan pilihan yang bersifat arbitrer, melainkan selaras dengan karakter data yang digunakan.

Meskipun demikian, performa prediktif yang tinggi tidak dengan sendirinya membuat model layak digunakan untuk kepentingan kelembagaan. Di sinilah interpretabilitas memperoleh arti penting, karena model yang kuat tetapi tidak dapat dijelaskan akan sulit digunakan untuk pembacaan institusional yang bertanggung jawab. SHAP menyediakan kerangka berbasis nilai Shapley untuk menjelaskan kontribusi fitur terhadap output model secara konsisten, termasuk pada model berbasis pohon seperti XGBoost (Lundberg & Lee, 2017). Dalam penelitian ini, SHAP digunakan bukan untuk menarik inferensi kausal, melainkan untuk membaca bagaimana model membangun prediksi dan mengapa kesalahan tertentu muncul pada kelas tertentu.

2.5 Posisi penelitian, gap, dan kebaruan

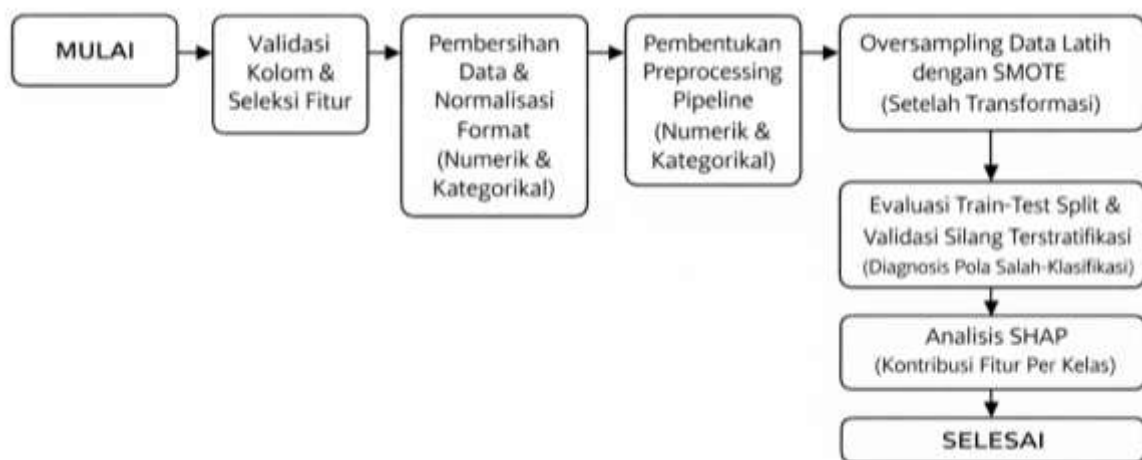
Berdasarkan komparasi literatur di atas, terdapat beberapa celah penelitian yang masih terbuka. Mayoritas studi terdahulu masih memodelkan luaran lulusan sebagai *binary employability* atau *binary employment outcome*, sehingga belum cukup menangkap heterogenitas jalur pasca-kelulusan yang lebih realistis dan lebih relevan bagi perguruan tinggi. Di saat yang sama, banyak penelitian masih berorientasi pada pemilihan model terbaik berdasarkan metrik agregat, sedangkan ketimpangan performa antarkelas belum ditempatkan sebagai pusat evaluasi. Keterbatasan ini membuat persoalan *tracer study* sering direduksi menjadi problem teknis klasifikasi, padahal secara substantif ia berkaitan dengan pembacaan diferensial atas beberapa jalur transisi lulusan yang berbeda. Kelemahan tersebut juga tampak pada studi-studi yang sudah bergerak ke arah *explainable AI*, tetapi masih berhenti pada identifikasi fitur penting secara umum dan belum cukup menghubungkan kontribusi fitur dengan struktur salah-klasifikasi antarkelas (Haque et al., 2025; Azis et al., 2026).

Penelitian ini menempati posisi pada irisan antara analitika *tracer study*, klasifikasi multikelas, evaluasi sensitif terhadap ketidakseimbangan kelas, dan interpretabilitas model. Kebaruannya tidak terletak semata pada penggunaan XGBoost atau SHAP, melainkan pada integrasi keduanya ke dalam

kerangka analitika kelembagaan multikelas yang memprediksi tiga jalur pasca-kelulusan alumni, yakni bekerja, melanjutkan pendidikan, dan berwirausaha, serta menilai model melalui metrik per kelas, *macro-based performance*, matriks konfusi, dan pembacaan kontribusi fitur per kelas. Dengan demikian, artikel ini mengisi celah antara *tracer study* sebagai pelaporan administratif dan *tracer study* sebagai infrastruktur analitika kelembagaan yang prediktif, interpretatif, dan relevan bagi pengembangan kebijakan internal.

3. Metode Penelitian

Bagian ini menjelaskan prosedur pemodelan *machine learning* untuk klasifikasi multikelas status_kerja alumni berbasis data tracer study tabular. Fokus metode tidak hanya pada pemilihan algoritma, tetapi pada konsistensi pra-pemrosesan, integrasi pipeline, penanganan ketidakseimbangan kelas pada data latih, evaluasi berbasis metrik per kelas, serta pembacaan interpretabilitas model. Alur kerja lengkap diringkas pada Gambar 1.



Gambar 1. Diagram alur pemodelan

Secara ringkas, tahapan yang dilakukan adalah: (1) validasi kolom dan seleksi fitur, (2) pembersihan data dan normalisasi format, (3) pembentukan preprocessing pipeline untuk fitur numerik dan kategorikal, (4) over-sampling pada data latih menggunakan SMOTE setelah fitur ditransformasikan, (5) pelatihan XGBoost, (6) evaluasi pada skema train-test split dan validasi silang terstratifikasi (untuk diagnosis pola salah-klasifikasi), dan (7) analisis SHAP untuk membaca kontribusi fitur per kelas.

3.1 Desain Penelitian dan Alur Pemodelan

Penelitian menggunakan pendekatan kuantitatif berbasis supervised learning untuk memprediksi status_kerja alumni dalam tiga kelas (WO, ST, WU). Pemodelan disusun sebagai pipeline terintegrasi agar proses transformasi fitur dan pelatihan model konsisten antara fase pelatihan, validasi, dan inferensi. Evaluasi dilakukan pada data uji (hold-out test set) serta dilengkapi prediksi validasi silang terstratifikasi pada data latih untuk membaca kestabilan pola kesalahan model pada tingkat kelas.

3.2 Sumber Data, Unit Analisis, dan Variabel

Data penelitian berasal dari dataset *tracer study* alumni Universitas Airlangga yang dihimpun melalui instrumen *tracer study* institusional dan kemudian diekspor ke format data tabular untuk keperluan analisis. Unit analisis adalah individu alumni, dengan satu baris data merepresentasikan satu alumni. Variabel yang tersedia mencakup identitas (misalnya nama, NIM), karakteristik akademik (tahun_lulus, prodi, fakultas, ipk, sks, elpt_akhir), informasi tambahan *tracer study* (termasuk deskripsi/judul tugas akhir), karakteristik pembiayaan (sumber_dana_kuliah, penghasilan_ortu, sumber_biaya, kelompok_biaya), serta variabel target status_kerja. Dalam pemodelan, kolom identitas

dan variabel yang berpotensi menimbulkan kebocoran informasi tidak digunakan sebagai prediktor agar evaluasi lebih mencerminkan kemampuan generalisasi model.

3.3 Pra-pemrosesan Data

Pra-pemrosesan dilakukan untuk memastikan kualitas dan konsistensi input. Tahapan meliputi: (1) validasi keberadaan kolom yang dibutuhkan, (2) penghapusan duplikasi, (3) penghapusan observasi tanpa label target, (4) pembersihan teks pada kolom kategorikal dengan normalisasi spasi (termasuk karakter non-breaking space) dan trimming, serta penyetaraan penanda missing (mis. string kosong, "nan", "None", "-") menjadi nilai hilang, (5) konversi kolom numerik menjadi tipe numerik; nilai yang tidak valid diperlakukan sebagai missing, lalu diimputasi menggunakan median kolom, dan (6) penghapusan observasi yang tidak memiliki nilai pada kolom kategorikal yang digunakan sebagai prediktor, agar one-hot encoding dan pembelajaran model berjalan stabil.

3.4 Representasi Fitur dan Preprocessing Pipeline

Karena dataset terdiri dari fitur numerik dan kategorikal, digunakan ColumnTransformer untuk membangun preprocessing yang eksplisit dan reproduisibel. Fitur numerik (tahun_lulus, ipk, sks, elpt_akhir, penghasilan_ortu) distandardisasi menggunakan StandardScaler. Fitur kategorikal (prodi, fakultas, sumber_dana_kuliah, sumber_biaya, kelompok_biaya) direpresentasikan dengan OneHotEncoder(handle_unknown='ignore'). Representasi one-hot diatur berbentuk dense agar kompatibel dengan SMOTE yang diterapkan setelah transformasi fitur. Parameter remainder='drop' memastikan hanya fitur yang ditetapkan yang masuk ke model.

3.5 Pengodean Target dan Pembagian Data

Label target status_kerja dikodekan menjadi bilangan bulat menggunakan LabelEncoder untuk kebutuhan pembelajaran multikelas, dan hasil prediksi dikembalikan ke label asli saat pelaporan agar interpretasi substantif tetap jelas. Data dibagi menjadi data latih dan data uji dengan skema stratified train-test split 80:20 (random_state=42) untuk mempertahankan proporsi kelas pada kedua subset.

3.6 Konfigurasi Model XGBoost

Model yang digunakan adalah XGBClassifier untuk klasifikasi multikelas dengan objective='multi:softprob' dan eval_metric='mlogloss'. Parameter utama meliputi n_estimators=400, max_depth=6, learning_rate=0.05, subsample=0.9, dan colsample_bytree=0.9, dengan random_state=42. Konfigurasi *softprob* dipilih agar model menghasilkan probabilitas per kelas, sehingga evaluasi dan interpretasi dapat memanfaatkan struktur probabilistik prediksi bila diperlukan pada tahap lanjutan. Dalam penelitian ini, konfigurasi parameter tersebut diperlakukan sebagai konfigurasi kerja utama, bukan sebagai hasil *benchmarking* komparatif antarmodel. Penelitian ini belum melakukan perbandingan antarmodel pada dataset yang sama; karena itu, pemilihan XGBoost dalam penelitian ini dibaca terutama sebagai keputusan metodologis berbasis kesesuaian terhadap karakter data serta konsistensi dengan temuan literatur komparatif yang relevan.

3.7 Pelatihan dan Evaluasi Performa

Pelatihan dilakukan menggunakan pipeline berurutan: preprocessing (ColumnTransformer) → SMOTE → XGBoost. SMOTE diterapkan hanya pada data latih (di dalam pipeline) untuk meningkatkan representasi kelas minoritas pada ruang fitur hasil transformasi, sehingga model tidak hanya belajar dari distribusi asli yang timpang. Evaluasi dilakukan melalui dua jalur. Pertama, prediksi validasi silang terstratifikasi (StratifiedKFold, 5 fold) pada data latih menggunakan cross_val_predict untuk memperoleh gambaran pola salah-klasifikasi yang lebih stabil dibanding satu kali split. Kedua, model dilatih pada seluruh data latih, kemudian dievaluasi pada data uji menggunakan classification report (precision, recall, F1-score, dan support per kelas) serta confusion matrix (bentuk hitungan dan normalisasi per kelas aktual). Visualisasi distribusi kelas pada subset latih dan uji juga disertakan untuk memastikan konsistensi stratifikasi dan konteks pembacaan metrik per kelas.

3.8 Analisis Interpretabilitas Menggunakan SHAP

Setelah model terlatih, interpretabilitas dianalisis menggunakan SHAP TreeExplainer karena XGBoost merupakan model berbasis pohon. Data uji ditransformasikan dengan preprocessing yang sama seperti saat pelatihan. Nama fitur hasil transformasi diambil dari `get_feature_names_out()` (atau fallback nama generik bila tidak tersedia) agar keluaran SHAP dapat ditelusuri. Analisis SHAP dilakukan pada seluruh data uji yang telah ditransformasikan; nilai SHAP positif diartikan mendorong prediksi ke kelas yang sedang dianalisis, sedangkan nilai SHAP negatif menurunkannya. Visualisasi yang dihasilkan mencakup summary plot (beeswarm) dan bar plot (mean absolute SHAP) untuk setiap kelas, sehingga pola kontribusi fitur dapat dibandingkan antar kelas.

Untuk memastikan keterlacakan eksperimen, pipeline terlatih dan label encoder disimpan dalam satu berkas model (joblib) agar proses prediksi ulang konsisten. Hasil evaluasi dan visualisasi (distribusi kelas, confusion matrix, serta plot SHAP) disimpan sebagai artefak keluaran untuk mendukung audit metodologis dan pelaporan.

4. Hasil dan Pembahasan

Bagian ini menyajikan hasil evaluasi model klasifikasi multikelas untuk `status_kerja` alumni dalam tiga kelas, yaitu Bekerja (*full time/part time*) / pernah bekerja, Melanjutkan pendidikan, dan Wiraswasta / pernah wiraswasta. Untuk menjaga keterbacaan pembahasan, ketiga kelas tersebut selanjutnya disingkat sebagai WO (*work*), ST (*study*), dan WU (*wiraswasta*). Pembahasan difokuskan pada dua hal utama, yaitu evaluasi kinerja model melalui metrik klasifikasi dan matriks konfusi, serta interpretabilitas model melalui SHAP. Struktur ini dipilih karena pada data dengan distribusi target yang timpang, akurasi agregat saja tidak cukup untuk menggambarkan kualitas model secara substantif.

4.1 Evaluasi Kinerja Model

Setelah proses pembersihan data, dataset akhir yang digunakan dalam penelitian ini berjumlah 25.391 observasi. Variabel target `status_kerja` menunjukkan distribusi yang timpang, dengan dominasi kuat kelas WO sebanyak 20.824 observasi atau 82,01% dari keseluruhan data. Dua kelas lainnya, yaitu ST dan WU, masing-masing berjumlah 2.868 observasi atau 11,30% dan 1.699 observasi atau 6,69%. Dengan skema *stratified train-test split* 80:20, data dibagi menjadi 20.312 observasi untuk pelatihan dan 5.079 observasi untuk pengujian, sehingga proporsi kelas pada kedua subset tetap terjaga. Struktur ini menunjukkan bahwa persoalan yang dihadapi model sejak awal bukan hanya klasifikasi multikelas, tetapi juga ketidakseimbangan distribusi target yang dapat memengaruhi performa apabila evaluasi hanya bertumpu pada satu metrik agregat.

Tabel 1. Distribusi Kelas Target (`status_kerja`)

<i>Kelas status_kerja</i>	<i>Jumlah Observasi</i>	<i>Persentase (%)</i>
Bekerja (<i>full time/part time</i>) / pernah bekerja	20.824	82,01
Melanjutkan pendidikan	2.868	11,30
Wiraswasta / pernah wiraswasta	1.699	6,69
Total	25.391	100,00

Berdasarkan Tabel 1, terlihat bahwa kelas WO merupakan kelas mayoritas yang sangat dominan, sedangkan kelas ST dan terutama WU merupakan kelas minoritas. Kondisi ini penting secara metodologis karena model cenderung lebih mudah mempelajari pola dari kelas mayoritas dibandingkan kelas minoritas. Oleh karena itu, evaluasi model perlu menempatkan perhatian pada *precision*, *recall*, dan *F1-score* per kelas, serta bukan hanya pada akurasi total.

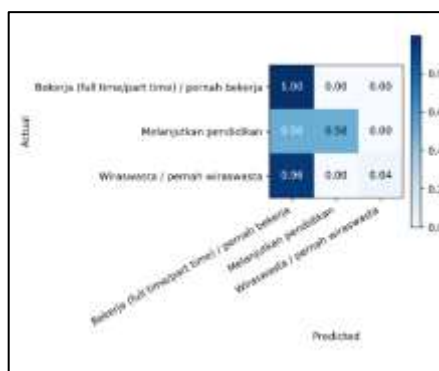
Evaluasi pada data uji menunjukkan bahwa model XGBoost mencapai akurasi sekitar 0,80. Secara agregat, nilai ini menunjukkan bahwa model memiliki kapasitas klasifikasi yang cukup baik. Namun, pembacaan yang lebih rinci memperlihatkan bahwa performa tersebut tidak merata antar kelas. Pada WO, model menghasilkan *precision* 0,89, *recall* 0,87, dan *F1-score* 0,88. Pada ST, model menghasilkan *precision* 0,62, *recall* 0,60, dan *F1-score* 0,61. Performa terendah muncul pada WU, dengan *precision* 0,21, *recall* 0,29, dan *F1-score* 0,25. Perbedaan antara *macro average* dan *weighted average* menunjukkan bahwa performa agregat masih sangat dipengaruhi oleh keberhasilan model pada kelas mayoritas. Dengan demikian, akurasi total tidak dapat dijadikan indikator tunggal kualitas model.

Tabel 2. Hasil Laporan Klasifikasi Model XGBoost pada Data Uji

Kelas / Ringkasan	Precision	Recall	F1-score	Support
Bekerja (full time/part time) / pernah bekerja	0,89	0,87	0,88	4165
Melanjutkan pendidikan	0,62	0,60	0,61	574
Wiraswasta / pernah wiraswasta	0,21	0,29	0,25	340
Accuracy			0,80	5079
Macro average	0,58	0,59	0,58	5079
Weighted average	0,82	0,80	0,81	5079

Tabel 2 memperlihatkan bahwa performa model sangat tidak merata antar kelas. Pada kelas WO, model menunjukkan performa relatif kuat. Pada ST, performa model berada pada tingkat moderat, sehingga model telah mampu menangkap sebagian pola alumni yang melanjutkan pendidikan, tetapi belum memisahkannya secara tajam dari WO. Kinerja terlemah muncul pada WU, baik dari sisi *precision* maupun *recall*. Rendahnya *precision* WU menunjukkan bahwa sebagian prediksi WU berasal dari observasi yang sebenarnya bukan WU, sedangkan rendahnya *recall* menunjukkan bahwa sebagian besar observasi WU aktual belum berhasil dikenali dengan baik. Pola ini menegaskan bahwa masalah utama model bukan hanya pada *under-detection* kelas minoritas, tetapi juga pada belum stabilnya batas keputusan antar kelas, terutama antara WO dan WU.

Matriks konfusi pada data uji menunjukkan bahwa kesalahan model tidak tersebar secara acak, melainkan terkonsentrasi terutama pada aliran ST ke WO dan, lebih kuat lagi, WU ke WO. Dengan kata lain, ketika model menghadapi observasi yang profil fiturnya ambigu, prediksi cenderung mengarah ke kelas mayoritas. Hal ini menandakan bahwa representasi fitur yang tersedia saat ini masih lebih efektif dalam menangkap pola alumni yang bekerja dibandingkan dalam membedakan jalur studi lanjut dan kewirausahaan. Dalam konteks analitika kelembagaan, pola seperti ini penting karena tujuan model bukan hanya memetakan *outcome* dominan, tetapi juga membedakan jalur pasca-kelulusan yang secara jumlah lebih kecil namun substantif.



Gambar 2. Matriks konfusi ternormalisasi

Sebagaimana terlihat pada Gambar 2, *recall* untuk WO berada pada tingkat tertinggi, *recall* ST berada pada tingkat menengah, dan *recall* WU merupakan yang terendah. Visualisasi ternormalisasi ini memperjelas bahwa kelemahan model pada WU tidak hanya berasal dari sedikitnya jumlah data, tetapi juga dari lemahnya pemisahan pola fitur yang membuat banyak observasi WU dialihkan ke WO. Dengan demikian, Gambar 2 menguatkan pembacaan dari Tabel 2 bahwa performa model harus dibaca melalui struktur salah-klasifikasi, bukan hanya melalui akurasi total.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model XGBoost cukup memadai untuk memetakan pola dominan status kerja alumni, terutama pada kelas WO, dan cukup berguna untuk membaca sebagian pola pada ST. Akan tetapi, model belum memadai untuk tugas klasifikasi yang menuntut pemisahan yang andal terhadap WU. Karena itu, kontribusi utama hasil penelitian ini bukan pada klaim bahwa model telah mencapai performa tinggi secara umum, melainkan pada penunjukan yang lebih teliti mengenai kelas mana yang telah dipelajari model secara relatif baik, kelas mana yang masih lemah, serta bagaimana struktur salah-klasifikasi tersebut harus dibaca dalam konteks data tracer study yang timpang. Jika dibandingkan dengan Shoaib et al. (2026), penelitian ini memperlihatkan pola yang sejalan dalam hal kelayakan XGBoost untuk klasifikasi luaran lulusan berbasis data tabular multikelas. Namun, performa agregat penelitian ini tetap lebih rendah daripada *F1-score* 0,89 dan *AUC-ROC* 0,93 yang dilaporkan Shoaib et al. (2026). Perbedaan tersebut perlu dibaca hati-hati karena kedua studi menggunakan definisi target, ukuran dataset, cakupan atribut, dan rancangan eksperimen yang berbeda. Shoaib et al. (2026) mengintegrasikan atribut akademik, keterampilan, latar sosial-ekonomi, dan tren pasar kerja serta membandingkan XGBoost dengan beberapa *baseline*, sedangkan penelitian ini berfokus pada data tracer study institusional dengan satu *pipeline* utama. Dengan demikian, komparasi ini lebih tepat dipahami sebagai penegasan bahwa XGBoost tetap kompetitif pada persoalan klasifikasi luaran lulusan, bukan sebagai pembuktian keunggulan universal atas semua konteks.

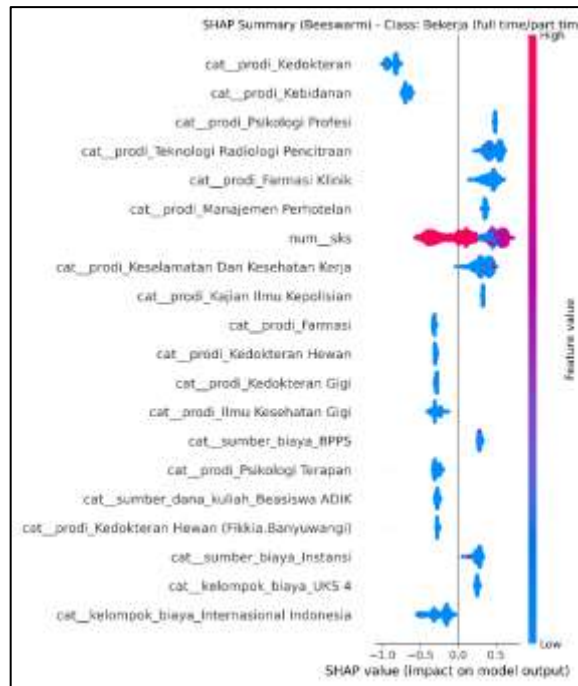
4.2 Analisis Interpretabilitas Model (SHAP)

Interpretabilitas dianalisis menggunakan SHAP untuk membaca kontribusi fitur terhadap prediksi model pada tiap kelas. Nilai SHAP positif ditafsirkan sebagai fitur yang mendorong prediksi menuju kelas yang sedang dianalisis, sedangkan nilai negatif menjauhkan prediksi dari kelas tersebut. Karena model bersifat multikelas, pembacaan dilakukan per kelas dan dibandingkan antar kelas untuk mengidentifikasi pola pemisahan yang dipelajari model. Dalam artikel ini, SHAP digunakan sebagai penjelasan perilaku model prediktif, bukan sebagai dasar inferensi kausal. Dibandingkan dengan Shoaib et al. (2026) yang juga menggunakan SHAP untuk menjelaskan model XGBoost multikelas, penelitian ini menempatkan interpretabilitas secara lebih langsung pada pembacaan *outcome* tracer study institusional, yaitu WO, ST, dan WU. Bila pada Shoaib et al. (2026) SHAP dipakai untuk menjelaskan kategori *vocational mobility* yang dibangun dari kombinasi atribut lulusan dan tren pasar kerja, maka di sini SHAP dipakai untuk memeriksa bagaimana fitur akademik, institusional, dan pembiayaan bekerja dalam membedakan tiga jalur pasca-kelulusan alumni.

4.2.1 Kelas WO (Bekerja / pernah bekerja)

Pada kelas WO, model memanfaatkan kombinasi fitur kategorikal, terutama kategori program studi, fakultas, dan beberapa kategori pembiayaan, serta fitur numerik tertentu, khususnya SKS. Sejumlah kategori prodi dan fakultas memberi kontribusi positif terhadap probabilitas WO, sementara kategori lain memberi kontribusi negatif. Hal ini menunjukkan bahwa model menangkap pola historis luaran karier yang berbeda antar konteks kelembagaan. Dengan demikian, sinyal prediksi WO tersebar pada beberapa kelompok fitur dan tidak ditentukan oleh satu variabel tunggal. Pola yang relatif jelas pada fitur numerik adalah SKS: nilai SKS yang lebih tinggi cenderung berkaitan dengan nilai SHAP negatif untuk WO, sedangkan nilai yang lebih rendah cenderung berkaitan dengan nilai SHAP positif. Pembacaan ini merefleksikan pola statistik yang dipelajari model, bukan hubungan sebab-akibat. Kestabilan sinyal pada WO konsisten dengan kinerja prediktif kelas ini yang paling kuat. Pola ini sejalan dengan Shoaib et al. (2026), yang menunjukkan bahwa *outcome* dengan pola mobilitas yang lebih stabil cenderung memiliki struktur kontribusi fitur yang lebih jelas dan karenanya lebih mudah diprediksi.

Dalam kedua studi, kelas dengan pola kontribusi yang lebih koheren menjadi kelas dengan performa prediktif relatif lebih tinggi.



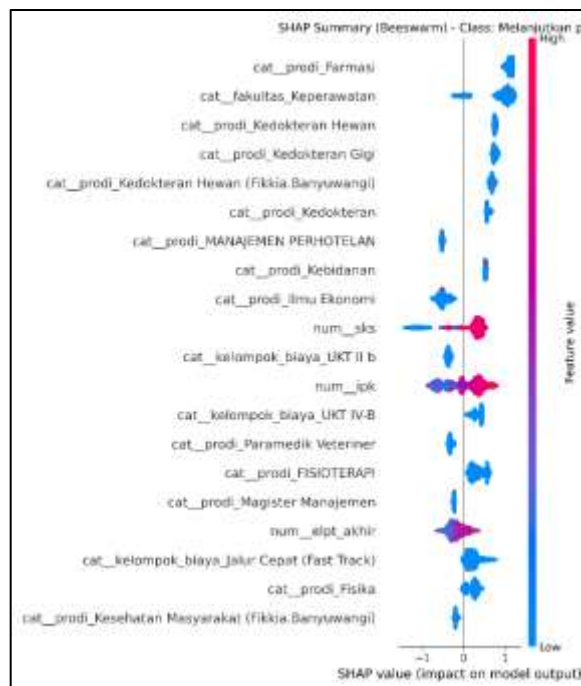
Gambar 3. SHAP Summary Plot untuk Kelas WO (Bekerja / pernah bekerja)

Gambar 3 memperlihatkan bahwa distribusi kontribusi fitur pada kelas WO relatif lebih stabil dibandingkan dua kelas lain. Sebaran nilai SHAP yang lebih jelas pada sejumlah fitur utama menunjukkan bahwa model memiliki dasar keputusan yang lebih kuat dalam mengenali alumni yang bekerja. Hal ini sejalan dengan performa WO yang paling tinggi pada evaluasi klasifikasi.

4.2.2 Kelas ST (Melanjutkan pendidikan)

Untuk kelas ST, SHAP menunjukkan peran yang lebih kuat dari fitur akademik numerik, terutama SKS dan IPK, di samping kategori prodi dan fakultas tertentu. Secara umum, nilai SKS dan IPK yang lebih tinggi cenderung mendorong prediksi ke ST, sedangkan nilai yang lebih rendah menahannya. Pola ini membantu menjelaskan mengapa model masih mampu mengenali sebagian alumni yang melanjutkan pendidikan, meskipun tidak seluruhnya dapat dipisahkan secara konsisten dari WO. Dengan demikian, model tampak belajar bahwa jalur studi lanjut lebih banyak dikaitkan dengan kombinasi sinyal akademik dan konteks institusional tertentu, walaupun separasinya belum sepenuhnya kuat. Jika dibandingkan dengan temuan Shoaib et al. (2026), kelas yang merepresentasikan kondisi transisional seperti ST menuntut kombinasi sinyal yang lebih berlapis daripada kelas yang lebih stabil. Hal ini menunjukkan bahwa *outcome* yang tidak sepenuhnya mapan di satu jalur memerlukan fitur yang lebih kaya agar separasinya meningkat.

Gambar 4 memperlihatkan bahwa distribusi kontribusi fitur pada ST cukup terstruktur, tetapi belum setajam WO. Sejumlah observasi ST tetap berada dekat dengan wilayah keputusan WO, yang menjelaskan mengapa *recall* ST berada pada tingkat menengah. Visual ini mendukung pembacaan bahwa model telah mempelajari sebagian pola ST, tetapi belum cukup kuat untuk memisahkan seluruh observasi ST secara stabil dari jalur bekerja.

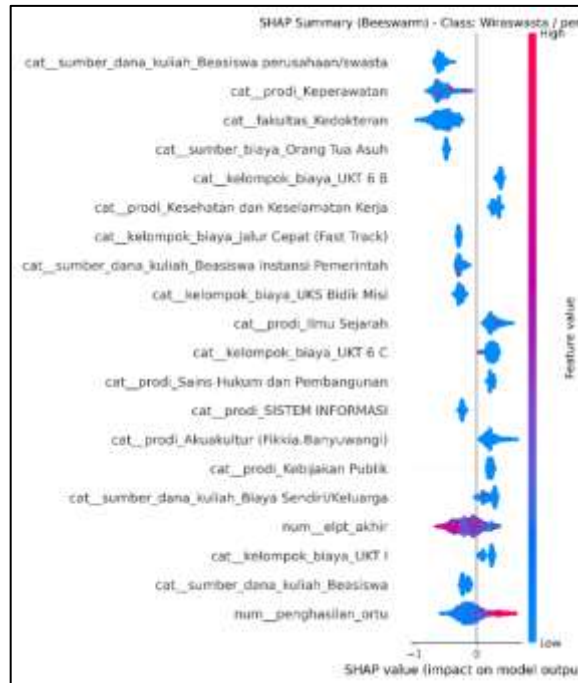


Gambar 4. SHAP Summary Plot untuk Kelas ST (Melanjutkan pendidikan)

4.2.3 Kelas WU (Wiraswasta / pernah wiraswasta)

Pada kelas WU, sebagian besar nilai SHAP terkonsentrasi di sekitar nol dengan kontribusi positif yang lemah dan tidak konsisten antar fitur. Hal ini menunjukkan bahwa, dengan fitur yang tersedia, model belum mempelajari pola pemisahan yang cukup kuat untuk membedakan jalur WU dari kelas lain, khususnya WO. Temuan ini konsisten dengan rendahnya performa WU pada evaluasi klasifikasi dan dengan dominannya salah-klasifikasi WU ke WO pada matriks konfusi. Dengan kata lain, problem utama WU bukan hanya jumlah observasinya yang lebih kecil, tetapi juga lemahnya daya pisah representasi fitur untuk *outcome* kewirausahaan. Perbedaan ini menjadi lebih jelas jika dibandingkan dengan Shoaib et al. (2026), yang mencapai performa agregat lebih tinggi dengan dukungan atribut yang lebih luas, termasuk keterampilan dan tren pasar kerja. Temuan tersebut mengisyaratkan bahwa kelemahan WU dalam penelitian ini kemungkinan berkaitan dengan keterbatasan representasi fitur institusional untuk menangkap dinamika kewirausahaan, bukan semata-mata dengan pilihan algoritma.

Gambar 5 menunjukkan bahwa distribusi nilai SHAP pada kelas WU lebih tersebar di sekitar nol dan tidak membentuk pola yang cukup kuat. Visual ini mendukung kesimpulan bahwa model mengalami kesulitan dalam membangun sinyal prediktif yang stabil untuk kewirausahaan. Oleh sebab itu, kelemahan WU harus dipahami bukan hanya sebagai persoalan jumlah data yang lebih sedikit, tetapi juga sebagai persoalan keterbatasan fitur dalam merepresentasikan karakteristik kewirausahaan secara lebih tajam.



Gambar 5. SHAP Summary Plot untuk Kelas WU (Wiraswasta / pernah wiraswasta)

4.2.4 Implikasi Interpretabilitas terhadap Pengembangan Model

Analisis SHAP menunjukkan dua arah pengembangan yang saling terkait. Pertama, strategi evaluasi dan pelatihan perlu dibuat lebih eksplisit terhadap tujuan kelas yang ditarget, misalnya apabila fokus kelembagaan adalah meningkatkan sensitivitas pada ST atau WU. Kedua, fitur untuk jalur kewirausahaan perlu diperkaya agar separabilitas WU meningkat, misalnya melalui indikator aktivitas usaha, pengalaman organisasi atau bisnis, status pekerjaan yang lebih rinci, pola penghasilan, atau pemanfaatan fitur tekstual tracer study apabila tata kelola data memungkinkan. Dengan demikian, interpretabilitas berbasis SHAP berfungsi bukan hanya sebagai alat baca terhadap perilaku model yang ada, tetapi juga sebagai dasar penentuan prioritas pengembangan model pada siklus berikutnya. Dalam perspektif komparatif, hasil ini konsisten dengan literatur yang menunjukkan bahwa peningkatan performa prediksi luaran lulusan tidak hanya bergantung pada algoritma, tetapi juga pada keluasan atribut dan rancangan evaluasi (Shoaib et al., 2026). Oleh karena itu, pengembangan berikutnya sebaiknya menggabungkan *benchmarking* antarmodel dan pengayaan fitur agar interpretabilitas model tidak berhenti pada pembacaan perilaku, tetapi juga membantu menjelaskan sumber keunggulan atau kelemahan performa secara lebih komparatif.

5. Penutup

5.1 Simpulan

Penelitian ini menunjukkan bahwa data tracer study dapat dimanfaatkan untuk membangun model klasifikasi multikelas status kerja alumni berbasis XGBoost, tetapi kualitas model tidak dapat dinilai secara memadai hanya melalui akurasi agregat. Pada artefak eksperimen yang digunakan dalam penelitian ini, model mencapai akurasi sekitar 0,80 pada data uji. Namun, evaluasi per kelas memperlihatkan ketimpangan performa yang jelas. Model menunjukkan kinerja relatif kuat pada kelas Bekerja (WO), dengan precision 0,89, recall 0,87, dan F1-score 0,88; kinerja moderat pada kelas Melanjutkan pendidikan (ST), dengan precision 0,62, recall 0,60, dan F1-score 0,61; serta kinerja terendah pada kelas Wiraswasta (WU), dengan precision 0,21, recall 0,29, dan F1-score 0,25. Temuan ini menegaskan bahwa pada data tracer study yang distribusi targetnya timpang, akurasi total tidak cukup untuk merepresentasikan kualitas model secara substantif.

Struktur matriks konfusi menunjukkan bahwa kesalahan prediksi model tidak tersebar secara acak, melainkan terkonsentrasi terutama pada aliran ST ke WO dan, secara lebih kuat, WU ke WO. Pola ini mengindikasikan bahwa model lebih berhasil mengenali outcome mayoritas yang memiliki representasi lebih dominan dan lebih stabil dalam data, tetapi masih mengalami kesulitan dalam memisahkan outcome minoritas yang secara empiris lebih heterogen. Dengan demikian, kelemahan utama model tidak hanya terletak pada rendahnya deteksi WU, tetapi juga pada belum stabilnya batas keputusan antara jalur bekerja dan jalur non-mayoritas, khususnya kewirausahaan.

Analisis interpretabilitas berbasis SHAP memperjelas bahwa performa model berkaitan erat dengan kualitas representasi fitur yang tersedia. Pada kelas WO, model memperlihatkan pola kontribusi fitur yang relatif lebih stabil, terutama melalui kombinasi variabel ekonomi keluarga, progres akademik, pembiayaan, dan konteks institusional. Pada kelas ST, model juga menunjukkan struktur kontribusi yang cukup bermakna, khususnya melalui fitur akademik dan beberapa kategori prodi dan fakultas. Sebaliknya, pada kelas WU, kontribusi fitur cenderung lemah dan tidak konsisten, sehingga model belum mampu membangun pemisahan yang kuat terhadap jalur kewirausahaan. Oleh sebab itu, kelemahan model pada WU harus dipahami bukan semata sebagai akibat dari ukuran kelas yang lebih kecil, melainkan juga sebagai akibat dari belum memadainya daya pisah fitur dalam merepresentasikan karakteristik kewirausahaan alumni.

Secara keseluruhan, kontribusi penelitian ini terletak pada penempatan data tracer study sebagai persoalan analitika kelembagaan multikelas, serta pada integrasi evaluasi per kelas, struktur salah-klasifikasi, dan interpretabilitas berbasis SHAP dalam menilai kelayakan model. Dalam bentuk saat ini, model dapat digunakan sebagai alat analitik awal untuk memetakan pola dominan status kerja alumni, terutama pada jalur bekerja, dan sampai tingkat tertentu pada jalur studi lanjut. Namun, model belum memadai untuk penggunaan kelembagaan yang menuntut sensitivitas tinggi terhadap jalur kewirausahaan. Dengan demikian, hasil penelitian ini lebih tepat dibaca sebagai dasar evaluasi metodologis dan pengembangan model lanjutan daripada sebagai sistem klasifikasi final yang siap digunakan tanpa perbaikan lebih lanjut.

5.2 Saran

Pengembangan penelitian berikutnya perlu diarahkan pada tiga poros utama, yaitu pengayaan data, penguatan strategi pemodelan, dan pengetatan evaluasi. Dari sisi data, prioritas utama adalah menambahkan fitur yang lebih spesifik terhadap jalur kewirausahaan, seperti indikator aktivitas usaha, pengalaman organisasi atau bisnis, keterlibatan pada program kewirausahaan, pola penghasilan, maupun informasi tekstual dari tracer study yang saat ini belum dimanfaatkan secara sistematis. Selama representasi kewirausahaan masih bergantung terutama pada variabel akademik, pembiayaan, dan latar institusional yang bersifat umum, model akan tetap mengalami kesulitan dalam membedakan WU dari WO secara andal.

Dari sisi pemodelan, penelitian lanjutan perlu menambahkan pembandingan model dan pengujian parameter secara eksplisit. Dalam versi eksperimen yang digunakan pada penelitian ini, evaluasi masih terbatas pada satu pipeline utama, yaitu preprocessing → SMOTE → XGBoost, sehingga belum memungkinkan klaim yang kuat mengenai model terbaik. Oleh karena itu, baseline seperti multinomial logistic regression atau Random Forest, serta model tabular lain seperti LightGBM atau CatBoost, perlu dimasukkan pada tahap berikutnya. Di samping itu, penggunaan SMOTE setelah one-hot encoding perlu dicatat sebagai pilihan pragmatis yang membantu proses pelatihan, tetapi tetap memiliki keterbatasan metodologis karena interpolasi pada fitur dummy kategorikal dapat menghasilkan representasi sintetis yang tidak sepenuhnya natural. Karena itu, pendekatan alternatif seperti SMOTENC, *sample weighting*, atau strategi penanganan imbalance lainnya layak diuji secara komparatif.

Dari sisi evaluasi, penelitian selanjutnya sebaiknya menempatkan *macro-based metrics*, recall per kelas, matriks konfusi lintas validasi, serta interpretabilitas model sebagai komponen standar penilaian. Pemilihan model tidak lagi sebaiknya ditentukan terutama oleh akurasi agregat, melainkan oleh kesesuaiannya dengan tujuan kelembagaan yang hendak dicapai. Jika tujuan analitik perguruan

tinggi adalah meningkatkan sensitivitas terhadap jalur studi lanjut atau kewirausahaan, maka desain evaluasi juga harus secara eksplisit diarahkan ke outcome tersebut, termasuk melalui *threshold tuning* dan pelaporan trade-off antar metrik. Bersamaan dengan itu, peningkatan kualitas data tracer study, terutama standardisasi kategori, konsistensi pengisian, dan relevansi variabel, tetap merupakan prasyarat utama bagi peningkatan performa model dan mutu interpretasinya.

Dengan demikian, pengembangan model tracer study ke depan tidak cukup dilakukan dengan mengganti algoritma semata, tetapi harus diarahkan pada peningkatan kualitas representasi data, perumusan tujuan evaluasi yang lebih eksplisit, dan integrasi yang lebih kuat antara performa prediktif dan kebutuhan substantif analitika kelembagaan.

Referensi

- Abdulloh, F. F., Rahardi, M., Aminuddin, A., Anggita, S. D., & Nugraha, A. Y. A. (2022). Observation of imbalance tracer study data for graduates employability prediction in Indonesia. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(8), 169–174. <https://doi.org/10.14569/IJACSA.2022.0130820>
- Albina, A. C., & Sumagaysay, L. P. (2020). Employability tracer study of Information Technology Education graduates from a state university in the Philippines. *Social Sciences & Humanities Open*, 2(1), 100055. <https://doi.org/10.1016/j.ssaho.2020.100055>
- Azis, A. L., Aziz, F., & Irmawati. (2026). Application of machine learning algorithms to predict alumni employability based on tracer study data. *Jurnal Cakrawala dan Integrasi Pendidikan*, 2(1), 26–33.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
- Dake, D. K. (2025). Information technology graduates employability prediction model in a low-income country using tree-based machine learning classifiers. *Discover Global Society*, 3, 158. <https://doi.org/10.1007/s44282-025-00304-3>
- Dzomeku, V. M., Amponsah, A. K., Boateng, E. A., Antwi, J., Amooba, P. A., Gracious, P. D., Armah, J., & Bam, V. (2024). Tracer study to assess the employability of graduates and quality of a nursing program: A descriptive cross-sectional survey. *International Journal of Africa Nursing Sciences*, 20, 100673. <https://doi.org/10.1016/j.ijans.2024.100673>
- ElSharkawy, G., Helmy, Y., & Yehia, E. (2022). Employability prediction of information technology graduates using machine learning algorithms. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(10), 359–367. <https://doi.org/10.14569/IJACSA.2022.0131043>
- Garcia, C. M., Sarsale, M. S., & Uy, I. M. (2024). Dimensions of program relevance towards employment success: Evidence from a graduate tracer study using principal component analysis. *Journal of Teaching and Learning for Graduate Employability*, 15(1), 205–224.
- Haque, R., Goh, H.-N., Ting, C.-Y., Quek, A., & Hasan, M. D. R. (2025). Leveraging LLMs for optimised feature selection and embedding in structured data: A case study on graduate employment classification. *Computers and Education: Artificial Intelligence*, 8, 100356. <https://doi.org/10.1016/j.caeai.2024.100356>
- Haque, R., Quek, A., Ting, C.-Y., Goh, H.-N., & Hasan, M. R. (2024). Classification techniques using machine learning for graduate student employability predictions. *International Journal on Advanced Science, Engineering and Information Technology*, 14(1), 45–56. <https://doi.org/10.18517/ijaseit.14.1.19549>
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, 5(4), 221–232. <https://doi.org/10.1007/s13748-016-0094-0>
- Lopena, G. L., & Madrigal, D. V. (2023). A tracer study of the business graduate programs of a Catholic university in the Philippines. *International Education Studies*, 16(2), 76–88. <https://doi.org/10.5539/ies.v16n2p76>

- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30).
- Mahdin, H., Nurwarsito, H., Baharum, Z., Kamri, K. A., Hassan, A., Haw, S.-C., & Arshad, M. S. (2025). Predictive analytics for employability in Malaysian TVET with a hybrid of regression and clustering methods. *JOIV: International Journal on Informatics Visualization*, 9(5), 1816–1821. <https://doi.org/10.62527/joiv.9.5.4516>
- Mansor, A., & Othman, Z. (2024). Malaysian community college graduates employability prediction model using machine learning approach. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 16(3), 1–7. <https://doi.org/10.54554/jtec.2024.16.03.001>
- Mezhoudi, N., Alghamdi, R., Aljunaid, R., Krichna, G., & Düşteğör, D. (2023). Employability prediction: A survey of current approaches, research challenges and applications. *Journal of Ambient Intelligence and Humanized Computing*, 14, 1489–1505. <https://doi.org/10.1007/s12652-021-03276-9>
- Munthali, C., Mkwinda, E., Mselema, C., Jere, J., Banda, M. K., Mugala, Z., Jere, T., Mfuno, D., Hara, M., & Chironda, G. (2025). A graduate tracer study of the diploma in nursing program at Saint John of God College of Health Sciences in Mzuzu, Malawi: A descriptive survey design. *International Journal of Africa Nursing Sciences*, 23, 100925. <https://doi.org/10.1016/j.ijans.2025.100925>
- Ramos, M. C. M. (2024). Graduate tracer study: Exploring the relationship between education and employment. *Asia Pacific Journal of Management and Sustainable Development*, 12(1), 108–117.
- Shoaib, M., Afridi, M. U., & Husnain, G. (2026). *Forecasting algorithm for the vocational mobility of graduates with multiple attributes*. *Expert Systems With Applications*, 297, 129463. <https://doi.org/10.1016/j.eswa.2025.129463>
- Vo, M.-T., Nguyen, T., & Le, T. (2023). Employability prediction of university graduates using optimized bagging model. *Computers, Materials & Continua*, 76(2), 1555–1568.