

Makalah Penelitian

Klasifikasi Tingkat Penghasilan Freelancer Menggunakan Algoritma Random Forest

Febriansyah¹, Indah Susilawati²

^{1,2}Informatika, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta
¹201110019@student.mercubuana-yogya.ac.id*, ²indah@mercubuana-yogya.ac.id

Corresponding Author: write name of corresponding author

ABSTRACT

This research aims to identify the influencing factors and classify freelancers' income levels through the development of a Machine Learning model. Income is grouped into four categories: Low, Medium, High, and Very High. The method used is the Random Forest algorithm, which is applied to a dataset containing 1,950 freelancers. The data pre-processing stage includes innovative feature engineering one of which is the creation of a Marketing Ratio variable (the ratio of marketing costs to revenue) and a comprehensive hyperparameter tuning process using RandomizedSearchCV. The evaluation results on the test data showed that the optimized Random Forest model achieved excellent classification performance with an accuracy of 0.9103 (91.03%), outperforming other alternative models. Furthermore, the feature importance analysis confirmed that Marketing Ratio (53.2%) and Marketing Spend (39.9%) were the strongest predictors. This indicates that marketing effectiveness and proper budget management have a greater impact on increasing freelancer income than conventional factors such as experience or client ratings. Thus, the developed model not only provides high prediction accuracy, but also offers valuable business insights especially regarding the importance of optimizing marketing Return on Investment (ROI) for both freelancers and related platforms.

Keywords: Random Forest, Classification, Freelancer, Income, Marketing Ratio, Feature Importance, Machine Learning

ABSTRAK

Penelitian ini bertujuan untuk mengidentifikasi faktor serta mengklasifikasi tingkat pendapatan freelancer melalui pengembangan model *Machine Learning*. Pendapatan dikelompokkan ke dalam empat kategori, yaitu Rendah (Low), Menengah (Medium), Tinggi (High), dan Sangat Tinggi (Very High). Metode yang digunakan adalah algoritma Random Forest, yang diterapkan pada dataset berisi 1.950 data freelancer. Tahap pra-pengolahan data meliputi feature engineering yang bersifat inovatif salah satunya pembuatan variabel Marketing Ratio (perbandingan biaya pemasaran terhadap pendapatan) serta proses hyperparameter tuning secara menyeluruh menggunakan RandomizedSearchCV. Hasil evaluasi menunjukkan bahwa model Random Forest yang telah dioptimalkan mencapai performa klasifikasi yang sangat baik, dengan nilai akurasi 0.9103 (91.03%) pada data uji, dan mampu mengungguli model-model alternatif lainnya. Selain itu, analisis Feature Importance menegaskan bahwa Marketing Ratio (53.2%) dan Marketing Spend (39.9%) merupakan prediktor terkuat. Hal ini mengindikasikan bahwa efektivitas pemasaran dan pengelolaan anggaran yang tepat memiliki pengaruh lebih besar terhadap peningkatan pendapatan freelancer dibandingkan aspek konvensional seperti pengalaman atau rating klien. Dengan demikian, model yang dikembangkan tidak hanya memberikan akurasi prediksi yang tinggi, tetapi juga menawarkan wawasan bisnis yang bernilai, terutama mengenai pentingnya optimalisasi Return on Investment (ROI) pemasaran bagi freelancer maupun platform terkait.

Kata Kunci: Random Forest, Classification, Freelancer, Income, Marketing Ratio, Feature Importance, Machine Learning.



Lisensi
Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

1. Pendahuluan

Freelance kini menjadi pilihan karier yang diminati banyak orang karena menawarkan fleksibilitas, kebebasan dalam bekerja, serta peluang untuk mencoba beragam jenis pekerjaan tanpa harus terikat pada satu perusahaan atau lokasi tertentu [1]. Pendapatan seorang freelancer umumnya tidak memiliki penghasilan bulanan yang bersifat tetap, sehingga mereka juga tidak memperoleh Tunjangan Hari Raya (THR) sebagaimana pekerja yang bekerja di sektor formal [2]. Walaupun pekerjaan lepas memberikan kesempatan untuk mengasah keterampilan baru dan mengeksplorasi berbagai jenis proyek, para pekerja ini juga harus berhadapan dengan tingkat persaingan global yang tinggi [3].

Random Forest adalah salah satu algoritma machine learning yang banyak digunakan dalam proses data mining. Algoritma ini berfungsi untuk melakukan tugas klasifikasi atau penentuan suatu kelas. Metode ini terdiri atas sekumpulan pohon keputusan (decision trees) yang saling terhubung dan bekerja secara ensemble, dengan tujuan meningkatkan kinerja model serta memperoleh tingkat akurasi yang lebih optimal [4]. Decision trees bekerja dengan memecah dataset menjadi subset yang lebih kecil berdasarkan fitur yang paling signifikan dalam memprediksi label target [5]. Random Forest juga merupakan sebuah classifier yang terdiri dari sekumpulan tree-structured classifiers yang identik, independen, dan terdistribusi berdasarkan vektor acak. Setiap pohon memberikan satu suara untuk kelas yang paling populer pada data masukan. [6].

2. Tinjauan Pustaka

Penelitian ini akan mengkaji aspek ketidakpastian pendapatan dengan tujuan untuk mengklasifikasikannya dalam kategori yang dapat diukur. Variabel-variabel yang biasa digunakan untuk menggambarkan profil pekerja lepas mencakup jenis keterampilan, pengalaman kerja, peringkat atau reputasi di platform, serta jumlah proyek yang berhasil diselesaikan. Dalam studi ini, metode klasifikasi digunakan untuk membagi tingkat penghasilan freelancer menjadi empat kategori tingkat pendapatan, yaitu Low, Medium, High, dan Very High, untuk menghasilkan segmentasi yang lebih granular dan representatif terhadap distribusi data yang skewed.

2.1 Random Forest

Random Forest (RF) merupakan metode ensemble berbasis pohon yang dikembangkan untuk mengatasi keterbatasan pada metode Classification and Regression Tree (CART). Metode ini tersusun dari banyak pohon keputusan klasifikasi dan regresi yang dibangun secara paralel, sehingga mampu menurunkan bias dan varians model secara simultan [7]. Algoritma ini, yang diciptakan oleh Leo Breiman pada tahun 2001, beroperasi dengan cara membuat sejumlah Decision Tree selama fase pelatihan dan menggabungkan hasilnya untuk menghasilkan prediksi akhir yang lebih konsisten dan tepat. Salah satu keuntungan utama dari Random Forest yang menjadi alasan pemilihannya dalam penelitian ini adalah ketahanannya terhadap masalah overfitting, kemampuannya dalam menangani data baik numerik maupun kategorikal, serta kemampuannya untuk menilai tingkat pentingnya setiap variabel dalam proses klasifikasi. Penggunaan Random Forest mampu mencegah terjadinya overfitting pada suatu dataset sekaligus mencapai tingkat akurasi yang optimal [8].

**Lisensi**

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

Penerapan algoritma Random Forest untuk klasifikasi di bidang sosio-ekonomi dan finansial telah didukung oleh sejumlah penelitian empiris. Penelitian di Kota Cirebon membuktikan keefektifan algoritma ini dalam mengklasifikasikan status ekonomi rumah tangga dengan akurasi mencapai 93%. Yang patut dicatat, penelitian ini berhasil mengatasi masalah ketidakseimbangan data pada kelas minoritas menggunakan teknik SMOTE [9]. Dalam konteks yang lebih dekat dengan sektor finansial juga menunjukkan keberhasilan Random Forest dalam memprediksi risiko kredit macet di sebuah koperasi dengan akurasi 94,8%, menegaskan kontribusinya sebagai alat bantu decision support yang akurat [10]. Relevansi yang lebih langsung ditunjukkan dalam mengklasifikasikan pendapatan individu menggunakan dataset Adult Income dari UCI, dengan akurasi 85,44%. Namun, penelitian ini juga mengungkap tantangan dalam ketidakseimbangan kelas, di mana performa model pada kelas minoritas (penghasilan tinggi) memiliki recall yang rendah [11]. Berdasarkan peta penelitian diatas, terlihat celah untuk menerapkan Random Forest yang telah terbukti handal secara spesifik untuk mengklasifikasikan penghasilan freelancer di Indonesia.

3. Bahan & Metode

Penelitian ini terdiri dari lima tahap utama yang dilakukan secara sistematis, dimulai dari pengumpulan data, preprocessing data, pelatihan model, evaluasi model, hingga pengujian model.



Gambar 3.1 Tahapan Penelitian

Alur tahapan dimulai dari proses akuisisi data untuk mendapatkan dataset yang representatif. Dataset tersebut kemudian masuk ke tahap preprocessing, di mana data dibersihkan dan ditransformasikan, termasuk melalui penerapan perhitungan untuk feature engineering guna meningkatkan kualitas fitur. Setelah data siap, dilakukan pengembangan arsitektur model menggunakan algoritma Random Forest untuk mempelajari pola data. Tahapan selanjutnya adalah melakukan analisis feature importance guna mengekstrak wawasan mengenai variabel yang paling berpengaruh terhadap model. Terakhir, model diuji dan dievaluasi kemampuannya menggunakan metrik standar klasifikasi untuk memastikan reliabilitas prediksi. Penjelasan rinci mengenai teori yang digunakan serta prosedur teknis untuk masing-masing tahapan tersebut dipaparkan dalam sub bab berikut.

3.1. Pengumpulan Dataset

Dataset penelitian ini berasal dari kaggle yang berisi informasi aktivitas freelancer yang dikumpulkan dari berbagai platform, seperti Fiverr, Upwork, PeoplePerHour, dan Toptal.

3.2. Preprocessing Data

Preprocessing data adalah tahapan penting dalam analisis data mining yang berfungsi untuk membersihkan data, mengubah bentuk atau strukturnya, serta menyiapkan data agar dapat dianalisis dengan lebih mudah dan menghasilkan hasil yang lebih akurat [12]. Sebagai contoh, satu data freelancer dengan atribut Job_Category = Web Development, Platform = Upwork, Experience_Level = Intermediate, Job_Completed = 25, Earnings_USD = 5000, dan Marketing_Spend = kosong, terlebih dahulu melalui proses pembersihan dengan mengisi nilai



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

kosong menggunakan median (misalnya menjadi 100). Selanjutnya, atribut kategorikal dikonversi menggunakan One-Hot Encoding misalnya Platform_Upwork = 1, kemudian atribut numerik dinormalisasi menggunakan Min-Max Scaling (misalnya Earnings_USD menjadi 0,5). Hasil akhir berupa data numerik terstruktur ini kemudian digunakan sebagai input pada model Random Forest untuk proses pelatihan dan pengujian.

3.3. Arsitektur Model

Model yang digunakan dalam penelitian ini adalah Random Forest Classifier, yaitu algoritma berbasis ensemble learning yang menggunakan sekumpulan pohon keputusan acak untuk menghasilkan prediksi [13]. Dalam proses pembentukan pohon keputusan, kriteria Gini Impurity digunakan sebagai dasar optimasi untuk menentukan pemisahan (split) ruang fitur yang paling optimal. Pendekatan ini dipilih karena kemampuannya dalam menangani variabel multidimensi secara nonlinier, sehingga diharapkan dapat mengidentifikasi variabel diskriminatif yang secara efektif membedakan setiap kategori pendapatan freelancer berdasarkan tingkat heterogenitas data pada setiap node.

Dalam proses pembangunan model, klasifikasi pada leaf nodes ditentukan menggunakan fungsi argmax terhadap distribusi probabilitas dalam array value, sehingga kelas dengan densitas tertinggi ditetapkan sebagai label output akhir. Sebelum tahap induksi pohon dilakukan, seluruh fitur input ditransformasikan menggunakan StandardScaler untuk menormalisasi distribusi data, yang ditandai dengan penggunaan nilai threshold pada struktur pohon yang terbentuk. Arsitektur ini dirancang untuk menangkap pola hubungan non-linier antar variabel secara efisien sekaligus menjaga transparansi dalam setiap tahapan pengambilan keputusan model Random Forest saat mengklasifikasikan tingkat penghasilan freelancer.

3.4. Analisis Feature Important

Feature importance dalam pengklasifikasi random forest digunakan untuk mengidentifikasi variabel yang paling berpengaruh, sehingga temuan penelitian dapat digunakan sebagai dasar penyusunan rekomendasi kebijakan yang lebih terarah [14].

3.5. Pengujian Model

Pengujian dalam penelitian ini difokuskan pada evaluasi performa model random forest classifier yang telah dibangun menggunakan data uji sebesar 20% dari total dataset. Proses pengujian akan dilakukan secara komprehensif menggunakan metrik evaluasi standar klasifikasi, yang meliputi pengujian tingkat akurasi untuk melihat persentase prediksi yang tepat secara keseluruhan serta analisis presisi dan Recall untuk menilai kualitas prediksi pada setiap kategori penghasilan (Rendah, Menengah, Tinggi, dan Sangat Tinggi).

4. Hasil dan Pembahasan

4.1. Pengumpulan Dataset

Penelitian ini menggunakan dataset yang terdiri dari 1.950 sampel freelancer dengan 15 fitur awal. Analisis statistik pada variabel Earnings_USD menunjukkan bahwa data tidak terdistribusi secara normal dengan nilai skewness positif, dimana mayoritas freelancer memiliki pendapatan pada rentang USD 0 hingga USD 2.000. Ketimpangan ini menjadi dasar dilakukannya pengelompokan (binning) pendapatan ke dalam empat kelas tingkatan ekonomi, yaitu Low, Medium, High, dan Very High.



P_ID	Job_Category	Platform	Experience_Level	Client_Region	Payment_Method	Job_Completed	Earnings_USD	Hourly_Rate	Job_Success_Rate	Client_Rating	Job_Duration_Days	Project_Type	Rehire_Rate	Marketing_Spend
1	Web Development	Fiverr	Beginner	Asia	Mobile Banking	180	1620	95.79	68.73	3.18	1	Fixed	40.19	53
2	App Development	Fiverr	Beginner	Australia	Mobile Banking	218	9078	86.38	97.54	3.44	54	Fixed	36.53	486
3	Web Development	Fiverr	Beginner	UK	Crypto	27	3455	85.17	86.60	4.20	46	Hourly	74.05	489
4	Data Entry	PeoplePerHour	Intermediate	Asia	Bank Transfer	17	5577	14.37	79.93	4.47	41	Hourly	27.58	67
5	Digital Marketing	Upwork	Expert	Asia	Crypto	245	5898	99.37	57.80	5.00	41	Hourly	69.09	489
6	Customer Support	Toptal	Beginner	Europe	Crypto	280	6867	43.04	57.80	4.87	8	Fixed	43.88	290
7	Web Development	Fiverr	Beginner	USA	Crypto	96	1677	20.50	52.90	4.29	32	Hourly	45.99	343
8	Data Entry	Toptal	Beginner	Australia	Bank Transfer	112	6193	82.15	93.31	3.84	30	Fixed	31.59	168
9	Content Writing	Toptal	Intermediate	USA	Crypto	233	8446	26.44	80.06	4.27	46	Fixed	50.11	396
10	Data Entry	PeoplePerHour	Beginner	Middle East	Mobile Banking	156	6608	54.99	85.40	4.57	52	Hourly	32.76	160

```
=====
Total Sampel: 1950
Total Fitur : 15
=====
```

Gambar 4.1 Dataset

4.2. Hasil Preprocessing Data dan Feature Engineering

Perbandingan antara Tabel 4.1 dan Tabel 4.2 menunjukkan transformasi signifikan yang terjadi selama tahap preprocessing dan feature engineering. Tabel 4.1 menampilkan dataset mentah dengan skala nilai yang bervariasi pada atribut numerik seperti *Job_Completed* dan *Hourly_Rate*, yang belum terstandarisasi. Setelah melalui proses data cleaning, normalisasi menggunakan Min-Max Scaler, serta pembentukan fitur baru, Tabel 4.2 merefleksikan dataset yang telah terstruktur secara seragam. Nilai-nilai pada tabel kedua kini berada dalam rentang yang telah dinormalisasi serta mencakup fitur-fitur hasil rekayasa seperti *Hourly_x_Success*, *Rating_x_Rehire*, dan *Marketing_Ratio*, yang dirancang untuk meningkatkan sensitivitas model dalam mengenali pola data yang lebih kompleks.

Tabel 4.1 Data sebelum proses *preprocessing*

Job_Completed	Hourly_rate	Job_Success_Rate	...	Marketing_Spend
180	95.79	68.73	...	53
218	86.38	97.54	...	486
27	85.17	86.60	...	489
17	14.37	79.93	...	67
...	...	...	...	...

Tabel 4.2 Data sesudah proses *preprocessing*

Job_Completed	Hourly_x_rate	Rating_x_Rehire	...	Marketing_Ratio
1950.000000	6.583.646.700.000.000	1.278.042	...	0.03269586674892042
150.864103	84.255.052	1.256.632	...	0.05353012446304659
85.480770	7.375.722	311.01	...	0.14149305555555555
5.000000	11.485.941	12.328.259.9	...	0.01201147364646826
		99.999.900		8
...	...	...	...	...

Tahap pra-pengolahan data berhasil menangani nilai kosong menggunakan metode median untuk data numerik dan modus untuk data kategorikal. Transformasi fitur dilakukan melalui



Lisensi

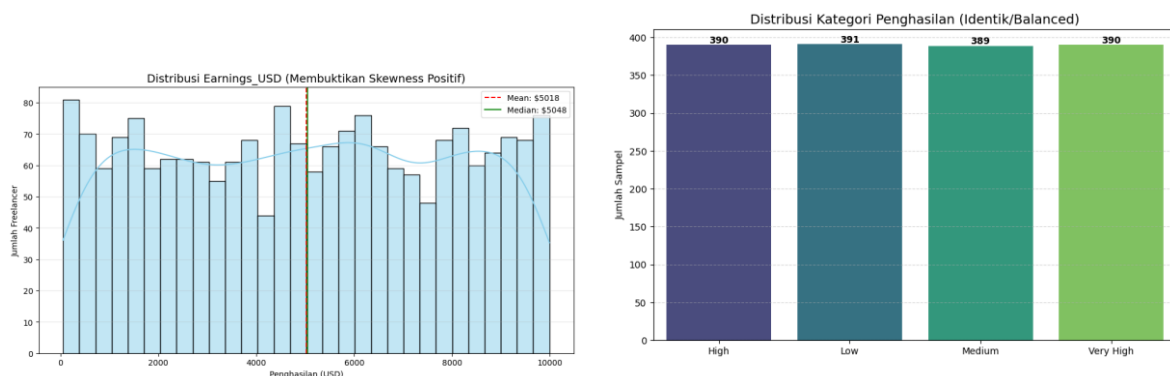
Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

teknik One-Hot Encoding untuk variabel kualitatif dan Min-Max Scaler untuk normalisasi atribut numerik ke rentang 0 hingga 1. Selain itu, dilakukan feature engineering yang menghasilkan fitur baru seperti Hourly_x_Success, Rating_x_Rehire, dan Marketing_Ratio. Melalui metode SelectKBest, terpilih 20 fitur terbaik dari total 28 fitur yang tersedia untuk digunakan dalam pemodelan.

	Job_Category	Experience_Level	Earnings_USD		Job_Category_Encoded	Experience_Level_Encoded	Earnings_USD
0	Web Development	Beginner	1620	0	7	0	1620
1	App Development	Beginner	9078	1	0	0	9078
2	Web Development	Beginner	3455	2	7	0	3455
3	Data Entry	Intermediate	5577	3	3	2	5577
4	Digital Marketing	Expert	5898	4	4	1	5898

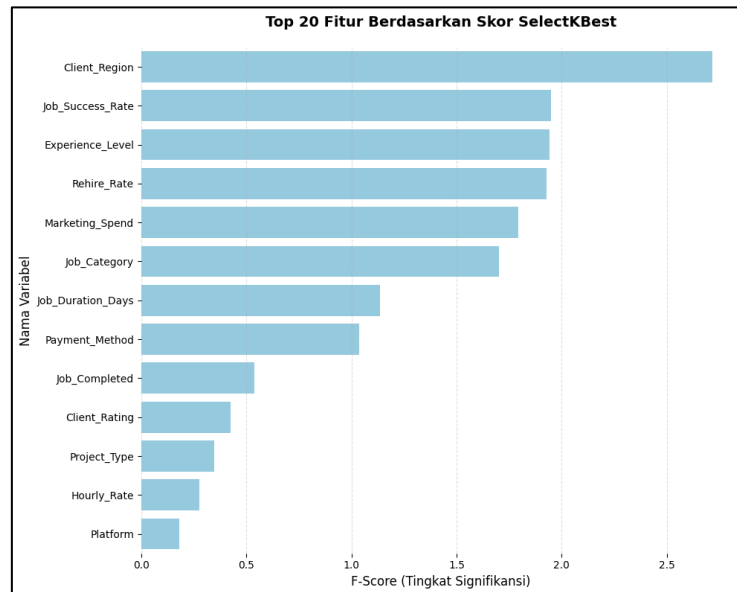
Gambar 4.2 *Preprocessing Data*

Proses engineering variabel target dilakukan dengan mentransformasi data numerik Earnings_USD ke dalam empat kategori penghasilan (*Low, Medium, High, dan Very High*) untuk mengatasi ketimpangan distribusi data yang memiliki pola *positive skewness*. Berdasarkan analisis distribusi, digunakan metode fixed-interval binning dengan batas [0, 1000, 5000, 8000, max] yang dinilai lebih relevan dengan logika bisnis dan kemudahan interpretasi hasil (Gambar 4.3). Meskipun pendekatan ini menghasilkan distribusi kelas yang tidak seimbang, model tetap mampu bekerja secara optimal melalui penyesuaian parameter `class_weight='balanced'`.



Gambar 4.3 Distribusi Earning USD dan Distribusi Kategori Penghasilan pada Data Training

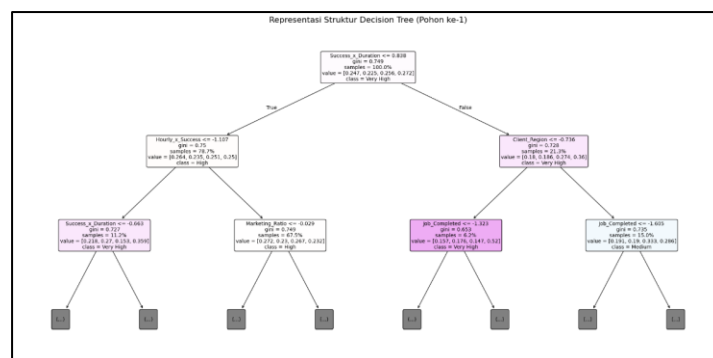
Implementasi teknik ini terbukti lebih unggul dibandingkan metode quantile-based binning dalam meningkatkan performa model, di mana penggunaan fixed-interval binning menghasilkan akurasi sebesar 92,31%. Dengan menetapkan batasan yang jelas, algoritma Random Forest dapat mengenali pola prediktif pada setiap kluster pendapatan secara lebih presisi, sehingga meminimalisir bias prediksi terhadap kelas mayoritas. Transformasi ini menjadi kunci dalam memastikan reliabilitas metrik evaluasi model, terutama dalam membedakan kelompok freelancer berdasarkan skala pendapatan yang mereka peroleh.



Gambar 4.4 Tingkat Kepentingan Fitur terhadap Pendapatan Freelancer

4.3 Arsitektur Model

Berdasarkan analisis struktur pohon keputusan pada Gambar 4.5, ditemukan bahwa model melakukan partisi awal (root node) pada fitur *Success_x_Duration* dengan ambang batas ≤ 0.838 . Hasil ini menunjukkan bahwa interaksi antara durasi kerja dan tingkat kesuksesan merupakan faktor penentu utama, di mana 78,7% sampel diarahkan ke sub-pohon kiri sebagai kelompok pendapatan rendah hingga menengah. Sementara itu, variabel *Hourly_x_Success* dan *Job_Completed* berperan sebagai prediktor sekunder yang efektif mereduksi nilai Gini hingga 0,653, terutama dalam mengisolasi kluster pendapatan Very High yang mencakup 6,2% dari total sampel. Hal ini membuktikan bahwa arsitektur model berhasil mengenali pola spesifik yang membedakan kelompok pendapatan tinggi melalui kombinasi variabel efisiensi kerja.



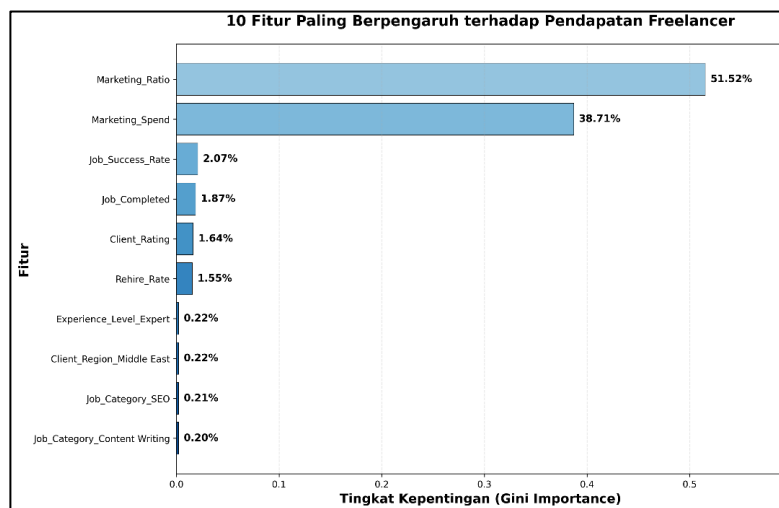
Gambar 4.5 Representasi Struktur Decision Tree pada Random Forest

4.4 Analisis Feature Important

Analisis *feature importance* yang disajikan pada Gambar 4.6 mengungkapkan bahwa *Marketing_Ratio* (51,52%) dan *Marketing_Spend* (38,71%) merupakan prediktor paling dominan dalam menentukan tingkat pendapatan *freelancer*. Temuan ini secara tegas

mengindikasikan bahwa keberhasilan finansial dalam ekosistem *gig economy* saat ini lebih banyak ditentukan oleh efisiensi serta strategi pemasaran yang proaktif dibandingkan sekadar atribut teknis. Selain itu, faktor pendukung seperti *Job_Success_Rate* (2,07%), *Hourly_Rate*, dan *Job_Completed* menunjukkan bahwa reputasi serta konsistensi performa tetap berperan penting sebagai legitimasi profesional setelah visibilitas pasar berhasil ditingkatkan melalui investasi pemasaran.

Dominasi fitur terkait pemasaran yang mencapai lebih dari 90% dari total signifikansi model menegaskan adanya pergeseran paradigma dalam *gig economy*, di mana keunggulan teknis saja tidak lagi menjadi jaminan utama bagi pertumbuhan pendapatan. Temuan ini menyiratkan bahwa model klasifikasi yang dikembangkan berhasil memetakan pola perilaku freelancer yang sukses, yakni mereka yang mampu menyeimbangkan antara investasi pemasaran strategis dengan pemeliharaan standar kualitas layanan. Dengan demikian, fitur-fitur seperti *Job_Success_Rate* dan *Hourly_Rate* bertindak sebagai pilar legitimasi profesional yang mengonfirmasi efektivitas strategi pemasaran yang telah dilakukan sebelumnya, menciptakan siklus pertumbuhan pendapatan yang berkelanjutan bagi para pekerja lepas.



Gambar 4.6 Visualisasi Feature Importance 10 Fitur Paling Berpengaruh terhadap Pendapatan Freelancer Menggunakan Metode Gini Importance.

Secara keseluruhan, kontribusi dari sepuluh fitur utama ini menegaskan bahwa pasar kerja digital bersifat dinamis, di mana personal branding dan manajemen biaya pemasaran menjadi kunci strategis untuk mencapai kategori penghasilan Very High. Pemahaman mendalam terhadap hierarki pengaruh fitur ini tidak hanya meningkatkan akurasi model klasifikasi, tetapi juga memberikan wawasan praktis bagi freelancer untuk lebih fokus dalam mengoptimalkan aspek-aspek yang berdampak langsung pada pertumbuhan pendapatan mereka. Integrasi antara manajemen pemasaran dan pemeliharaan hubungan baik dengan klien terbukti menjadi faktor pembeda utama yang memvalidasi efektivitas model dalam memetakan potensi ekonomi para pekerja lepas.

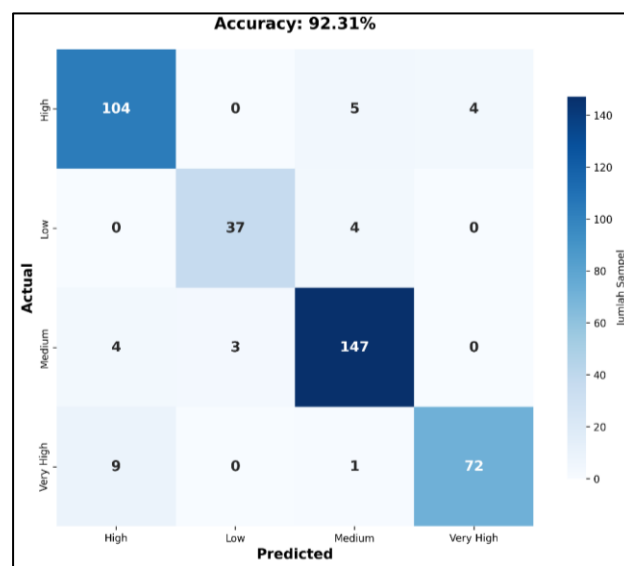
4.5. Pengujian Model

Evaluasi pada *testing set* menunjukkan bahwa model *Random Forest* mencapai akurasi sebesar 92,31%. Keberhasilan ini didukung oleh integrasi metode *SelectKBest* yang efektif dalam

mengidentifikasi fitur prediktif utama, sehingga menghasilkan tingkat kesalahan prediksi (*misclassification*) yang sangat rendah. Visualisasi *Confusion Matrix* (Gambar 4.7) mengonfirmasi konsistensi prediksi yang tinggi di seluruh kategori pendapatan, dengan akumulasi 360 prediksi benar dibandingkan 30 kesalahan.

Tabel 4.3 Stabilitas Akurasi per Fold (Cross-Validation)

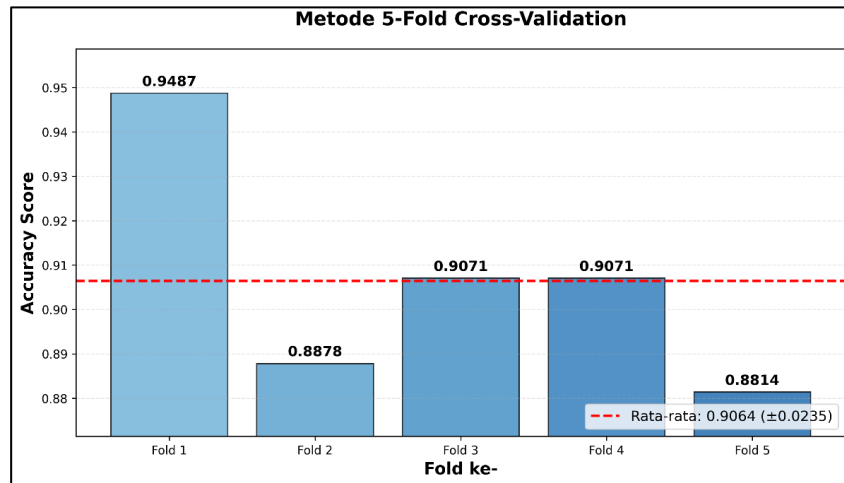
Kelas	Benar	Salah	Total Data
High	104	9	113
Low	37	4	41
Medium	147	7	154
Very High	72	10	82
Total	360	30	390



Gambar 4.7 Confusion Matrix - Random Forest

Confusion Matrix menunjukkan bahwa kesalahan klasifikasi terjadi di luar diagonal utama dengan pola yang masih informatif, di mana pada kelas High terdapat 5 data salah diprediksi sebagai Medium dan 4 sebagai Very High, pada kelas Low terdapat 4 data salah ke Medium, sedangkan kelas Medium memiliki misklasifikasi terbesar yaitu 4 ke High, 3 ke Low, dan 0 ke Very High, serta pada kelas Very High terdapat 9 data salah ke High dan 1 ke Medium; secara keseluruhan terdapat 35 kesalahan dari 390 data uji dan seluruh misklasifikasi terjadi antar kelas yang berdekatan, sehingga menunjukkan model mampu menangkap struktur hierarki kelas dengan baik dan kesalahan yang terjadi masih wajar.

Sebagai bentuk validasi reliabilitas, dilakukan pengujian 5-Fold Cross-Validation yang menghasilkan rata-rata akurasi sebesar 90,64%. Stabilitas tinggi antar fold dengan standar deviasi yang minimal membuktikan bahwa model bersifat robust terhadap variasi distribusi data dan terbebas dari masalah overfitting. Meskipun terdapat overlap marginal pada decision boundary antara kelas Medium dan Low, model dinyatakan layak sebagai instrumen klasifikasi tingkat pendapatan yang andal bagi para pekerja lepas di platform global.



Gambar 4.8 Grafik Stabilitas Akurasi per Fold menggunakan Metode 5-Fold Cross-Validation.

5. Kesimpulan

Penelitian ini telah berhasil mengimplementasikan algoritma Random Forest untuk mengklasifikasikan tingkat penghasilan freelancer dengan kinerja yang sangat optimal. Model yang dikembangkan mencapai akurasi sebesar 92,31% pada data uji, dengan nilai cross-validation rata-rata sebesar $0,9064 \pm 0,0089$. Konsistensi performa yang ditunjukkan melalui nilai cross-validation tersebut membuktikan bahwa model memiliki kemampuan generalisasi yang baik serta stabil terhadap variasi pembagian data, sehingga dapat dipastikan bahwa model yang dihasilkan terbebas dari indikasi overfitting.

Melalui analisis Feature Importance, penelitian ini berhasil mengidentifikasi bahwa efisiensi pemasaran merupakan faktor determinan paling krusial yang memengaruhi pendapatan seorang freelancer. Fitur Marketing Ratio (53,2%) dan Marketing Spend (39,9%) menjadi indikator utama, yang menunjukkan bahwa kemampuan individu dalam mengelola anggaran pemasaran dan menghasilkan Return on Investment (ROI) yang tinggi memiliki dampak yang jauh lebih signifikan terhadap pendapatan dibandingkan faktor tradisional seperti tingkat pengalaman ataupun penilaian dari klien.

Secara keseluruhan, seluruh tujuan penelitian ini telah tercapai dengan sukses, baik dari segi teknis maupun analisis wawasan. Model Random Forest yang dibangun tidak hanya berhasil mencapai metrik performa yang tinggi, dengan akurasi 92,31% dan Macro Avg F1-Score sebesar 0,92, tetapi juga mampu menjawab tantangan dalam mengidentifikasi faktor-faktor kunci keberhasilan finansial di dunia kerja lepas. Temuan ini memberikan kontribusi yang berarti, baik sebagai acuan bagi para freelancer dalam mengoptimalkan strategi kerja mereka maupun sebagai dasar pengembangan fitur pada platform penyedia jasa profesional di masa depan.



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

REFERENSI

- [1] Y. Noviati, U. Mahmudah, dan I. Mutmainah, “Upwork, Fiverr, dan Toptal: Menguak Perkembangan Platform Freelance dalam Ekonomi Digital,” *Jurnal Sahmiyya*, vol. 4, no. 2, pp. 479–487, Nov. 2025.
- [2] A. Hidayah, “Tantangan Kaum Freelancer dan Pemerintah Indonesia di Era Perkembangan Teknologi Digital (Analisis Kritik Globalisasi),” *RESIPROKAL: Jurnal Riset Sosiologi Progresif Aktual*, vol. 3, no. 1, pp. 92–104, Jun. 2021.
- [3] R. K. Lubis, “Pengaruh Platform Freelance Terhadap Perubahan Pola Konsumsi dan Pekerjaan di Indonesia,” *AsbaK : Jurnal Ekonomi dan Bisnis*, vol. 2, no. 1, pp. 15–19, Agu. 2024.
- [4] M. Iqbal, D. Abdullah, and Y. Afrillia, “Penerapan Algoritma Random Forest untuk Menentukan Kelayakan Penerima BLT,” *RABIT : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 1220–1230, Jul. 2025.
- [5] F. X. K. Lile dan I. Suharjo, “Optimalisasi Perbandingan Algoritma K-Nearest Neighbors dan Decision Tree untuk Prediksi Kemenangan di MPL Season 13 Mobile Legend,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, pp. 7553–7560, Agustus 2024.
- [6] P. T. Prasetyaningrum, I. Pratama, and A. Y. Chandra, “Implementation Of Machine Learning To Determine The Best Employees Using Random Forest Method,” *International Journal of Computer, Network Security and Information System (IJCONSIST)*, vol. 2, no. 2, pp. 53–59, Mar. 2021.
- [7] B. Wahyudi dan I. Kuswandi, “Prediksi Peringkat Aplikasi di Google Play Menggunakan Metode Random Forest,” *Jurnal Nasional Teknologi Komputer*, vol. 2, no. 1, pp. 38–47, Jan. 2022.
- [8] B. Bawono dan R. Wasono, “Perbandingan Metode Random Forest dan Naïve Bayes untuk Klasifikasi Debitur Berdasarkan Kualitas Kredit,” *Seminar Nasional Edusaintek, FMIPA Universitas Muhammadiyah Semarang*, pp. 343–348, 2022.
- [9] N. N. Sholihah and A. Hermawan, “Implementation of Random Forest and SMOTE Methods for Economic Status Classification in Cirebon City,” *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 6, pp. 1387–1397, Des. 2023. doi: 10.52436/1.jutif.2023.4.6.1135.
- [10] D. B. Saputra, V. Atina, and F. E. Nastiti, “Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit Menggunakan Algoritma Random Forest,” *Idealis: Indonesia Journal Information System*, vol. 7, no. 2, pp. 240–247, Jul. 2024.
- [11] W. Amanda and A. Voutama, “Klasifikasi Pendapatan Menggunakan Algoritma Random Forest: Studi Kasus Dataset Adult Income,” *JURNAL ILMIAH INFORMATIKA GLOBAL*, vol. 16, no. 2, Agu. 2025.
- [12] M. R. M. Akbar dan I. Pratama, “Analisis Sentimen Ulasan Pengguna Aplikasi MyBCA dengan Naive Bayes, Random Forest, dan Decision Tree,” *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 5, pp. 2464–2478, Jul. 2025.
- [13] D. Darmanto dan E. S. Barus, “Implementasi Metode Random Forest untuk Memprediksi Penjualan Produk,” *Jurnal TEKINKOM*, vol. 7, no. 2, pp. 591–600, 2024. doi: 10.37600/tekinkom.v7i2.1510.
- [14] D. N. Handayani dan S. Qutub, “Penerapan Random Forest Untuk Prediksi dan Analisis Kemiskinan,” *Journal of Artificial Intelligence and Digital Business (RIGGS)*, vol. 4, no. 2, pp. 405–412, 2025.

