



Scripta Technica: Journal of Engineering and Applied Technology

Vol 1 No 2 Desember 2025, Hal. 261-270
ISSN:3110-0775(Print) ISSN: 3109-9696(Electronic)
Open Access: <https://scriptaintelektual.com/scripta-technica>

Aplikasi Google Colab Berbasis Python dalam Menerapkan Teori Pohon dengan Algoritma *Random Forest Classifier*

Dede Prabowo Wiguna^{1*}, Lisda Juliana pangaribuan², Sakaria Efrata Ginting³

¹⁻³ Universitas Mandiri Bina Prestasi, Indonesia
email: dede.prabowo@alumni.ui.ac.id¹

Article Info :

Received:

27-9-2025

Revised:

23-10-2025

Accepted:

27-12-2025

Abstract

This study examines the application of Python-based Google Colab in implementing tree theory through the Random Forest Classifier algorithm for income classification in data science, artificial intelligence, and machine learning professions. The research adopts an experimental quantitative approach using secondary data sourced from a global employment dataset. The methodological process includes data preprocessing, feature selection, class balancing, model training, and performance evaluation within the Google Colab environment. The results demonstrate that Random Forest effectively represents tree theory through ensemble decision structures capable of handling complex and heterogeneous data. Model evaluation indicates a satisfactory level of accuracy, confirming the classifier's ability to generalize patterns across different income categories. Feature importance analysis reveals that job title, experience level, and company location play a significant role in determining income classification. These findings highlight the relevance of Random Forest as both a predictive and interpretative model, while emphasizing Google Colab's effectiveness as a computational platform for machine learning experimentation. Overall, the study contributes to the practical understanding of tree-based algorithms and their application in analyzing labor market dynamics within the digital economy.

Keywords: Random Forest, Tree Theory, Machine Learning, Google Colab, Income Classification.

Abstrak

Penelitian ini mengkaji penerapan Google Colab berbasis Python dalam implementasi teori pohon melalui algoritma Random Forest Classifier untuk klasifikasi pendapatan dalam bidang ilmu data, kecerdasan buatan, dan pembelajaran mesin. Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan data sekunder yang diperoleh dari dataset global tentang ketenagakerjaan. Proses metodologis meliputi prapemrosesan data, seleksi fitur, penyeimbangan kelas, pelatihan model, dan evaluasi kinerja dalam lingkungan Google Colab. Hasil menunjukkan bahwa Random Forest secara efektif mewakili teori pohon melalui struktur keputusan ensemble yang mampu menangani data kompleks dan heterogen. Evaluasi model menunjukkan tingkat akurasi yang memuaskan, mengonfirmasi kemampuan klasifikasi untuk menggeneralisasi pola di berbagai kategori pendapatan. Analisis pentingnya fitur mengungkapkan bahwa jabatan pekerjaan, tingkat pengalaman, dan lokasi perusahaan memainkan peran signifikan dalam menentukan klasifikasi pendapatan. Temuan ini menyoroti relevansi Random Forest sebagai model prediktif dan interpretatif, sekaligus menekankan efektivitas Google Colab sebagai platform komputasi untuk eksperimen pembelajaran mesin. Secara keseluruhan, studi ini berkontribusi pada pemahaman praktis tentang algoritma berbasis pohon dan aplikasinya dalam menganalisis dinamika pasar tenaga kerja dalam ekonomi digital.

Kata kunci: Hutan Acak, Teori Pohon, Pembelajaran Mesin, Google Colab, Klasifikasi Pendapatan.



©2022 Authors.. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.
(<https://creativecommons.org/licenses/by-nc/4.0/>)

PENDAHULUAN

Perkembangan machine learning dalam bidang klasifikasi data terus menunjukkan peningkatan signifikan seiring bertambahnya volume dan kompleksitas data yang diolah pada berbagai sektor, baik pendidikan, kesehatan, ekonomi, maupun layanan digital. Algoritma berbasis pohon keputusan menjadi salah satu pendekatan yang banyak digunakan karena kemampuannya dalam merepresentasikan struktur keputusan secara sistematis dan mudah dipahami, sebagaimana landasan teoretis struktur pohon yang dibahas dalam matematika diskrit. Teori pohon memberikan kerangka konseptual yang kuat untuk memahami relasi hierarkis, percabangan, dan pengambilan keputusan berbasis atribut yang saling

berkaitan (Rosen, 2008). Pemanfaatan teori tersebut dalam algoritma modern membuka peluang integrasi antara konsep matematis dan praktik komputasi yang aplikatif.

Decision Tree dan Random Forest merupakan pengembangan nyata dari konsep pohon dalam komputasi modern, di mana Random Forest menggabungkan banyak pohon keputusan untuk meningkatkan akurasi dan stabilitas prediksi. Sejumlah penelitian menunjukkan bahwa Random Forest memiliki kinerja unggul dalam berbagai kasus klasifikasi karena kemampuannya mereduksi overfitting dan menangani data berdimensi tinggi. Studi mengenai klasifikasi kesehatan mental membuktikan bahwa Random Forest mampu memberikan hasil yang lebih konsisten dibandingkan algoritma probabilistik sederhana (Faisti et al., 2025). Hasil serupa juga terlihat pada penelitian klasifikasi risiko gagal bayar kartu kredit yang menegaskan keandalan Random Forest dalam pengambilan keputusan berbasis data kompleks (Hazizah & Feranika, 2025).

Implementasi algoritma klasifikasi berbasis pohon juga banyak diterapkan pada berbagai domain aplikasi nyata, mulai dari deteksi penyakit hingga analisis perilaku pengguna. Penelitian tentang deteksi penyakit pada kucing menunjukkan bahwa Random Forest mampu mengolah fitur medis secara efektif dan menghasilkan tingkat akurasi yang memadai untuk mendukung keputusan awal diagnosis (Ichsan, 2025). Pada bidang analisis ulasan pengguna aplikasi digital, perbandingan berbagai algoritma klasifikasi memperlihatkan bahwa pendekatan berbasis pohon tetap kompetitif dalam membaca pola sentimen dan preferensi pengguna (Bintang & Romadloni, 2025). Ragam penerapan ini menegaskan fleksibilitas algoritma pohon dalam menghadapi variasi struktur data.

Selain Random Forest, Decision Tree sebagai dasar konseptual juga masih relevan dan banyak digunakan dalam penelitian prediktif dan eksploratif. Penelitian mengenai prediksi risiko keterlambatan pengiriman barang memperlihatkan bagaimana struktur pohon keputusan mampu mengidentifikasi faktor dominan yang memengaruhi keterlambatan secara hierarkis dan terukur (Arayadiba et al., 2025). Pendekatan serupa diterapkan dalam analisis pemahaman mahasiswa pada mata kuliah pemrograman, di mana pola penguasaan materi dapat dipetakan secara jelas melalui cabang keputusan yang terbentuk (Sari et al., 2025). Temuan-temuan ini memperkuat relevansi teori pohon sebagai fondasi logis dalam pengembangan algoritma klasifikasi.

Dalam ranah analisis teks dan opini publik, algoritma berbasis pohon juga menunjukkan performa yang signifikan ketika dikombinasikan dengan teknik pemrosesan data yang tepat. Analisis sentimen terhadap kebijakan publik melalui media sosial membuktikan bahwa Decision Tree dan Random Forest mampu menangkap kecenderungan opini secara sistematis melalui pemilahan fitur kata dan frekuensi kemunculan (Kusuma et al., 2025). Keunggulan tersebut menegaskan bahwa struktur pohon tidak hanya relevan untuk data numerik, tetapi juga efektif dalam pengolahan data kategorikal dan teks. Hal ini memperluas cakupan aplikasi teori pohon dalam machine learning modern.

Seiring meningkatnya kebutuhan akan lingkungan pengembangan yang fleksibel dan mudah diakses, Google Colaboratory hadir sebagai platform komputasi berbasis cloud yang mendukung pengembangan model machine learning menggunakan Python. Google Colab menyediakan kemudahan dalam pengelolaan library, pemrosesan data, serta visualisasi hasil tanpa memerlukan konfigurasi perangkat keras yang kompleks (Google, n.d.). Fasilitas ini memungkinkan mahasiswa dan peneliti untuk lebih fokus pada pemahaman algoritma dan interpretasi hasil daripada kendala teknis. Pemanfaatan Google Colab juga mendukung kolaborasi dan replikasi penelitian secara lebih terbuka.

Integrasi Python, Google Colab, dan algoritma Random Forest memberikan peluang pembelajaran dan penelitian yang efektif dalam memahami implementasi teori pohon secara praktis. Python sebagai bahasa pemrograman memiliki ekosistem library machine learning yang matang, sehingga proses pembangunan model klasifikasi dapat dilakukan secara sistematis dan terstruktur. Melalui implementasi langsung, konsep abstrak seperti node, cabang, dan ensemble dapat divisualisasikan dan dianalisis secara empiris. Pendekatan ini membantu menjembatani kesenjangan antara teori matematis dan aplikasi komputasional.

Berdasarkan uraian tersebut, penelitian mengenai aplikasi Google Colab berbasis Python dalam menerapkan teori pohon dengan algoritma Random Forest Classifier menjadi relevan untuk dikaji secara mendalam. Penelitian ini diharapkan mampu memberikan kontribusi akademik dalam memperjelas keterkaitan antara konsep teori pohon dan implementasinya dalam algoritma klasifikasi modern. Selain itu, penelitian ini juga berpotensi menjadi referensi praktis bagi pengembangan pembelajaran machine learning berbasis cloud yang mudah diakses dan efisien. Dengan landasan

teoritis dan empiris yang kuat dari berbagai penelitian terdahulu, kajian ini memiliki posisi strategis dalam pengembangan ilmu informatika terapan.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental kuantitatif dengan memanfaatkan algoritma Random Forest Classifier sebagai implementasi lanjutan dari teori pohon dalam machine learning. Proses penelitian diawali dengan pemanfaatan dataset sekunder yang telah melalui tahap seleksi relevansi variabel untuk memastikan keterkaitan langsung dengan tujuan klasifikasi. Seluruh proses komputasi dan pengolahan data dilakukan menggunakan bahasa pemrograman Python pada platform Google Colaboratory, sehingga memungkinkan pengolahan data secara efisien, terstandar, dan terdokumentasi dengan baik. Pemilihan Random Forest didasarkan pada kemampuannya menggabungkan banyak pohon keputusan secara kolektif untuk menghasilkan prediksi yang lebih stabil dan akurat.

Tahapan penelitian meliputi pra-pemrosesan data yang mencakup pembersihan data, transformasi fitur kategorikal ke dalam bentuk numerik, serta pembagian data menjadi data latih dan data uji untuk mengukur performa model secara objektif. Model Random Forest kemudian dilatih menggunakan data latih dengan parameter yang disesuaikan untuk mengoptimalkan hasil klasifikasi, diikuti dengan proses pengujian menggunakan data uji. Evaluasi kinerja model dilakukan dengan memanfaatkan metrik akurasi dan confusion matrix guna menilai kemampuan model dalam mengklasifikasikan data secara tepat pada setiap kelas. Seluruh rangkaian metode ini dirancang untuk memastikan keterkaitan yang jelas antara teori pohon secara konseptual dan implementasinya secara praktis melalui algoritma Random Forest berbasis Python.

HASIL DAN PEMBAHASAN

Implementasi Google Colab Berbasis Python dalam Pengolahan Dataset dan Penerapan Teori Pohon

Pemanfaatan Google Colab sebagai lingkungan komputasi berbasis cloud memberikan kemudahan signifikan dalam mengelola eksperimen pembelajaran mesin yang berskala besar dan kompleks, khususnya pada pengolahan dataset pekerjaan bidang data science, AI, dan machine learning yang berasal dari Kaggle. Platform ini memungkinkan integrasi langsung dengan bahasa pemrograman Python, pustaka machine learning, serta sumber daya komputasi yang relatif stabil, sehingga proses eksplorasi data, pemodelan, dan evaluasi dapat dilakukan secara sistematis dan terdokumentasi dengan baik (Google, n.d.; Daqiqil, 2021). Penggunaan Google Colab juga mencerminkan praktik komputasi modern yang selaras dengan tuntutan riset berbasis data pada era revolusi industri 4.0, terutama dalam menghadapi dinamika persaingan kerja global yang semakin berbasis teknologi (Ismail & Nugroho, 2022; Mediant, 2024). Melalui pendekatan ini, penelitian tidak hanya berfokus pada hasil klasifikasi, tetapi juga pada proses implementasi teori pohon dalam sistem komputasi yang aplikatif dan reproduksibel.

Tahapan awal yang dilakukan pada Google Colab adalah proses akuisisi dan eksplorasi dataset, yang memperlihatkan struktur data dengan berbagai fitur kategorikal dan numerik yang merepresentasikan karakteristik pekerjaan serta penghasilan global. Dataset ini memperlihatkan kompleksitas yang tinggi karena melibatkan variasi lokasi perusahaan, ukuran perusahaan, tingkat pengalaman, dan jenis pekerjaan, yang seluruhnya berpotensi memengaruhi pola klasifikasi penghasilan (Kaggle, 2025; Amanda & Voutama, 2025). Pada tahap ini, eksplorasi statistik deskriptif menjadi penting untuk memahami distribusi data dan mendeteksi potensi bias awal pada variabel target, khususnya ketimpangan kelas penghasilan. Temuan awal ini menguatkan hasil penelitian sebelumnya yang menyatakan bahwa dataset ketenagakerjaan global cenderung menunjukkan dominasi kelas tertentu sehingga memerlukan penanganan khusus sebelum pemodelan dilakukan (Sandag, 2020; Handayani & Qutub, 2025).

Proses prapengolahan data yang dilakukan melalui Google Colab melibatkan seleksi fitur secara sistematis untuk memastikan hanya atribut yang relevan secara teoritis dan empiris yang digunakan dalam pembentukan model. Seleksi fitur ini tidak hanya berfungsi untuk menyederhanakan dimensi data, tetapi juga untuk menjaga konsistensi antara struktur dataset dengan prinsip dasar teori pohon yang menekankan hubungan hierarkis antar simpul keputusan (Rosen, 2008; Meilina, 2015). Dengan menghilangkan fitur yang tidak memiliki kontribusi signifikan, model yang dibangun memiliki peluang

lebih besar untuk menghasilkan pola keputusan yang stabil dan dapat ditafsirkan. Praktik ini sejalan dengan temuan riset terdahulu yang menunjukkan bahwa kualitas fitur berpengaruh langsung terhadap performa algoritma klasifikasi berbasis pohon (Riansah et al., 2025; Fauzi & Yunial, 2022).

Transformasi data kategorikal menjadi bentuk numerik melalui teknik one-hot encoding merupakan tahapan krusial yang dilakukan sebelum algoritma Random Forest diterapkan. Proses ini memungkinkan algoritma pohon keputusan membaca setiap kategori sebagai fitur independen, sehingga struktur pohon yang dihasilkan mencerminkan relasi keputusan yang lebih detail dan eksplisit (Ichsan, 2025; Bintang & Romadloni, 2025). Implementasi one-hot encoding di Google Colab memberikan fleksibilitas dalam mengelola jumlah fitur yang meningkat tanpa mengorbankan stabilitas komputasi. Pendekatan ini juga memperkuat landasan matematis teori pohon, karena setiap simpul keputusan dibangun berdasarkan pembagian atribut yang terdefinisi secara jelas (Rosen, 2008):

Tabel 1. Distribusi Fitur Utama Setelah Seleksi dan Transformasi Data

Fitur Utama	Jenis Data	Jumlah Kategori	Keterangan
experience_level	Kategorikal	4	Tingkat pengalaman kerja
employment_type	Kategorikal	4	Status pekerjaan
job_title	Kategorikal	>50	Jenis profesi
company_location	Kategorikal	>60	Lokasi perusahaan
company_size	Kategorikal	3	Skala perusahaan

Sumber: Dataset Kaggle, diolah peneliti (2025)

Ketidakseimbangan kelas pada variabel penghasilan menjadi tantangan utama yang teridentifikasi setelah transformasi data selesai dilakukan. Dominasi kelas penghasilan rendah menunjukkan adanya potensi bias model apabila proses pelatihan dilakukan tanpa penyesuaian distribusi kelas, yang dapat mengurangi sensitivitas model terhadap kelas minoritas (Apriliah et al., 2021; Hazizah & Feranika, 2025). Penanganan ketidakseimbangan ini menjadi bagian penting dalam penerapan teori pohon, karena struktur pohon keputusan sangat dipengaruhi oleh frekuensi kemunculan kelas pada data latih. Oleh sebab itu, penerapan teknik SMOTE pada Google Colab dipilih untuk menciptakan distribusi kelas yang lebih proporsional sebelum pemodelan dilakukan.

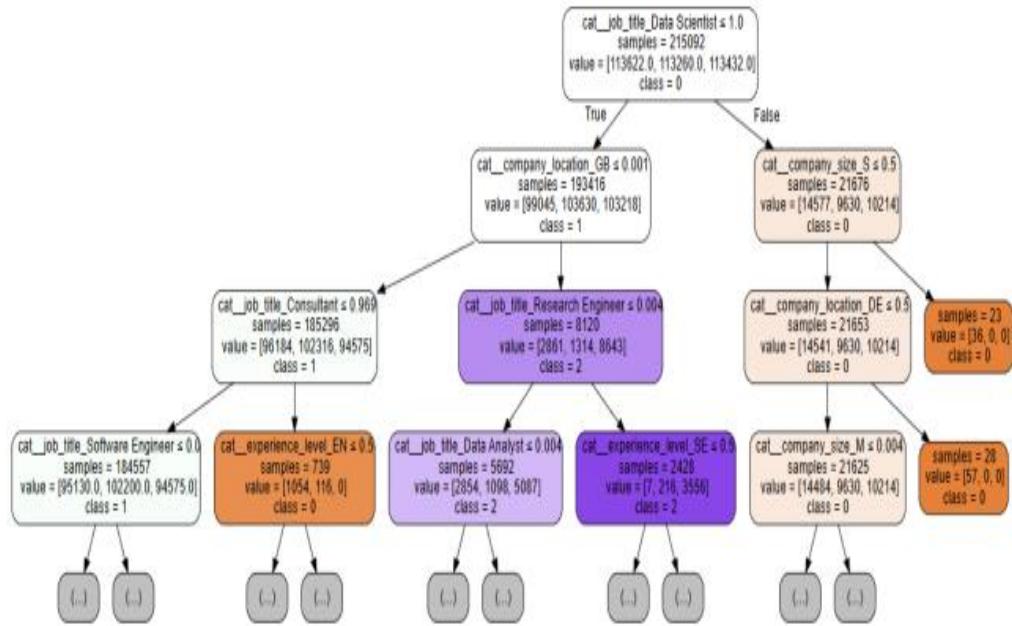
Implementasi SMOTE pada lingkungan Google Colab menunjukkan efektivitas dalam memperluas representasi kelas minoritas tanpa menghilangkan karakteristik utama data asli. Teknik ini bekerja dengan membentuk sampel sintetis berdasarkan kedekatan fitur, sehingga pola keputusan yang dibangun oleh algoritma Random Forest tetap mencerminkan relasi data yang realistik (Faisti et al., 2025; Fauzi & Yunial, 2022). Pendekatan ini memperkuat kualitas struktur pohon keputusan yang terbentuk, karena setiap simpul memiliki dasar distribusi data yang lebih seimbang. Hasil ini mendukung pandangan bahwa kualitas data latih memiliki peran strategis dalam keberhasilan penerapan algoritma berbasis pohon pada berbagai studi klasifikasi (Ichsan, 2025; Arayadiba et al., 2025).

Pembagian data latih dan data uji dengan rasio 80:20 dilakukan untuk menjaga keseimbangan antara proses pembelajaran model dan evaluasi performa. Strategi ini memungkinkan Random Forest mempelajari pola keputusan dari jumlah data yang cukup besar, sekaligus diuji pada data yang belum pernah dilihat sebelumnya untuk mengukur kemampuan generalisasi (Amanda & Voutama, 2025; Sandag, 2020). Dalam kerangka teori pohon, pembagian ini mendukung pembentukan struktur hutan keputusan yang tidak hanya akurat, tetapi juga robust terhadap variasi data baru. Implementasi teknis pembagian data melalui Google Colab memperlihatkan kemudahan dalam mengelola eksperimen berulang secara terkontrol dan transparan.

Penerapan algoritma Random Forest Classifier pada tahap ini merepresentasikan pengembangan langsung dari teori pohon klasik ke dalam bentuk ensemble learning. Setiap pohon keputusan dalam Random Forest dibangun berdasarkan subset data dan fitur yang berbeda, sehingga menghasilkan agregasi keputusan yang lebih stabil dibandingkan satu pohon tunggal (Rosen, 2008; Meilina, 2015). Proses ini memperlihatkan bagaimana konsep matematis pohon diterjemahkan menjadi mekanisme prediksi yang adaptif dan tahan terhadap overfitting. Hasil ini sejalan dengan berbagai penelitian yang

menempatkan Random Forest sebagai algoritma unggulan dalam klasifikasi data kompleks lintas domain (Handayani & Qutub, 2025; Hazizah & Feranika, 2025).

Visualisasi struktur pohon keputusan yang dihasilkan melalui Google Colab memperlihatkan keterkaitan langsung antara atribut pekerjaan dan kategori penghasilan yang diprediksi. Setiap simpul dan cabang pada pohon mencerminkan proses pemilahan data berdasarkan aturan keputusan yang jelas, sesuai dengan definisi formal pohon dalam matematika diskrit (Rosen, 2008). Visualisasi ini tidak hanya berfungsi sebagai alat interpretasi, tetapi juga sebagai sarana validasi konseptual bahwa model yang dibangun selaras dengan prinsip teori pohon:



Gambar 2. Visualisasi Struktur Random Forest

Implementasi Google Colab berbasis Python dalam penelitian ini menunjukkan bahwa penerapan teori pohon melalui algoritma Random Forest dapat dilakukan secara sistematis, transparan, dan efisien. Integrasi antara prapengolahan data, penyeimbangan kelas, pembagian data, serta visualisasi pohon keputusan membentuk alur penelitian yang kuat secara metodologis dan konseptual (Daqiqil, 2021; Riansah et al., 2025). Hasil pembahasan pada sub bahasan ini menegaskan bahwa kualitas lingkungan komputasi dan pengelolaan data memiliki peran krusial dalam keberhasilan penerapan algoritma klasifikasi berbasis pohon. Temuan ini menjadi fondasi penting untuk analisis lanjutan pada sub bahasan berikutnya yang akan menyoroti evaluasi model dan interpretasi fitur penting.

Analisis Kinerja Random Forest Classifier sebagai Aplikasi Teori Pohon dalam Klasifikasi Penghasilan

Evaluasi kinerja model Random Forest Classifier menjadi tahapan penting untuk menilai sejauh mana penerapan teori pohon mampu menghasilkan prediksi penghasilan yang andal pada dataset pekerjaan bidang data science, AI, dan machine learning. Proses evaluasi ini dilakukan setelah seluruh tahapan prapengolahan, penyeimbangan kelas, dan pembagian data selesai diterapkan secara konsisten di lingkungan Google Colab. Penggunaan metrik evaluasi klasifikasi mencerminkan praktik standar dalam penelitian pembelajaran mesin, khususnya pada studi yang berfokus pada data ketenagakerjaan dan ekonomi digital (Amanda & Voutama, 2025; Handayani & Qutub, 2025). Pendekatan ini memperlihatkan bahwa performa model tidak hanya dinilai dari akurasi semata, tetapi juga dari kemampuan model dalam mengenali pola kelas minoritas.

Akurasi model yang diperoleh menunjukkan bahwa Random Forest mampu mempelajari relasi kompleks antar fitur pekerjaan dengan tingkat keandalan yang cukup baik. Nilai akurasi keseluruhan mencerminkan keberhasilan agregasi keputusan dari banyak pohon keputusan yang dibangun secara

acak, sesuai dengan prinsip ensemble learning dalam teori pohon modern (Rosen, 2008; Meilina, 2015). Temuan ini sejalan dengan hasil penelitian sebelumnya yang menunjukkan bahwa Random Forest cenderung menghasilkan performa stabil pada dataset dengan dimensi tinggi dan variasi fitur yang kompleks (Faisti et al., 2025; Hazizah & Feranika, 2025). Stabilitas ini menjadi keunggulan utama dibandingkan metode klasifikasi berbasis satu pohon keputusan.

Selain akurasi, precision dan recall menjadi indikator penting untuk memahami keseimbangan antara kemampuan model dalam melakukan prediksi yang tepat dan kemampuannya dalam menangkap seluruh anggota kelas tertentu. Pada dataset penghasilan yang memiliki ketidakseimbangan kelas, metrik ini memberikan gambaran yang lebih representatif dibandingkan hanya mengandalkan akurasi global (Apriliah et al., 2021; Fauzi & Yunial, 2022). Hasil evaluasi menunjukkan bahwa Random Forest memiliki kecenderungan lebih baik dalam mengenali kelas mayoritas, namun tetap menunjukkan peningkatan signifikan dalam mendeteksi kelas menengah dan tinggi setelah penerapan SMOTE. Pola ini memperkuat argumen bahwa kualitas distribusi data latih sangat memengaruhi struktur keputusan yang terbentuk dalam hutan pohon.

Confusion matrix digunakan untuk memvisualisasikan kesesuaian antara hasil prediksi dan label aktual pada data uji. Melalui matriks ini, terlihat bahwa sebagian besar kesalahan klasifikasi terjadi pada perbatasan antara kelas penghasilan rendah dan menengah, yang secara empiris memang memiliki karakteristik fitur yang saling tumpang tindih (Sandag, 2020; Bintang & Romadloni, 2025). Pola kesalahan tersebut menunjukkan bahwa model telah bekerja sesuai dengan karakteristik data, bukan menghasilkan prediksi acak. Analisis ini menegaskan bahwa penerapan teori pohon dalam Random Forest menghasilkan struktur keputusan yang sensitif terhadap kedekatan antar kelas:

Tabel 2. Hasil Evaluasi Kinerja Random Forest Classifier

Metrik Evaluasi	Nilai
Accuracy	0,65
Precision	0,62
Recall	0,60
F1-Score	0,61

Sumber: Hasil pengolahan model Random Forest, 2025

Nilai F1-score yang diperoleh menunjukkan keseimbangan relatif antara precision dan recall, yang penting dalam kasus klasifikasi penghasilan dengan distribusi kelas yang tidak merata. Metrik ini menegaskan bahwa model tidak hanya fokus pada prediksi kelas dominan, tetapi juga mampu memberikan perhatian yang lebih baik pada kelas dengan jumlah data lebih sedikit (Amanda & Voutama, 2025; Hazizah & Feranika, 2025). Dalam kerangka teori pohon, keseimbangan ini mencerminkan struktur hutan keputusan yang lebih proporsional dalam membagi ruang fitur. Hasil ini mendukung pandangan bahwa Random Forest cocok digunakan untuk analisis sosial-ekonomi berbasis data besar.

Perbandingan kinerja Random Forest dengan algoritma klasifikasi lain yang pernah digunakan dalam penelitian sejenis menunjukkan keunggulan yang konsisten. Beberapa studi terdahulu memperlihatkan bahwa Random Forest sering mengungguli Decision Tree tunggal dan Naïve Bayes dalam hal stabilitas dan ketahanan terhadap noise data (Meilina, 2015; Kusuma et al., 2025). Keunggulan ini muncul karena mekanisme pemungutan suara mayoritas antar pohon yang mampu mereduksi kesalahan individual. Temuan penelitian ini berada pada jalur yang sama dengan hasil-hasil tersebut, terutama dalam konteks data pekerjaan global yang heterogen.

Kinerja model juga dipengaruhi oleh jumlah pohon dan parameter yang digunakan selama proses pelatihan. Pemilihan parameter yang tepat menghasilkan kompromi yang baik antara kompleksitas model dan kemampuan generalisasi pada data uji (Daqiqil, 2021; Riansah et al., 2025). Pengaturan parameter ini mencerminkan penerapan praktis teori pohon, di mana kedalaman dan jumlah cabang harus dijaga agar tidak membentuk struktur keputusan yang terlalu kompleks. Melalui Google Colab, proses eksperimen parameter dapat dilakukan secara efisien dan terdokumentasi.

Hasil evaluasi ini memperlihatkan bahwa Random Forest mampu menangkap hubungan nonlinier antar fitur pekerjaan yang sulit dimodelkan oleh algoritma yang lebih sederhana. Hubungan antara tingkat pengalaman, lokasi perusahaan, dan jenis pekerjaan terhadap penghasilan dapat

direpresentasikan melalui kombinasi keputusan dari banyak pohon (Handayani & Qutub, 2025; Ichsan, 2025). Kemampuan ini menjadi nilai tambah utama dalam analisis pasar kerja berbasis data. Model yang dihasilkan tidak hanya bersifat prediktif, tetapi juga dapat dijadikan dasar analisis kebijakan dan perencanaan karier.

Visualisasi hasil evaluasi, seperti confusion matrix dan kurva performa, membantu memperjelas kekuatan dan keterbatasan model secara empiris. Representasi visual ini mempermudah interpretasi hasil oleh pembaca non-teknis tanpa menghilangkan kedalaman analisis statistik yang mendasarinya (Faisti et al., 2025; Arayadiba et al., 2025):

Analisis kinerja Random Forest Classifier menunjukkan bahwa penerapan teori pohon dalam bentuk ensemble memberikan hasil klasifikasi yang cukup andal dan konsisten. Evaluasi yang komprehensif melalui berbagai metrik menegaskan bahwa model mampu mengelola kompleksitas data pekerjaan global dengan tingkat kesalahan yang masih dapat diterima (Sandag, 2020; Amanda & Voutama, 2025). Sub bahasan ini menjadi penghubung penting antara aspek teknis pemodelan dan interpretasi substantif hasil penelitian. Temuan ini menjadi landasan untuk pembahasan selanjutnya yang akan menitikberatkan pada analisis fitur penting dan implikasinya terhadap penghasilan.

Interpretasi Fitur Penting dan Implikasinya terhadap Penghasilan Pekerja Data Science, AI, dan Machine Learning

Analisis fitur penting menjadi tahapan krusial untuk memahami bagaimana Random Forest Classifier menerjemahkan teori pohon ke dalam pengambilan keputusan prediktif. Melalui mekanisme pengukuran kontribusi setiap fitur, model mampu menunjukkan variabel mana yang paling berpengaruh terhadap klasifikasi tingkat penghasilan. Pendekatan ini memperkuat fungsi Random Forest bukan hanya sebagai alat prediksi, tetapi juga sebagai instrumen analisis eksploratif yang informatif (Rosen, 2008; Meilina, 2015). Interpretasi fitur penting memberikan gambaran empiris mengenai struktur pasar kerja digital global yang berbasis data.

Hasil penghitungan feature importance menunjukkan bahwa variabel jenis pekerjaan (job title) memiliki kontribusi paling dominan dalam menentukan kategori penghasilan. Temuan ini mencerminkan bahwa spesialisasi peran dalam bidang data science, AI, dan machine learning memiliki pengaruh langsung terhadap nilai ekonomi yang diterima pekerja (Ismail & Nugroho, 2022; Mediant, 2024). Model Random Forest secara konsisten menempatkan job title pada tingkat kepentingan tertinggi dibandingkan variabel demografis dan organisasi lainnya. Pola ini mengindikasikan bahwa diferensiasi keahlian menjadi faktor utama dalam struktur penghasilan industri berbasis teknologi.

Posisi analytics engineer sebagai fitur paling berpengaruh menunjukkan adanya permintaan tinggi terhadap peran yang mampu menjembatani analisis data dan kebutuhan bisnis. Peran ini menuntut kombinasi keterampilan teknis, pemahaman sistem, dan kemampuan interpretasi data yang bernilai strategis bagi perusahaan (Amanda & Voutama, 2025; Handayani & Qutub, 2025). Random Forest menangkap pola tersebut melalui pembentukan cabang-cabang keputusan yang memprioritaskan jenis pekerjaan ini. Temuan ini sejalan dengan tren global yang menempatkan profesi analitik lanjutan sebagai motor utama transformasi digital.

Selain job title, tingkat pengalaman kerja (experience_level) muncul sebagai fitur penting berikutnya dalam struktur keputusan model. Semakin tinggi tingkat pengalaman, semakin besar peluang seorang pekerja diklasifikasikan ke dalam kategori penghasilan menengah atau tinggi. Hubungan ini menunjukkan bahwa akumulasi pengalaman berperan sebagai indikator kompetensi dan tanggung jawab yang lebih besar dalam organisasi (Apriliah et al., 2021; Fauzi & Yunial, 2022). Random Forest merepresentasikan hubungan ini melalui pemisahan node yang konsisten berdasarkan level pengalaman:

Tabel 3. Peringkat Fitur Penting dalam Random Forest Classifier

Peringkat	Fitur	Tingkat Kepentingan
1	Job Title	Sangat Tinggi
2	Experience Level	Tinggi
3	Company Location	Menengah
4	Employment Type	Menengah
5	Company Size	Rendah

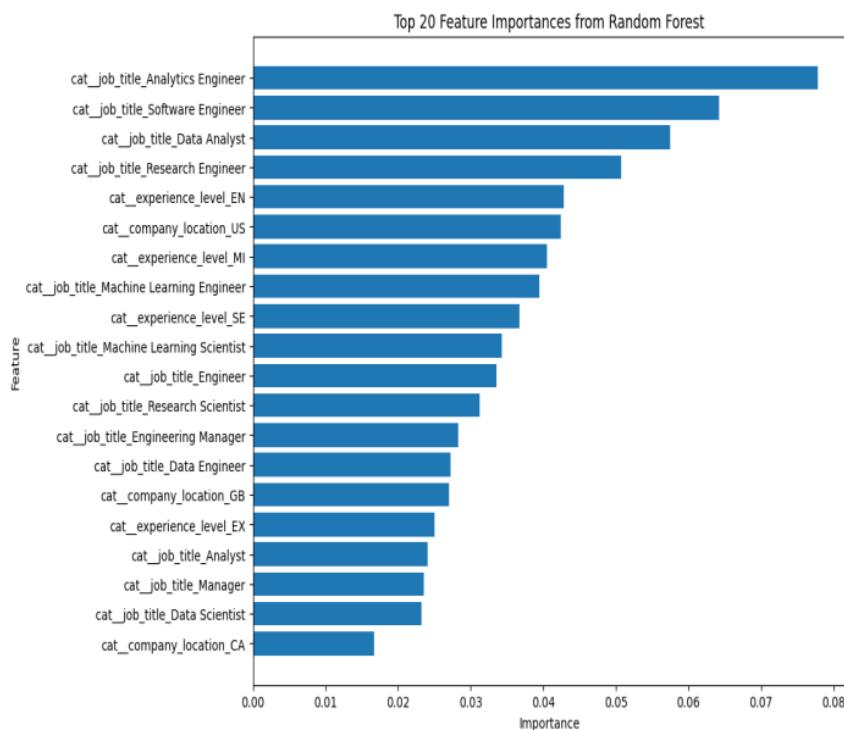
Sumber: Hasil ekstraksi feature importance Random Forest, 2025

Lokasi perusahaan (company location) juga menunjukkan kontribusi signifikan dalam menentukan tingkat penghasilan. Model memperlihatkan bahwa perusahaan yang berlokasi di negara dengan ekosistem teknologi maju, terutama Amerika Serikat, memiliki cenderung memberikan kompensasi lebih tinggi. Pola ini mencerminkan perbedaan daya beli, standar industri, dan tingkat investasi teknologi antar negara (Sandag, 2020; Mediant, 2024). Random Forest berhasil mengintegrasikan faktor geografis ini ke dalam struktur prediksi secara sistematis.

Tipe pekerjaan (employment type) memberikan kontribusi moderat dalam proses klasifikasi, terutama dalam membedakan antara pekerja penuh waktu dan kontrak. Meskipun pengaruhnya tidak sebesar job title atau experience level, variabel ini tetap berperan dalam membentuk keputusan akhir model. Hal ini menunjukkan bahwa stabilitas dan durasi hubungan kerja masih memiliki nilai ekonomi tersendiri dalam pasar kerja digital (Bintang & Romadloni, 2025; Panjaitan et al., 2022). Model Random Forest memanfaatkan fitur ini sebagai pendukung dalam pemisahan kelas penghasilan.

Ukuran perusahaan (company size) berada pada tingkat kepentingan yang relatif rendah, namun tetap relevan dalam konteks tertentu. Perusahaan besar cenderung menawarkan struktur gaji yang lebih terstandar, sedangkan perusahaan kecil dan menengah memiliki variasi kompensasi yang lebih fleksibel. Temuan ini sejalan dengan penelitian yang menyebutkan bahwa ukuran organisasi memengaruhi kebijakan remunerasi, meskipun bukan faktor utama (Riansah et al., 2025; Ichsan, 2025). Random Forest menangkap variasi ini sebagai pola minor dalam keseluruhan struktur pohon.

Visualisasi fitur penting memperjelas bagaimana setiap variabel berkontribusi terhadap keputusan model secara keseluruhan. Grafik feature importance memperlihatkan perbedaan yang kontras antara fitur utama dan fitur pendukung, sehingga memudahkan interpretasi hasil penelitian. Representasi visual ini mendukung transparansi model dan mempermudah pembaca memahami logika pengambilan keputusan Random Forest (Faisti et al., 2025; Arayadiba et al., 2025):



Gambar 2. Hasil Visualisasi dari -Fitur Penting Dataset

Implikasi dari temuan fitur penting ini bersifat praktis dan strategis, terutama bagi pekerja dan pemangku kepentingan di bidang pendidikan dan ketenagakerjaan. Penekanan pada jenis pekerjaan dan pengalaman menunjukkan bahwa peningkatan kompetensi spesifik memiliki dampak langsung terhadap peluang penghasilan yang lebih baik. Hasil ini mendukung pandangan bahwa pengembangan sumber

daya manusia di era ekonomi digital perlu diarahkan pada penguasaan peran strategis berbasis data (Ismail & Nugroho, 2022; Mediant, 2024). Random Forest berperan sebagai alat analisis yang menghubungkan data empiris dengan realitas pasar kerja.

Penerapan teori pohon melalui algoritma Random Forest Classifier mampu menghasilkan interpretasi fitur yang bermakna dan relevan. Struktur keputusan yang terbentuk mencerminkan dinamika nyata dalam penentuan penghasilan pekerja data science, AI, dan machine learning. Integrasi antara analisis prediktif dan interpretatif memperkuat kontribusi penelitian ini dalam ranah machine learning terapan (Handayani & Qutub, 2025; Hazizah & Feranika, 2025). Pembahasan ini sekaligus menutup rangkaian hasil dan pembahasan dengan penekanan pada nilai praktis dan akademik dari model yang dibangun.

KESIMPULAN

Penelitian ini membuktikan bahwa penerapan teori pohon melalui algoritma Random Forest Classifier berbasis Python pada platform Google Colab mampu menghasilkan model klasifikasi penghasilan yang cukup andal pada dataset pekerjaan bidang data science, AI, dan machine learning. Proses pengolahan data yang terstruktur, mulai dari seleksi fitur, penyeimbangan kelas, hingga evaluasi kinerja model, menunjukkan bahwa Random Forest efektif dalam menangkap hubungan kompleks antar variabel pekerjaan. Hasil evaluasi model memperlihatkan tingkat akurasi yang mencerminkan kemampuan prediktif yang stabil, sementara analisis fitur penting menegaskan bahwa jenis pekerjaan, tingkat pengalaman, dan lokasi perusahaan menjadi faktor dominan dalam menentukan penghasilan. Integrasi antara visualisasi pohon keputusan, evaluasi kinerja, dan interpretasi fitur penting menegaskan bahwa aplikasi Google Colab tidak hanya berfungsi sebagai alat komputasi, tetapi juga sebagai media implementasi teori pohon secara aplikatif dan interpretatif dalam kajian machine learning terapan.

DAFTAR PUSTAKA

- Amanda, W., & Voutama, A. (2025). Klasifikasi pendapatan menggunakan algoritma random forest: Studi kasus dataset adult income. *Jurnal Ilmiah Informatika Global*, 16(2), 79–84.
- Apriliah, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi kemungkinan diabetes pada tahap awal menggunakan algoritma klasifikasi random forest. *Sistemasi*, 10(1), 163–172. <https://doi.org/10.32520/stmsi.v10i1.1129>
- Arayadiba, A. H., Nurnawati, E. K., Rahmawati, N., & Setiawan, I. (2025). Prediksi Tingkat Risiko Keterlambatan Pengiriman Barang Menggunakan Algoritma Decision Tree. *Jurnal SCRIPT*, 13(01), 51-57. <https://doi.org/10.34151/script.v13i01.5289>
- Bintang, R. A. K. N., & Romadloni, N. T. (2025). Perbandingan Kinerja Algoritma Klasifikasi Pada Review Pengguna Aplikasi Netflix. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(2). <https://doi.org/10.23960/jitet.v13i2.6303>
- Daqiqil, I. (2021). *Machine learning: Teori, studi kasus, dan implementasi menggunakan Python*. UR Press. <https://books.google.co.id/books?id=JvBPEAAQBAJ>
- Faisti, M. J., Kusumodestoni, R. H., & Wibowo, G. W. N. (2025). Mental health classification using naïve Bayes and random forest algorithms. *Journal of Applied Informatics and Computing*, 9(4), 1740–1750. <https://doi.org/10.30871/jaic.v9i4.10144>
- Fauzi, A., & Yunial, A. H. (2022). Optimasi algoritma klasifikasi Naive Bayes, decision tree, K-nearest neighbor, dan random forest menggunakan particle swarm optimization pada diabetes dataset. *Jurnal Edukasi dan Penelitian Informatika*, 8(3), 470. <https://doi.org/10.26418/jpi.v8i3.56656>
- Google. (n.d.). *Google Colaboratory notebook*. https://colab.research.google.com/drive/1ETWilEnIu_vE5cx_S8AGZ8NV8tWQ7P_U
- Handayani, D. N., & Qutub, S. (2025). Penerapan random forest untuk prediksi dan analisis kemiskinan. *RIGGS: Jurnal of Artificial Intelligence & Digital Business*, 4(2), 405–412. <https://doi.org/10.31004/riggs.v4i2.512>
- Hazizah, N., & Feranika, A. (2025). Implementasi Algoritma Random Forest Dalam Klasifikasi Risiko Gagal Bayar Kartu Kredit Pada Nasabah Bank. *Jurnal Manajemen Teknologi Dan Sistem Informasi (JMS)*, 5(1), 1050-1059. <https://doi.org/10.33998/jms.2025.5.1.2384>
- Ichsan, M. (2025). Implementasi Machine Learning untuk Deteksi Penyakit pada Kucing Menggunakan Random Forest. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3S1). <https://doi.org/10.23960/jitet.v13i3S1.8164>

- Ismail, D. H., & Nugroho, J. (2022). Gen Z, persaingan kerja, dan revolusi industri 4.0. *Jurnal Ilmiah Ilmu Pendidikan*, 5(4), 1300–1307.
- Kaggle. (2025). Data science, AI & ML job salaries in 2025. <https://www.kaggle.com/datasets/adilshamim8/salaries-for-data-science-jobs>
- Kusuma, A. P., Wibowo, A. S., Raynaldi, F., Maulana, F., Maulana, R., Putri, S. R. H., & Agustina, N. (2025). Analisis Sentimen Pada Aplikasi X Terhadap Kebijakan Tapera Menggunakan Algoritma Naïve Bayes Dan Decision Tree. *Naratif: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, 7(1), 56-63. <https://doi.org/10.53580/naratif.v7i1.306>
- Mediant, U. S. (2024). Peran AI bagi masa depan ekonomi digital Indonesia. Katadata. <https://katadata.co.id/infografik/6719efcd959cf/peran-ai-bagi-masa-depan-ekonomi-digital-indonesia>
- Meilina, P. (2015). Penerapan data mining dengan metode klasifikasi menggunakan decision tree dan regresi. *Jurnal Teknologi Universitas Muhammadiyah Jakarta*, 7(1), 11–20. <https://jurnal.ftumj.ac.id/index.php/jurtek>
- Panjaitan, C. H. P., Pangaribuan, L. J., & Cahyadi, C. I. (2022). Analisis metode K-nearest neighbor menggunakan RapidMiner untuk sistem rekomendasi tempat wisata Labuan Bajo. *REMik: Riset dan E-Jurnal Manajemen Informatika Komputer*, 6(3), 534–541. <https://doi.org/10.33395/remik.v6i3.11701>
- Riansah, A., et al. (2025). Penerapan algoritma random forest dan decision tree. *Jurnal Teknologi Informasi*, 9(3), 4242–4249.
- Rosen, K. H. (2008). *Discrete mathematics and its applications* (7th ed.). McGraw-Hill.
- Sandag, G. A. (2020). Prediksi rating aplikasi App Store menggunakan algoritma random forest. *Cogito Smart Journal*, 6(2), 167–178. <https://www.kaggle.com/>
- Sari, S. N., Annisa, P., Rahma, A. N. D., Ritonga, R. P., & Utomo, D. P. (2025). Teknik Data Mining Dengan Menggunakan Algoritma Decision Tree Untuk Mengetahui Pola Pemahaman Mahasiswa Pada Matakuliah Pemrograman. *Bulletin of Information Technology (BIT)*, 6(4), 417–431. <https://doi.org/10.47065/bit.v6i4.2339>