ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

Klasifikasi Pendapatan Menggunakan Algoritma Random Forest: Studi Kasus Dataset Adult Income

Widia Amanda[1)], **Apriade Voutama**[2)]

[1), 2)] *Sistem Informasi, Universitas Singaperbangsa Karawang*
Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361
Email : widiaamanda03@gmail.com[1)], *apriade.voutama@staff.unsika.ac.id*[2)]

ABSTRACT

This research aims to classify a person's income based on demographic attributes using Random Forest algorithm, which is one of the popular ensemble learning methods in the field of machine learning. The dataset used is Adult Income from the UCI Machine Learning Repository, which consists of more than 32 thousand data with 15 attributes such as age, gender, education, education level, employment type, marital status and others. The research process includes data preprocessing, model pipeline creation, training, and performance evaluation. Preprocessing was done through the removal of irrelevant attributes, normalization of numerical data, and application of one-hot encoding on categorical data. The model was trained with default parameters and evaluated using accuracy, precision, recall, F1-score, and confusion matrix metrics. The evaluation results show that the model achieved an accuracy of 85.44%, with higher performance in classifying income classes ≤50K than >50K. The low recall value in the >50K class indicates that the model tends to be biased towards the majority class, which could be caused by data imbalance. Therefore, it is necessary to improve the model through hyperparameter tuning techniques, handling data imbalance, or exploring other algorithms such as Gradient Boosting. This research is expected to be the basis for developing accurate and applicable data-based prediction systems in the fields of economics, policy planning, and decision support systems that require analysis of individual income potential.

Keywords : *Machine Learning, Random Forest, Income Prediction, Adult Income Dataset*

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan pendapatan seseorang berdasarkan atribut demografis menggunakan algoritma Random Forest, yang merupakan salah satu metode ensemble learning populer dalam bidang machine learning. Dataset yang digunakan adalah Adult Income dari UCI Machine Learning Repository, yang terdiri atas lebih dari 32 ribu data dengan 15 atribut seperti usia, jenis kelamin, pendidikan, tingkat pendidikan, jenis pekerjaan, status pernikahan dan lain-lain. Proses penelitian meliputi tahapan preprocessing data, pembuatan pipeline model, pelatihan, dan evaluasi kinerja. Preprocessing dilakukan melalui penghapusan atribut yang tidak relevan, normalisasi data numerik, serta penerapan one-hot encoding pada data kategorikal. Model dilatih dengan parameter default dan dievaluasi menggunakan metrik akurasi, precision, recall, F1-score, serta confusion matrix. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 85,44%, dengan performa yang lebih tinggi dalam mengklasifikasikan kelas pendapatan ≤50K dibandingkan >50K. Nilai recall yang rendah pada kelas >50K menunjukkan bahwa model cenderung bias terhadap kelas mayoritas, yang dapat disebabkan oleh ketidakseimbangan data. Untuk itu, diperlukan peningkatan model melalui teknik hyperparameter tuning, penanganan data imbalance, atau eksplorasi algoritma lain seperti Gradient Boosting. Penelitian ini diharapkan dapat menjadi dasar dalam pengembangan sistem prediksi berbasis data yang akurat dan aplikatif dalam bidang ekonomi, perencanaan kebijakan, serta sistem pendukung keputusan yang memerlukan analisis terhadap potensi pendapatan individu

Kata Kunci : *Machine Learning, Random Forest, Prediksi Pendapatan, Adult Income Dataset*

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

1. Pendahuluan

Prediksi pendapatan individu merupakan salah satu tugas penting dalam bidang analisis data demografi yang memiliki beragam manfaat praktis. Informasi mengenai pendapatan seseorang tidak hanya bermanfaat dalam konteks ekonomi makro, tetapi juga sangat relevan untuk mendukung perencanaan kebijakan sosial, strategi pemasaran yang lebih tertarget, serta pengambilan keputusan dalam berbagai sektor publik maupun swasta. Dalam konteks pemerintahan, data pendapatan dapat digunakan untuk merancang kebijakan subsidi atau bantuan sosial secara lebih tepat sasaran. Sementara itu, di sektor komersial, data ini dapat dimanfaatkan untuk melakukan segmentasi pasar dan menentukan strategi pemasaran yang lebih efektif.

Dalam bidang ilmu data dan kecerdasan buatan, klasifikasi merupakan salah satu metode utama dalam menganalisis dan menginterpretasikan data. Klasifikasi didefinisikan sebagai proses menemukan model atau fungsi yang dapat membedakan antara kelas data. Tujuan utamanya adalah untuk memprediksi kelas suatu objek yang labelnya tidak diketahui sebelumnya. (Samudra et al., 2022). Random Forest adalah salah satu algoritma klasifikasi yang paling umum digunakan. Ini adalah pengklasifikasi yang membuat lebih dari satu pohon keputusan dengan menggunakan subset sampel dan variabel pelatihan yang dipilih secara acak untuk meningkatkan akurasi. (Chyntia Morama et al., 2022).

Random Forest adalah sebuah metode ensemble yang memanfaatkan kumpulan pohon keputusan untuk membentuk sebuah "hutan. ". Hal ini bertujuan untuk meningkatkan nilai akurasi secara optimal (Nurfitriana et al., 2024). Algoritma ini biasanya digunakan untuk menyelesaikan berbagai masalah, baik dalam klasifikasi maupun regresi (Almufqi & Voutama, 2023).

Sebagai salah satu teknik dalam machine learning, Random Forest dapat dipandang sebagai kumpulan berbagai pohon keputusan. Tujuan dari metode ini adalah untuk meningkatkan keakuratan dalam mengklasifikasikan data dengan mengkombinasikan berbagai jenis metode yang sama melalui proses penyesuaian untuk menghasilkan prediksi akhir (Azhima et al., 2023). Random Forest juga dapat memperbaiki hasil prediksi jika dibandingkan dengan banyak metode pembelajaran terawasi lainnya, khususnya pohon keputusan. Dalam model Random Forest, sejumlah teknik penggabungan pohon diperkenalkan untuk mengurangi keterkaitan antara pohon keputusan dalam sebuah hutan (Kurniawan et al., 2023).

Salah satu contoh dataset yang sering digunakan untuk klasifikasi dengan Random Forest adalah Adult Income Dataset. Dataset ini digunakan untuk mengklasifikasikan seseorang berdasarkan pendapatannya, apakah lebih dari $50.000 per tahun atau tidak, dengan mempertimbangkan sejumlah variabel demografis dan pekerjaan. Data ini membantu dalam memahami tingkat kemakmuran di masyarakat Amerika Serikat serta memberikan wawasan bagi pemerintah dalam merancang kebijakan ekonomi (Clara et al., 2021).

Berbagai penelitian telah dilakukan untuk mengeksplorasi efektivitas Random Forest dalam berbagai konteks. Misalnya, penelitian oleh (Fitri, 2023) menunjukkan bahwa Random Forest Regression mencapai akurasi tertinggi sebesar 81,5% dalam memprediksi harga rumah dibandingkan metode regresi lainnya. (Prasojo & Haryatmi, 2021) menemukan bahwa model Random Forest mencapai akurasi sebesar 83% dalam memprediksi kelayakan kredit. Penelitian oleh (Fabian Azmi et al., 2024) menunjukkan bahwa model Random Forest lebih akurat dari pada Decision Tree dalam memprediksi hilangnya pelanggan bank dengan 78%.

Selanjutnya, penelitian oleh (Arisusanto et al., 2023) mengklasifikasikan harga handphone menggunakan Random Forest dengan akurasi 88,50%. (Sinambela et al., 2023) menemukan bahwa Random Forest lebih efektif dalam memprediksi perdarahan pascasalin dengan akurasi 83%. Penelitian oleh (Irma Purnamasari & Ali, 2024) menunjukkan bahwa kedua algoritma, Decision Tree dan Random Forest, mencapai akurasi 97,86% dalam mengidentifikasi penerima bantuan social di masa depan. Penelitian oleh (Rosta Br Sebayang et al., 2023) mengklasifikasikan data kesehatan mental di industri teknologi menggunakan algoritma Random Forest, yang memiliki tingkat akurasi mencapai 84%. Hasil penelitian ini menunjukkan bahwa metode Random Forest dapat digunakan secara efektif untuk menemukan gangguan kesehatan mental di antara karyawan yang bekerja di industri teknologi.

Terakhir, penelitian oleh (Yoga Religia et al., 2021) yang mengklasifikasikan data Bank Marketing menggunakan Random Forest dan optimasi dengan Bagging serta Genetic Algorithm mencapai akurasi sebesar 88,30%. Hasil penelitian ini menunjukkan bahwa meskipun Random Forest berfungsi dengan baik untuk mengklasifikasikan data pengajuan pinjaman, pengoptimalan yang digunakan tidak menghasilkan peningkatan akurasi.

Penelitian ini bertujuan untuk:
1. Menganalisis hubungan antara variabel demografis dan pendapatan individu.
2. Mengimplementasikan algoritma Random Forest Classifier untuk klasifikasi pendapatan.
3. Mengevaluasi kinerja model menggunakan metrik accuracy dan classification report.

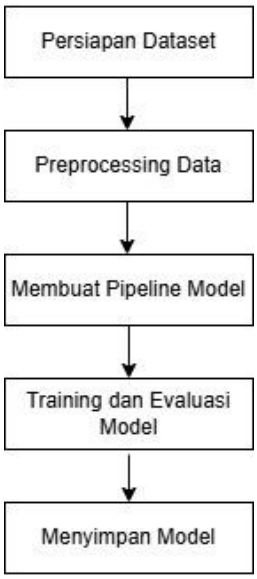
Melalui penelitian ini, diharapkan dapat diperoleh wawasan mengenai bagaimana teknik pembelajaran

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

mesin, khususnya Random Forest, dapat digunakan secara efektif dalam memprediksi pendapatan individu. Selain itu, hasil penelitian ini juga diharapkan dapat memberikan kontribusi positif terhadap pengembangan aplikasi berbasis data untuk mendukung pengambilan keputusan yang lebih cerdas di berbagai sektor.

2. Pembahasan

Penelitian ini bertujuan untuk mengklasifikasikan pendapatan individu berdasarkan atribut demografis menggunakan algoritma Random Forest. Proses klasifikasi dilakukan melalui lima tahapan utama, yaitu: (1) persiapan dataset, (2) preprocessing data, (3) pembuatan pipeline model, (4) pelatihan dan evaluasi model, serta (5) penyimpanan model. Gambar 1 menunjukkan tahapan alur penelitian secara menyeluruh.



Gambar 1. *Tahapan Penelitian*

2.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah Adult Income yang diperoleh dari UCI Machine Learning Repository. Dataset ini berisi sebanyak 32.561 data dengan 15 atribut utama, Target klasifikasi dalam penelitian ini adalah variabel "income" yang terbagi menjadi dua kelas: ≤50K dan >50K. Dataset ini sangat relevan digunakan karena mewakili kondisi sosial-ekonomi individu dan banyak digunakan dalam studi klasifikasi pendapatan.

Dataset yang digunakan memiliki kolom sebagai berikut:

- age: Usia individu
- workclass: Jenis pekerjaan
- fnlwgt: Berat final individu
- education: Tingkat Pendidikan
- education.num: Kategori numerik Pendidikan
- marital.status: Status pernikahan
- occupation: Jenis pekerjaan
- relationship: Hubungan keluarga
- race: Ras individu
- sex: Jenis kelamin
- capital.gain: Keuntungan modal
- capital.loss: Kerugian modal
- hours.per.week: Jam kerja per minggu
- native.country: Negara asal
- income: Kategori pendapatan (>50K atau <=50K)

2.2 Preprocessing Data

Tahap preprocessing menjadi komponen penting dalam pembelajaran mesin karena kualitas data sangat menentukan hasil model. Langkah pertama adalah menghapus atribut yang tidak relevan seperti fnlwgt, karena atribut ini tidak memberikan kontribusi berarti dalam proses prediksi pendapatan. Kemudian dilakukan normalisasi pada fitur numerik seperti usia, capital gain, dan capital loss agar model tidak bias terhadap nilai-nilai skala besar.Selanjutnya, fitur kategorikal seperti jenis pekerjaan, status pernikahan, dan ras diubah ke dalam bentuk numerik menggunakan teknik One-Hot Encoding, yang menghasilkan representasi biner untuk setiap kategori. Hal ini penting karena algoritma Random Forest tidak dapat memproses string secara langsung. Setelah data dibersihkan dan direpresentasikan, dataset dibagi menjadi data latih sebesar 70% dan data uji sebesar 30% untuk memastikan validitas hasil evaluasi model.

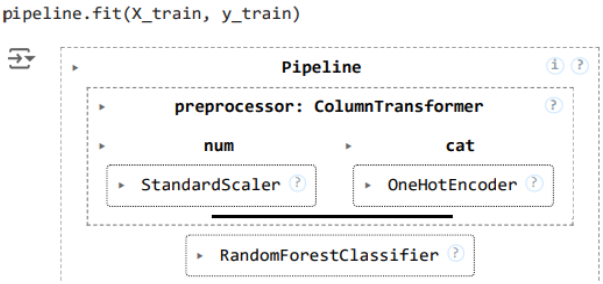
2.3 Model dan Pipeline

Penggunaan Pipeline dalam pembangunan model bertujuan untuk menyusun tahapan preprocessing dan pelatihan secara otomatis dan konsisten. Pipeline yang digunakan dalam penelitian ini dibangun dengan bantuan pustaka scikit-learn, dan terdiri dari dua komponen utama:

- ColumnTransformer: digunakan untuk mengatur transformasi kolom, baik numerik maupun kategorikal, sesuai metode preprocessing masing-masing.
- RandomForestClassifier: digunakan sebagai model klasifikasi utama, dengan parameter random_state=42 untuk menjamin reproducibility hasil.

Penggunaan pipeline mempermudah integrasi proses preprocessing dan pelatihan model dalam satu kesatuan langkah yang efisien dan terstruktur.

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786



Gambar 2. *Diagram alur dari pipeline*

2.4 Evaluasi Model

Model klasifikasi yang digunakan dalam penelitian ini adalah Random Forest Classifier dari pustaka scikit-learn dengan parameter default. Algoritma ini membentuk banyak decision tree secara acak dengan menggunakan subset dari data dan atribut, lalu hasil prediksi ditentukan berdasarkan voting mayoritas. Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, F1-score, serta visualisasi melalui confusion matrix

Model Random Forest Classifier dievaluasi menggunakan classification report dengan hasil sebagai berikut:

- Akurasi: 85,44%
- Precision untuk kelas ≤50K: 0,88
- Recall untuk kelas ≤50K: 0,93
- F1-score untuk kelas ≤50K: 0,91
- Precision untuk kelas >50K: 0,73
- Recall untuk kelas >50K: 0,61
- F1-score untuk kelas >50K: 0,67

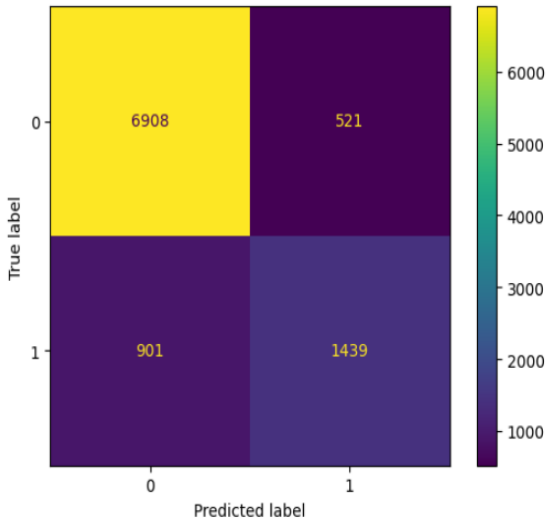
```
Accuracy: 0.8544
Classification Report:
              precision    recall  f1-score   support

       <=50K       0.88      0.93      0.91      7429
        >50K       0.73      0.61      0.67      2340

    accuracy                           0.85      9769
   macro avg       0.81      0.77      0.79      9769
weighted avg       0.85      0.85      0.85      9769
```

Gambar 3. *Hasil Evaluasi*

Hasil evaluasi memperlihatkan bahwa model memiliki kinerja yang sangat baik dalam mengklasifikasikan individu dengan pendapatan ≤50K, yang merupakan kelas mayoritas dalam dataset. Hal ini tercermin dari nilai recall yang tinggi, yaitu sebesar 0,93, yang menunjukkan bahwa sebagian besar individu pada kelas ini berhasil diklasifikasikan dengan benar. Sebaliknya, performa model terhadap kelas >50K (kelas minoritas) masih kurang optimal, dengan nilai recall yang hanya mencapai 0,61. Hal ini mengindikasikan bahwa banyak individu berpendapatan tinggi diklasifikasikan secara keliru ke dalam kelas pendapatan rendah.
Berdasarkan hasil pengujian, belum terdapat indikasi kuat bahwa model mengalami overfitting. Hal ini ditunjukkan oleh nilai akurasi yang cukup tinggi dan konsisten pada data uji, yaitu sebesar 85,44%. Overfitting umumnya ditandai dengan performa yang sangat tinggi pada data pelatihan namun menurun secara drastis pada data pengujian. Namun demikian, karena evaluasi performa terhadap data pelatihan tidak disajikan dalam penelitian ini, maka analisis terhadap kemungkinan overfitting masih bersifat terbatas. Oleh sebab itu, disarankan pada pengembangan selanjutnya untuk menyertakan evaluasi terhadap data pelatihan serta menerapkan teknik validasi silang (cross-validation) guna memperoleh gambaran yang lebih menyeluruh terhadap kemampuan generalisasi model. Selain itu, pengaturan parameter model seperti jumlah pohon (n_estimators) dan kedalaman pohon (max_depth) pada algoritma Random Forest juga penting untuk dioptimalkan guna meminimalkan risiko overfitting dan meningkatkan kinerja model secara keseluruhan.
Untuk memperjelas performa model, digunakan confusion matrix yang menunjukkan jumlah prediksi benar dan salah dari setiap kelas. Confusion Matrix adalah pengukuran performa untuk klasifikasi data berdasarkan nilai prediksi dan nilai aktual. Confusion matrix merupakan tabel dengan beberapa kombinasi dari nilai hasil klasifikasi (Pratiwi & Nugroho, 2024).



Gambar 4. *Hasil Perfoma Confusion Matrix*

Visualisasi confusion matrix pada Gambar 3 mendukung temuan ini. Terdapat 901 kesalahan klasifikasi (False Negative) pada kelas >50K, yang artinya banyak individu yang seharusnya diklasifikasikan sebagai berpendapatan tinggi justru diprediksi memiliki pendapatan rendah. Ketidakseimbangan kelas (class imbalance) menjadi salah satu penyebab utama rendahnya recall ini. Dalam dataset, proporsi data kelas ≤50K jauh lebih banyak dibandingkan >50K, sehingga model cenderung belajar dari pola mayoritas. Penyebab kesenjangan kelas ini berasal dari distribusi alami dalam data sensus yang digunakan, yaitu Adult Income Dataset, yang mencerminkan kondisi riil populasi di mana mayoritas individu memiliki pendapatan tahunan kurang

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

dari atau sama dengan 50K. Karena data ini dikumpulkan dari lingkungan masyarakat umum tanpa penyesuaian proporsi kelas, maka secara alami jumlah sampel dari kelas pendapatan rendah jauh lebih besar dibandingkan kelas pendapatan tinggi. Akibatnya, model mesin pembelajaran seperti Random Forest lebih dilatih pada pola mayoritas dan memiliki performa yang rendah dalam mengenali kelas minoritas.

3. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Random Forest memiliki performa yang baik dalam mengklasifikasikan pendapatan individu berdasarkan atribut demografis, sebagaimana dibuktikan melalui pengujian terhadap dataset Adult Income dari UCI Machine Learning Repository. Model yang dikembangkan dalam penelitian ini mampu mencapai akurasi sebesar 85,44%, dengan nilai precision, recall, dan F1-score yang cukup tinggi untuk kelas mayoritas (≤50K). Hal ini menunjukkan bahwa Random Forest mampu mengenali pola-pola umum dalam data demografi yang berkaitan dengan tingkat pendapatan.

Namun, performa model dalam mengklasifikasikan kelas minoritas (>50K) masih tergolong rendah, terutama dari sisi recall yang hanya mencapai 0,61. Rendahnya recall ini menunjukkan bahwa banyak individu berpendapatan tinggi diklasifikasikan secara salah sebagai berpendapatan rendah. Kesalahan ini disinyalir disebabkan oleh ketidakseimbangan kelas pada dataset, di mana jumlah data untuk kelas ≤50K jauh lebih besar dibandingkan >50K. Situasi ini menyebabkan model cenderung bias terhadap kelas yang lebih dominan dan kurang sensitif terhadap pola-pola dari kelas minoritas.

Berdasarkan temuan tersebut, disarankan beberapa langkah untuk meningkatkan kualitas dan akurasi model di masa yang akan datang. Pertama, penanganan terhadap ketidakseimbangan data menjadi prioritas, dengan menerapkan teknik seperti SMOTE, undersampling, atau penyesuaian class weight untuk memastikan bahwa kedua kelas dapat dipelajari dengan proporsi yang seimbang. Kedua, dilakukan optimasi parameter model melalui hyperparameter tuning, baik dengan menggunakan metode grid search maupun random search, untuk mendapatkan konfigurasi terbaik dalam pelatihan model.

Selanjutnya, eksplorasi terhadap algoritma klasifikasi alternatif seperti Gradient Boosting, XGBoost, atau LightGBM dapat menjadi opsi untuk dibandingkan dengan Random Forest, terutama karena algoritma tersebut memiliki kinerja yang baik dalam menangani data yang tidak seimbang. Evaluasi model juga dapat diperluas dengan metrik tambahan seperti ROC-AUC, log loss, dan Matthews Correlation Coefficient (MCC) guna memperoleh pemahaman yang lebih mendalam mengenai kinerja model secara keseluruhan.

Terakhir, penelitian ini dapat dikembangkan lebih lanjut dengan menggunakan model serupa pada data demografi lokal atau data survei dari masyarakat Indonesia, agar hasil prediksi menjadi lebih aplikatif dan relevan untuk digunakan dalam konteks sosial-ekonomi nasional. Selain itu, model yang dikembangkan juga berpotensi untuk diintegrasikan dalam sistem informasi berbasis data seperti sistem pendukung keputusan (Decision Support System), yang dapat digunakan dalam perumusan kebijakan publik, penyaluran bantuan sosial, atau analisis pasar pada sektor bisnis.

Dengan melakukan perbaikan dan pengembangan lebih lanjut seperti yang disarankan, diharapkan hasil penelitian ini dapat memberikan kontribusi nyata dalam pemanfaatan teknologi machine learning untuk mendukung pengambilan keputusan yang lebih tepat, akurat, dan berbasis data di berbagai bidang.

Daftar Pustaka

Almufqi, F. M., & Voutama, A. (2023). PERBANDINGAN METODE DATA MINING UNTUK MEMPREDIKSI PRESTASI AKADEMIK SISWA. *Jurnal Teknika*, *15*(1), 61–66. https://doi.org/10.30736/jt.v15i1.929

Arisusanto, A., Suarna, N., & Dwilestari, G. (2023). Analisa Klasifikasi Data Harga Handphone Menggunakan Algoritma Random Forest Dengan Optimize Parameter Grid. *Jurnal Teknologi Ilmu Komputer*, *1*(2), 43–47. https://doi.org/10.56854/jtik.v1i2.51

Azhima, T., Siswa, Y., & Pranoto, W. J. (2023). IMPLEMENTASI SELEKSI FITUR INFORMATION GAIN RATIO PADA ALGORITMA RANDOM FOREST UNTUK MODEL DATA KLASIFIKASI PEMBAYARAN KULIAH. *Dinamika Informatika*, *15*(1), 41–49.

Chyntia Morama, H., Ratnawati, D. E., & Arwani, I. (2022). *Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier* (Vol. 6, Issue 4). http://j-ptiik.ub.ac.id

Clara, S., Laksmi Prianto, D., Al Habsi, R., Friscila Lumbantobing, E., Chamidah, N., Kom, S., Kom, M., Informatika, J., Ilmu Komputer, F., Pembangunan Nasional Veteran Jakarta Jl Fatmawati Raya, U. R., Labu, P., Cilandak, K., & Depok, K. (2021). Implementasi Seleksi Fitur Pada Algoritma Klasifikasi Machine Learning Untuk Prediksi Penghasilan Pada Adult Income Dataset. In *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*.

Fabian Azmi, A., Voutama, A., Karawang Jl HSRonggo Waluyo, S., & Timur, T. (2024). *PREDIKSI CHURN NASABAH BANK MENGGUNAKAN KLASIFIKASI RANDOM FOREST DAN*

ISSN PRINT : 2302-500X
ISSN ONLINE : 2477-3786

DECISION TREE DENGAN EVALUASI CONFUSION MATRIX. *13*(1).

Fitri, E. (2023). Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah. *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)*, *4*(1), 2723–1453. https://doi.org/10.52158/jacost.491

Irma Purnamasari, A., & Ali, I. (2024). PERBANDINGAN TINGKAT AKURASI ALGORITMA DECISION TREE DAN RANDOM FOREST DALAM MENGKLASIFIKASI PENERIMA BANTUAN SOSIAL BPNT DI DESA SLANGIT. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 1).

Kurniawan, I., Cahya Putri Buani, D., Apriliah, W., Amegia Saputra, R., & Korespondensi, P. (2023). *IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK MENENTUKAN PENERIMA BANTUAN RASKIN IMPLEMENTATION OF RANDOM FOREST ALGORITHM FOR DETERMINING RECIPIENTS OF RASKIN*. *10*(2), 421–428. https://doi.org/10.25126/jtiik.202396225

Nurfitriana, D., Ridwan, T., & Voutama, A. (2024). Analisis Opini Terhadap Aplikasi Riliv di Twitter Menggunakan Algoritma Naïve Bayes dan Random Forest. *Jurnal SAINTEKOM*, *14*(1), 26–37. https://doi.org/10.33020/saintekom.v14i1.526

Prasojo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *7*(2), 79–89. https://doi.org/10.25077/teknosi.v7i2.2021.79-89

Pratiwi, G. E., & Nugroho, A. (2024). IMPLEMENTASI METODE RANDOM FOREST UNTUK KLASIFIKASI PENJUALAN PRODUK SABUN PALING LARIS. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, *7*(2), 541. https://doi.org/10.37600/tekinkom.v7i2.1610

Rosta Br Sebayang, E., Herry Chrisnanto, Y., Jenderal Achmad Yani Cimahi, U., Terusan Jend Sudirman, J., Selatan, C., & Barat, J. (2023). Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest. In *IJESPG Journal* (Vol. 1, Issue 3). http://ijespgjournal.org

Samudra, J. T., Hayadi, B. H., & Ramadhan, P. S. (2022). Nomor 2. *Volume*, *5*, 127–133. https://ojs.trigunadharma.ac.id/index.php/jsk/index

Sinambela, D. P., Naparin, H., Zulfadhilah, M., & Hidayah, N. (2023). Implementasi Algoritma Decision Tree dan Random Forest dalam Prediksi Perdarahan Pascasalin. *Jurnal Informasi Dan Teknologi*, *5*(3), 58–64. https://doi.org/10.60083/jidt.v5i3.393

Yoga Religia, Agung Nugroho, & Wahyu Hadikristanto. (2021). Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(1), 187–192. https://doi.org/10.29207/resti.v5i1.2813