

Analisis Retensi Tenaga Kerja Panen Kelapa Sawit dengan Random Forest pada Perusahaan XYZ

Haidar Zakki Jumali^{*1}, Vinna Rahmayanti Setyaning Nastiti¹, Yufis Azhar¹

^{*}Informatika, Universitas Muhammadiyah Malang

haidarzjumm1@webmail.umm.ac.id^{*}

Abstrak

Sektor kelapa sawit berperan krusial dalam perekonomian Indonesia. Perusahaan XYZ turut berkontribusi dalam bidang ini, maka retensi, yaitu kemampuan perusahaan mempertahankan tenaga kerja panen kelapa sawit, menjadi penting untuk menjaga stabilitas dan kinerjanya. Random Forest diusulkan sebagai algoritma untuk mengklasifikasikan retensi tenaga kerja panen kelapa sawit. Berisi data 370 pekerja selama satu bulan, dataset diperoleh dari HRD dengan memadukan fitur kategorikal serta numerik. Tahap preprocessing meliputi data cleaning, pembuatan dan seleksi fitur, transformasi data, serta mengatasi distribusi kelas yang tidak seimbang dengan SMOTE. Rasio splitting data yaitu 70% untuk data train dan 30% untuk data test. Evaluasi model dilakukan melalui dua pendekatan: parameter default menghasilkan akurasi 98%, sedangkan hyperparameter tuning dengan GridSearchCV memperoleh akurasi yang lebih tinggi, yaitu 99%. Efektivitas Random Forest dalam mengklasifikasikan retensi tenaga kerja panen kelapa sawit terbukti melalui hasil ini. Wawasan dari penelitian ini memungkinkan HRD untuk memperkuat taktik retensi, mengoptimalkan produktivitas tenaga kerja, dan mendukung pengambilan keputusan berbasis data.

Kata Kunci: Random Forest, Retensi Tenaga Kerja Panen Kelapa Sawit, HRD, Klasifikasi, Machine Learning

Abstract

The palm oil sector plays a crucial role in Indonesia's economy. Company XYZ also contributes to this field, making retention that is the company's ability to maintain palm harvest workers are essential for maintaining its stability and performance. Random Forest is proposed as an algorithm to classify the retention of palm harvest workers. Containing data from 370 workers over one month, this dataset was obtained from HRD and combines both categorical and numerical features. The preprocessing stages include data cleaning, feature selection and creation, data transformation, and addressing imbalanced distribution using SMOTE. The data splitting ratio is 70% for training data and 30% for test data. The model evaluation is conducted through two approaches: the default parameters produce an accuracy of 98%, while hyperparameter tuning using GridSearchCV achieves a higher accuracy of 99%. The effectiveness of Random Forest in classifying the retention of palm harvest workers is proven through these results. The insights from this research enable HRD to strengthen retention tactics, optimize worker productivity, and support data-driven decision-making.

Keywords: Random Forest, Palm Harvest Worker Retention, HRD, Classification, Machine Learning

1. Pendahuluan

Kelapa sawit memiliki peran krusial bagi perekonomian Indonesia, sebagai salah satu produsen minyak kelapa sawit terbesar di dunia. Selama 2016–2022, sektor perkebunan—terutama kelapa sawit—berkontribusi sekitar 3,65% terhadap PDB. Kelapa sawit juga menjadi komoditas ekspor nonmigas terbesar kedua setelah batu bara, dengan nilai mencapai 27.418,2 juta USD atau 10,81% dari total ekspor nonmigas. Pada 2022, produksi minyak kelapa sawit Indonesia tercatat sebesar 46,50 juta ton (1), (2). Maka optimalisasi sektor ini melalui peningkatan retensi tenaga kerja penting untuk meningkatkan kompetisi global.

Perusahaan XYZ bergerak di bidang perkebunan kelapa sawit di Kalimantan Tengah, memiliki hingga 5.000 tenaga kerja. Tenaga kerja memegang peranan krusial dalam berbagai proses, mulai dari pembukaan lahan hingga pengolahan hasil (3). Khususnya, tenaga kerja panen

berkontribusi menjaga kualitas dan kontinuitas hasil panen, yang memengaruhi stabilitas operasional perusahaan (4). Retensi tenaga kerja menjadi faktor utama dalam menjaga efisiensi produksi, karena turnover yang tinggi dapat meningkatkan biaya rekrutmen, pelatihan, dan adaptasi pekerja baru, serta mengganggu proses produksi (5), (6). Maka dari itu tingkat retensi yaitu kemampuan perusahaan mempertahankan tenaga kerja menjadi kunci bagi stabilitas perusahaan (5).

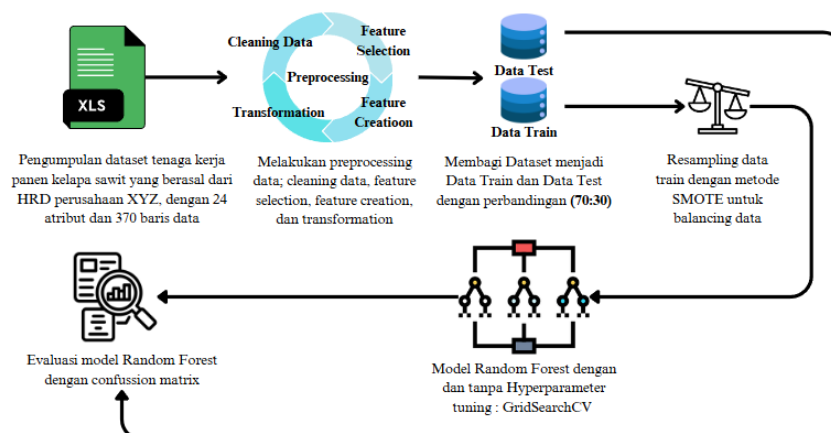
Machine learning semakin banyak diterapkan di berbagai sektor, namun penggunaannya dalam industri kelapa sawit masih terbatas, meskipun berpotensi mendukung keberlanjutan sektor ini (7). Machine learning bertujuan mengenali pola dari data untuk menyelesaikan tugas tertentu (8). Memiliki potensi untuk mengoptimalkan berbagai aspek, termasuk klasifikasi retensi tenaga kerja panen di Perusahaan XYZ. Salah satu algoritma yang relevan untuk klasifikasi retensi adalah Random Forest yang mampu menangani variabel kompleks yang memengaruhi keputusan retensi melalui pendekatan ensemble learning dengan pendekatan bagging dan majority votes, sehingga cocok digunakan dalam kasus klasifikasi(9) (10).

Efektivitas Random Forest dalam memprediksi turnover tenaga kerja di mayoritas sektor perkantoran telah dibuktikan melalui beberapa penelitian terdahulu. Parmod Kumar et al. menekankan pentingnya memantau turnover menggunakan model machine learning untuk menjaga stabilitas tenaga kerja dan meminimalisir risiko attrisi, dengan akurasi mencapai 83,24% (11). Meskipun turnover paper tersebut tetap relevan bagi klasifikasi retensi, yakni mengidentifikasi faktor atau kategori atribut yang memengaruhi keputusan karyawan untuk bertahan. Jin et al. mengusulkan model hibrida RFRSF yang mengombinasikan Random Forest dengan analisis survival, mencapai akurasi 84,65%(12), memperlihatkan potensi Random Forest dalam mengklasifikasikan retensi, terutama terkait umur kerja karyawan di organisasi. Chakraborty dan Mridha menemukan bahwa Random Forest menghasilkan akurasi tertinggi sebesar 90,20% dalam memprediksi turnover dibandingkan algoritma lainnya (13). Sementara itu, Budiman et al. melaporkan akurasi 98% saat menguji Random Forest untuk mengkaji pergantian karyawan di organisasi lain(14). Namun, riset terkait klasifikasi retensi tenaga panen kelapa sawit menggunakan Random Forest masih terbatas, sehingga studi ini bertujuan mengisi kekosongan tersebut.

Penelitian ini berfokus pada pengembangan dan evaluasi model dengan klasifikasi retensi tenaga kerja panen kelapa sawit dengan algoritma Random Forest. Diharapkan, hasil studi ini dapat memberikan wawasan baru bagi manajemen SDM Perusahaan XYZ dalam mengelola tenaga kerja, meningkatkan retensi, dan memperkuat stabilitas operasional.

2. Metode Penelitian

Tahapan penelitian disusun secara sistematis untuk mencapai tujuan yang ditetapkan. Gambar 1 menunjukkan alur proses yang dilakukan untuk mengevaluasi efektivitas algoritma Random Forest dalam mengklasifikasikan retensi tenaga kerja panen kelapa sawit di Perusahaan XYZ.



Gambar 1. Tahapan Penelitian

Dataset penelitian bersumber dari Human Resource Development (HRD) Perusahaan XYZ, terdiri dari 370 instance yang merepresentasikan tenaga kerja individu, lengkap dengan

berbagai atribut terkait karakteristik pribadi, riwayat kerja, kinerja panen harian, mingguan, dan bulanan, serta status kerja.

Preprocessing data meliputi, cleaning data seperti penanganan *missing values* dengan *fillna*, memperbaiki tipe data, ekstraksi nilai numerik. Feature selection bertujuan mengurangi dimensi data dengan memilih fitur relevan berdasarkan pengetahuan domain (15). Feature creation tahap mengembangkan variabel yang baru yaitu "Produktivitas" dikategorikan ke dalam tiga tingkat: "rendah", "sedang", "tinggi" berdasarkan "Produktivitas Aktual". Kemudian transformasi data menggunakan LabelEncoder untuk mengonversi variabel kategorikal (16). Proses splitting data membagi data total menjadi rasio,70% yakni 259 data train untuk melatih model dan 30% yaitu 111 data test untuk evaluasi performa model pada data yang belum pernah dilihat sebelumnya (17). Resampling melibatkan SMOTE untuk penyeimbangan guna meningkatkan hasil akurasi model (18).

Random forest yaitu algoritma ensemble dengan cara kerja membangun banyak pohon keputusan dengan metode bagging, di mana setiap pohon menerima subset acak data dan fitur, yang efektif menangani hubungan non-linear antar fitur, mengurangi overfitting, dan menghasilkan prediksi yang stabil melalui voting mayoritas serta sesuai untuk kasus klasifikasi (9), (10). Model diuji melalui dua skenario: pertama, menggunakan pengaturan default tanpa tuning hyperparameter, dan kedua, dengan tuning hyperparameter menggunakan GridSearchCV, yang mengevaluasi kombinasi parameter secara menyeluruh berdasarkan metrik evaluasi yang ditentukan (19), dengan random state untuk replikasi

Evaluasi performa model dilakukan menggunakan confusion matrix pada Tabel 2. berbentuk N x N, di mana N adalah jumlah kelas target. Visualisasi matriks ini membantu mengukur akurasi model dengan memerhatikan nilai diagonal utama yang merepresentasikan jumlah klasifikasi benar (20). Metrik evaluasi beserta rumus yang digunakan dijelaskan lebih lanjut pada bagian berikut:

1. Akurasi (Accuracy): Proporsi prediksi yang benar dibandingkan total prediksi.
2. Precision: Proporsi prediksi positif yang benar dibandingkan dengan seluruh prediksi positif.
3. Recall: Proporsi kasus positif yang diprediksi dengan benar oleh model.
4. F1-Score: Rata-rata harmonik dari precision dan recall, memberikan keseimbangan antara keduanya.

Tabel 1. Confusion Matrix

	Negatif	Positif
Negatif	True Negative(TN)	False Positive (FP)
Positif	False Negative (FN)	True Positive (TP)

$$Precision = \frac{TP}{(TP + FN)} \quad (1)$$

$$Recall = \frac{TP}{(TP + FP)} \quad (2)$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (3)$$

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (4)$$

Keterangan :

TP : Jumlah data positif yang tervalidasi benar

TN : Jumlah data negatif yang tervalidasi benar

FN : Jumlah data negatif namun tervalidasi salah

FP : Jumlah data positif namun tervalidasi salah

3. Hasil Penelitian dan Pembahasan

3.1 Pengumpulan Data

Data tenaga kerja panen kelapa sawit yang disajikan dalam Tabel 2 diperoleh secara resmi dari departemen HRD Perusahaan XYZ. Data dalam format .xlsx ini berisi 370 instance yang merepresentasikan individu tenaga kerja, dilengkapi dengan berbagai atribut terkait karakteristik pribadi, riwayat kerja, kinerja panen harian, mingguan, bulanan, serta status kerja mereka. Data tersebut dimuat ke dalam dataframe untuk diproses lebih lanjut.

Tabel 2. Sampel Data Tenaga Kerja Panen

No.	Atribut	Instance	Instance	Instance	Keterangan
1	No.	1	2	11	Nomor urut pekerja lapangan.
2	Nama	FRAXXX	BUXXX	PETRXXX	Nama pekerja
3	Umur	35 Tahun 1 Bulan	51 Tahun 1 Bulan	34 Tahun 10 Bulan	Usia pekerja saat data diambil.
4	Suku	Kupang	Jawa	Kupang	Suku atau etnis pekerja.
5	Pendidikan	SD	SMP	SD	Tingkat pendidikan terakhir pekerja.
6	DOJ	18/10/2013	17/04/2014	21/06/2017	Tanggal mulai bergabung di perusahaan.
7	Masa Kerja	10 Tahun 7 Bulan	10 Tahun 1 Bulan	6 Tahun 11 Bulan	Lama waktu bekerja di perusahaan.
8	Status Tenaga Kerja	KHT	KHT	KHL	Status pekerja, harian tetap (KHT) atau harian lepas (KHL).
9	Total Salary	3.097.609	2.645.199	2.976.785	Total gaji bulanan pekerja.
10	Insentif	464.641			Insentif berdasarkan kinerja panen.
11	Hari Kerja	24	21	23	Jumlah hari aktif dalam satu bulan.
12	panen sehari	271	231	260	Hasil panen rata-rata per hari.
13	panen seminggu	1.274	1.088	1.224	Hasil panen rata-rata per minggu.
14	panen satu bulan	5.285	4.513	5.079	Total hasil panen dalam satu bulan.
15	Rp/janjang	674	586	586	Upah per tandan sawit yang dipanen.
16	Produktivitas				Kategori produktivitas (rendah, sedang, tinggi).
17	Target Produktivitas	5000	5000	5000	Target hasil panen bulanan perusahaan.
18	Produktivitas Aktual	105,68	90,24	101,56	Persentase pencapaian target panen bulanan.
19	Status Pernikahan	Menikah	Menikah	Lajang	Status perkawinan pekerja ("Lajang" atau "Menikah.")
20	Jumlah Tanggungan	1	2	0	Jumlah anggota keluarga yang menjadi tanggungan.
21	Jarak Tempuh	4	5	7	Jarak dari tempat tinggal ke lokasi kerja(km)
22	Kelayakan Promosi	Layak	Tidak Layak	Layak	Evaluasi potensi promosi pekerja.
23	Retensi	Stay	Leave	Stay	Status keberlanjutan kerja "stay" atau "leave"

3.2 Preprocessing Data

3.2.1 Cleaning Data

Langkah yang dilakukan untuk memastikan kualitas data dapat digunakan dalam analisis dan pembelajaran pola model. Pada Gambar 2 disajikan mencakup rangkuman informasi untuk identifikasi dan ditangani sesuai kondisinya, seperti missing value pada kolom insentif dengan imputasi nilai 0 agar tidak null dan penanganan "Produktivitas" pada tahap creation, tidak ditemukan duplikat.

```

Data columns (total 23 columns):
#   Column              Non-Null Count  Dtype
---  -
0   No                   370 non-null    int64
1   Nama                 370 non-null    object
2   Umur                 370 non-null    object
3   Suku                 370 non-null    object
4   Pendidikan           370 non-null    object
5   DOJ                  370 non-null    datetime64[ns]
6   Masa Kerja           370 non-null    object
7   Status Tenaga Kerja 370 non-null    object
8   Total salary         370 non-null    float64
9   insentif             222 non-null    float64
10  Hari Kerja           370 non-null    float64
11  panen sehari         370 non-null    float64
12  panen seminggu      370 non-null    float64
13  panen satu bulan    370 non-null    float64
14  Rp/janjang           370 non-null    float64
15  productivity         0 non-null      float64
16  Target Produktivitas 370 non-null    int64
17  Produktivitas Aktual 370 non-null    float64
18  Status Pernikahan   370 non-null    object
19  Jumlah Tanggungan   370 non-null    int64
20  Jarak Tempuh         370 non-null    int64
21  Kelayakan Promosi    370 non-null    object
22  Retensi              370 non-null    object
dtypes: datetime64[ns](1), float64(9), int64(4), object(9)

```

Gambar 2. Rangkuman Informasi

Tabel 3. Hasil Transformasi

Kolom	Sebelum Transformasi	Setelah Transformasi
Umur	35 Tahun 1 Bulan	35
	51 Tahun 1 Bulan	51
	33 Tahun 5 Bulan	33
Masa Kerja	10 Tahun 7 Bulan	10
	10 Tahun 1 Bulan	10
	9 Tahun 11 Bulan	9

Pada Tabel 3 disajikan atribut Umur dan Masa Kerja, terdapat nilai yang ditulis dengan format string dan tidak seragam, seperti "35 tahun". Dilakukan ekstraksi angka dari setiap nilai dalam kolom tersebut sehingga hanya menyisakan informasi numerik agar terstruktur dan menghindari kesalahan saat analisis statistik atau pemodelan dilakukan.

Tabel 4. Konversi Tipe Data

Atribut	Sebelum	Setelah
Umur, Masa Kerja	object	int64
Total salary, Insentif, Hari Kerja, Panen Sehari, Panen Seminggu, Panen Satu Bulan, Rp/janjang	float64	int64
Atribut	Sebelum	Setelah

Tabel 4 menunjukkan beberapa atribut, seperti Total Salary, Insentif, dan Hari Kerja, memiliki tipe data float yang seharusnya integer. Hal ini ditangani dengan mengonversi tipe data pada atribut tersebut, kecuali Produktivitas Aktual yang tetap dalam bentuk aslinya agar lebih detail dengan float. Langkah ini memastikan data menjadi lebih terstruktur, seragam, dan siap untuk proses pemodelan.

3.2.2 Feature Selection

Beberapa penelitian terdahulu kategori variabel yang dapat menjelaskan retensi dibentuk dengan pengetahuan domain, fitur yang digunakan dalam pelatihan model adalah sebagai berikut:

*Analisis Retensi Tenaga Kerja Panen Kelapa Sawit...
Haidar Zakki Jumali, Vinna Rahmayanti Setyaning Nastiti, Yufis Azhar*

- Karakteristik pemanen: Umur, Suku, Pendidikan, Status Pernikahan, dan Jumlah Tanggungan.
- Riwayat pekerjaan: Masa Kerja, Status Tenaga Kerja, dan Jarak Tempuh ke lokasi kerja.
- Insentif dan kompensasi: Insentif, Total Gaji (Total Salary), dan Kelayakan Promosi.
- Produktivitas kerja: Panen per hari, Panen per minggu, Panen per bulan, dan Rp per janjang.
- Target : Retensi, seorang tenaga panen tetap bekerja atau keluar dari perusahaan.

3.2.3 Feature Creation

Transformasi ini dilakukan dengan membagi nilai Produktivitas Aktual ke dalam tiga kategori untuk mempermudah analisis dan pemodelan. Tabel 5 menunjukkan pembentukan fitur ini bertujuan untuk menyederhanakan pemahaman produktivitas tenaga kerja serta meningkatkan efektivitas model dalam mengelompokkan pekerja berdasarkan tingkat produktivitas mereka.

Tabel 5. Feature Creation

Interval (Produktivitas Aktual)	Kategori Produktivitas
0 - 97	Rendah
98 - 135	Sedang
> 135	Tinggi

3.2.4 Transformation

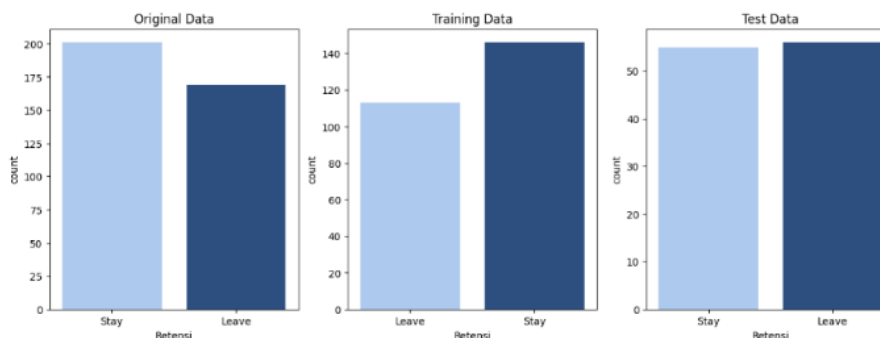
Hasil transformasi atau encoding fitur kategorikal ke dalam bentuk numerik menggunakan Label Encoder bertujuan mempermudah model mengenali pola dari variabel kategori. Pada Tabel 6 tersebut menampilkan kategori asli beserta representasi numeriknya.

Tabel 6. Encoded Features

Fitur	Kategori & Kode
Suku	0 = Bugis, 1 = Dayak, 2 = Flores, 3 = Jawa, 4 = Kupang, 5 = Madura, 6 = Melayu, 7 = Sumba, 8 = Sunda
Pendidikan	0 = SD, 1 = SMA, 2 = SMP
Status Tenaga Kerja	0 = KHL, 1 = KHT
Retensi	0 = Leave, 1 = Stay
Status Pernikahan	0 = Lajang, 1 = Menikah
Kelayakan Promosi	0 = Layak, 1 = Tidak Layak
Produktivitas	0 = Rendah, 1 = Sedang, 2 = Tinggi

3.3. Splitting Data

Dataset dibagi menjadi 70% yaitu 259 sampel untuk pelatihan dan 30% yaitu 111 sampel untuk pengujian, dengan proporsi kelas asli tetap terjaga. Data latih diproses menggunakan SMOTE untuk mengatasi ketidakseimbangan kelas, sementara data uji tetap dalam bentuk aslinya agar mencerminkan kondisi nyata.



Gambar 3. Sebelum dan Setelah Splitting Data

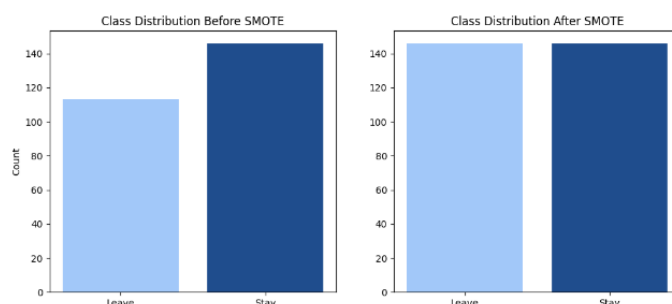
3.4. Resampling

Dataset awal mengungkap adanya ketidakseimbangan kelas, dengan 201 data pada kelas Stay (1) dan 169 pada kelas Leave (0). Ketimpangan ini berpotensi membuat model lebih condong memprediksi kelas mayoritas dengan akurasi tinggi, namun berisiko gagal mengenali pola dari kelas minoritas secara efektif, sehingga mengganggu keseimbangan performa klasifikasi.



Gambar 4. Distribusi Kelas Sebelum SMOTE

Guna mengatasi ketidakseimbangan kelas, seperti yang ditunjukkan pada Gambar 4 dan Gambar 5, diterapkan teknik Synthetic Minority Over-sampling Technique (SMOTE). Proses ini dilakukan setelah dataset dibagi menjadi data latih dan data uji, memastikan bahwa SMOTE hanya diterapkan pada data latih untuk mencegah kebocoran data. Hasilnya, distribusi kelas menjadi lebih seimbang, memungkinkan model mempelajari pola dari kedua kelas secara lebih efektif dan meningkatkan kemampuan klasifikasinya.



Gambar 5. Distribusi Setelah SMOTE

3.5 Hasil Skenario Model Random Forest & Evaluasi

3.5.1 Confusion Matrix

Evaluasi kinerja kedua model Random Forest dalam mengklasifikasikan retensi tenaga panen dilakukan dengan menggunakan confusion matrix dan metrik-metrik terkait seperti Precision, Recall, F1-Score, dan Akurasi. Analisis ini mencakup model dengan parameter default dan model yang telah dioptimasi melalui hyperparameter tuning.

3.5.1.1 Evaluasi Random Forest Default

Model menghasilkan 55 True Positive, 2 False Positive, 54 True Negative, dan 0 False Negative.

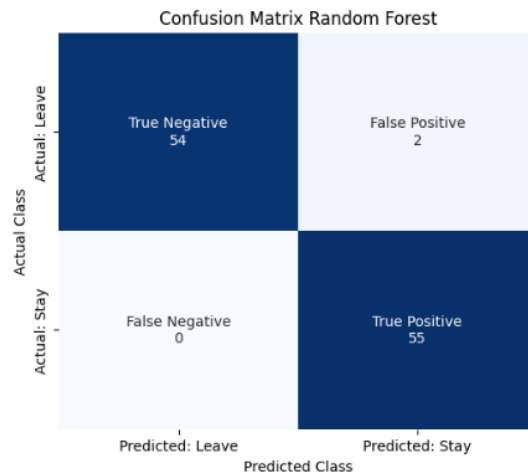
$$Precision = \frac{TP}{(TP + FN)} = \frac{55}{(55 + 2)} = 0.965 \quad (1)$$

$$Recall = \frac{TP}{(TP + FN)} = \frac{55}{(55 + 0)} = 1.00 \quad (2)$$

$$F1 - Score = 2 \times \frac{0.965 \times 1.00}{0.965 + 1.00} = 2 \times \frac{0.965}{1.965} = 0.98 \quad (3)$$

$$Accuracy = \frac{55 + 54}{(55 + 54 + 2 + 0)} = \frac{109}{111} = 0.98 \quad (4)$$

Model Random Forest dengan parameter default mencapai akurasi 98.1%, sebagaimana ditunjukkan pada Confusion Matrix (Gambar 6) dan laporan klasifikasi (Tabel 7). Nilai Precision, Recall, dan F1-Score untuk kedua kelas (0 dan 1) masing-masing sebesar 0.98 atau lebih. Kemampuan model dalam mengklasifikasikan tenaga kerja yang bertahan dan keluar terlihat dari jumlah False Negative (FN) yang sebesar 0 dan False Positive (FP) sebanyak 2, menunjukkan kesalahan prediksi pekerja yang keluar berada pada tingkat yang sangat rendah.



Gambar 6. Confusion Matrix Random Forest

Tabel 7. Classification Report Random Forest

Class	Precision	Recall	F1-Score	Support
Leave	1.00	0.96	0.98	56
Stay	0.96	1.00	0.98	55
Accuracy			0.98	111
Macro Avg	0.98	0.98	0.98	111
Weighted Avg	0.98	0.98	0.98	111

3.5.1.2 Evaluasi Model Random Forest Dengan Hyperparameter Tuning

Model menghasilkan 55 True Positive, 1 False Positive, 55 True Negative, dan 0 False Negative.

$$Precision = \frac{TP}{(TP + TN)} = \frac{55}{(55 + 1)} = 0.992 \quad (1)$$

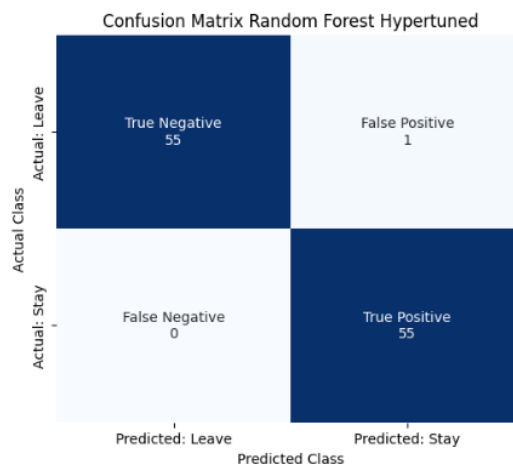
$$Recall = \frac{TP}{(TP + FN)} = \frac{55}{(55 + 0)} = 1.00 \quad (2)$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} = 2 \times \frac{0.982 \times 1.00}{0.982 + 1.00} = 2 \times \frac{0.982}{1.982} = 0.991 \quad (3)$$

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} = \frac{55+55}{(55+55+1+0)} = \frac{110}{111} = 0.99 \quad (4)$$

Hasil hyperparameter tuning menggunakan GridSearchCV (Gambar 7 dan Tabel 8) menunjukkan bahwa akurasi model meningkat hingga 99%, disertai dengan peningkatan kecil pada Precision dan F1-Score. Jumlah False Positive menurun dari 2 menjadi 1, sedangkan False Negative tetap berada di angka 0. Detail parameter hasil tuning adalah sebagai berikut:

- Bootstrap: “False” — Menggunakan seluruh dataset tanpa resampling, meningkatkan stabilitas tetapi mengurangi variasi antar pohon (rentang uji: [True, False]).
- Criterion: “Entropy” — Mempertimbangkan ketidakpastian informasi saat membagi node, berguna untuk distribusi kelas yang tidak seimbang (rentang uji: ['gini', 'entropy']).
- Max_depth: “None” — Mengizinkan pohon tumbuh hingga maksimal, sampai semua sampel dalam node terminal memiliki kelas yang sama, meskipun berisiko overfitting (rentang uji: [None, 10, 20]).
- Max_features: “log2” — Membatasi jumlah fitur yang dipertimbangkan tiap pohon, menyeimbangkan akurasi dan efisiensi (rentang uji: ['sqrt', 'log2']).
- Min_samples_leaf: “1” — Memungkinkan model menangkap pola detail, meskipun meningkatkan potensi overfitting (rentang uji: [1, 2, 4]).
- Min_samples_split: “2” — Mengizinkan node terbagi jika memiliki minimal dua sampel, memberi fleksibilitas dalam proses pemisahan node (rentang uji: [2, 5, 10]).
- N_estimators: “50” — Menghasilkan 50 pohon keputusan, cukup untuk meningkatkan stabilitas model tanpa memperpanjang waktu pelatihan (rentang uji: [50, 100, 200]).



Gambar 7. Confusion Matrix RF Hyperparameter Tuning

Tabel 8 Classification Report Random Forest Dengan Hyperparameter Tuning

Class	Precision	Recall	F1-Score	Support
Leave	1.00	0.98	0.99	56
Stay	0.98	1.00	0.99	55
Accuracy			0.99	111
Macro avg	0.99	0.99	0.99	111
Weighted avg	0.99	0.99	0.99	111

Berdasarkan evaluasi pada Tabel 9, Random Forest terbukti efektif dalam mengklasifikasikan retensi tenaga kerja panen di Perusahaan XYZ, dengan akurasi 98% pada model default dan meningkat menjadi 99% setelah hyperparameter tuning. Meskipun terdapat peningkatan, selisihnya kecil sehingga model default tetap relevan untuk diterapkan. Hasil tinggi ini juga didukung oleh pemilihan fitur yang konsisten pada kedua model, yang berperan penting dalam membentuk performa klasifikasi.

Tabel 9. Perbandingan Performa Model

Model	Akurasi
Random Forest	98%
Random Forest Hyperparameter Tuning	99%

4. Kesimpulan

Kesimpulan penelitian “Analisis Retensi Tenaga Kerja Panen Kelapa Sawit dengan Random Forest pada Perusahaan XYZ” ini membuktikan bahwa algoritma Random Forest efektif dalam mengklasifikasikan retensi tenaga kerja panen kelapa sawit di Perusahaan XYZ. Model

dengan parameter default mencapai akurasi 98%, sementara optimasi melalui GridSearchCV meningkatkan akurasi menjadi 99%. Hasil ini menunjukkan bahwa Random Forest mampu membedakan tenaga kerja yang bertahan dan keluar dengan tingkat kesalahan minimal. Temuan ini memungkinkan HRD di Perusahaan XYZ untuk memanfaatkan model Random Forest dalam memperkuat strategi retensi, mengoptimalkan produktivitas tenaga kerja, dan mendukung pengambilan keputusan berbasis data.

Referensi

- [1] Hidayatullah T, Kajian P, Keahlian AB, Jenderal S. Artikel Analisis Empiris Produksi Kelapa Sawit Terhadap Tingkat Kesejahteraan Petani An Empirical Analysis of Palm Oil Production on The Welfare of Farmers. Vol. 8, Jurnal Budget: Isu dan Masalah Keuangan Negara.
- [2] Sulaiman AA, Amruddin A, Bahrun AH, Yuna K, Keela M. New Challenges and Opportunities of Indonesian Crude Palm Oil in International Trade. Caraka Tani: Journal of Sustainable Agriculture. 2024 Apr 1;39(1):94–106.
- [3] Tassew T, Nie X. A Comprehensive Review of the Application of Machine Learning in Medicine and Health Care [Internet]. 2022. Available from: <https://www.techrxiv.org/doi/full/10.36227/techrxiv.21204779.v1>
- [4] Tasya Billa M, Iswarini H. Faktor yang Mempengaruhi Produktivitas Tenaga Kerja Buruh Panen Kelapa Sawit Di PT. Patri Agung Perdana Estate Rambutan Desa Suka Pindah Kabupaten Banyuasin.
- [5] Serlia Raflesia Lamandasa dan Frizka Paulin Lumentah Strategi Retensi Karyawan PT. Sawit Jaya Abadi dalam Meningkatkan Employee Engagement.
- [6] Prawitasari A. Faktor-Faktor Yang Mempengaruhi Turnover Intention Karyawan Pada PT. Mandiri Tunas Finance Bengkulu.
- [7] Khan N, Kamaruddin MA, Sheikh UU, Yusup Y, Bakht MP. Oil palm and machine learning: Reviewing one decade of ideas, innovations, applications, and gaps. Vol. 11, Agriculture (Switzerland). MDPI; 2021.
- [8] Machine Learning.
- [9] Gusnina M. Penerapan Metode Random Forest pada Klasifikasi Student Academics Performance di Universitas Sebelas Maret. 2022;
- [10] Azhar Y, Mahesa GA, Mustaqim MochC. Prediction of hotel bookings cancellation using hyperparameter optimization on Random Forest algorithm. Jurnal Teknologi dan Sistem Komputer. 2021 Jan 31;9(1):15–21.
- [11] Kumar P, Gaikwad SB, Ramya ST, Tiwari T, Tiwari M, Kumar B. Predicting Employee Turnover: A Systematic Machine Learning Approach for Resource Conservation and Workforce Stability †. Engineering Proceedings. 2023;59(1).
- [12] Jin Z, Shang J, Zhu Q, Ling C, Xie W, Qiang B. RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Science and Business Media Deutschland GmbH; 2020. p. 503–15.
- [13] Chakraborty R, Mridha K, Shaw RN, Ghosh A. Study and Prediction Analysis of the Employee Turnover using Machine Learning Approaches. In: 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies, GUCON 2021. Institute of Electrical and Electronics Engineers Inc.; 2021.
- [14] Surya Budiman A. Uji Akurasi Algoritma Machine Learning Untuk Pemodelan Prediksi Faktor Pendorong Pergantian Karyawan. JSI: Jurnal Sistem Informasi (E-Journal) [Internet]. 2023;15(1). Available from: <http://ejournal.unsri.ac.id/index.php/jsi/index>
- [15] Made I, Adnyana B, Jln SB, Puputan R. Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa.
- [16] Purba W, Sembiring GA, Saputra A, Turnip T, Jua B, Manihuruk I, et al. Penerapan Data Mining Untuk Pengelolaan Data Rekam Medis Menggunakan Metode K-Means Clustering Pada Rumah Sakit Royal Prima Medan. Jurnal TEKINKOM [Internet]. 2023;6(1). Available from: <https://jurnal.murnisadar.ac.id/index.php/Tekinkom/158>
- [17] Wiratama Y, Abdul Aziz R. Perbandingan Prediksi Penyakit Stunting Balita Menggunakan Algoritma Support Vektor Machine dan Random Forest. Technology and Science (BITS). 2024;6(2):1159–68.

- [18] Haryawan C, Muria Kusuma Ardhana Y, Raya Janti Karang Jambe No J, Bantul K. Analisa Perbandingan Teknik Oversampling SMOTE Pada Imbalanced Data [Internet]. Vol. 6, Jurnal Informatika & Rekayasa Elektronika). 2023. Available from: <http://e-journal.stmiklombok.ac.id/index.php/jire>
- [19] Ramadanti E, Dinathi DA, Sri C, Aditya K, Chandranegara R. Diabetes Disease Detection Classification Using Light Gradient Boosting (LightGBM) With Hyperparameter Tuning. Jurnal dan Penelitian Teknik Informatika [Internet]. 2024;8(2). Available from: <https://doi.org/10.33395/v8i2.13530>
- [20] Jasman TZ, Fadhlullah MA, Pratama AL, Rismayani R. Analisis Algoritma Gradient Boosting, Adaboost dan Catboost dalam Klasifikasi Kualitas Air. Jurnal Teknik Informatika dan Sistem Informasi. 2022 Aug 26;8(2).

