



## Optimasi Algoritma Random Forest dalam Mengukur Kepuasan Peserta Pelatihan Guru pada Lembaga HAFECS

**Radhitya Abdi Nurhafiz<sup>1\*</sup>, Finki Dona Marleny<sup>2</sup>, Ayu Ahadi Ningrum<sup>3</sup>**

<sup>1,2,3</sup>Teknik Informatika, Universitas Muhammadiyah Banjarmasin, Kalimantan Selatan, Indonesia

Email: <sup>1,\*</sup>[radhitya\\_abdi\\_2255201110029@umbjm.ac.id](mailto:radhitya_abdi_2255201110029@umbjm.ac.id), <sup>2</sup>[finkidona@umbjm.ac.id](mailto:finkidona@umbjm.ac.id), <sup>3</sup>[ayuhadi@umbjm.ac.id](mailto:ayuhadi@umbjm.ac.id)

Email Penulis Korespondensi: <sup>1</sup>[radhitya\\_abdi\\_2255201110029@umbjm.ac.id](mailto:radhitya_abdi_2255201110029@umbjm.ac.id)

**Abstrak**—Peningkatan kualitas layanan pelatihan guru memerlukan evaluasi kepuasan peserta yang akurat dan berbasis data, mengingat hasil evaluasi menjadi dasar penting dalam perbaikan mutu program dan pengambilan keputusan lembaga. Kompleksitas data survei kepuasan serta ketidakseimbangan distribusi kelas merupakan tantangan utama yang membatasi kemampuan pendekatan evaluasi konvensional dalam mengidentifikasi pola kepuasan secara mendalam. Penelitian ini bertujuan mengoptimasi algoritma Random Forest untuk klasifikasi kepuasan peserta pelatihan guru di Lembaga HAFECS. Metode penelitian menggunakan pendekatan *Cross-Industry Standard Process for Data Mining (CRISP-DM)* hingga tahap evaluasi, dengan penerapan transformasi data, penanganan nilai ekstrem, *feature engineering*, serta optimasi hyperparameter menggunakan *RandomizedSearchCV*. Evaluasi kinerja model dilakukan menggunakan *accuracy*, *F1-score*, *recall* kelas minoritas, AUC-ROC, stabilitas *cross-validation*, dan *overfitting gap*. Hasil penelitian menunjukkan bahwa model Random Forest teroptimasi mencapai akurasi 70,00%, F1-score 70,10%, AUC-ROC 69,42%, serta *overfitting gap* terendah sebesar 4,20%. *Feature engineering* meningkatkan kemampuan generalisasi model, sementara *feature selection* agresif justru menurunkan performa. Model yang dihasilkan berpotensi mendukung evaluasi kepuasan peserta pelatihan secara objektif dan sistematis.

**Kata Kunci:** Kepuasan Peserta, Random Forest, *Hyperparameter Tuning*, *Feature Engineering*, Klasifikasi

**Abstract**—Improving the quality of teacher training services requires accurate and data-driven evaluation of participant satisfaction, given that evaluation results provide an important basis for program improvement and institutional decision-making. The complexity of satisfaction survey data and class imbalance present major challenges that limit the ability of conventional evaluation approaches to identify satisfaction patterns in detail. This study aims to optimize the Random Forest algorithm for classifying participant satisfaction at HAFECS. The research method follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework up to the evaluation stage, incorporating data transformation, outlier handling, feature engineering, and hyperparameter optimization using RandomizedSearchCV. Model performance was evaluated using accuracy, F1-score, minority-class recall, area under the receiver operating characteristic curve, cross-validation stability, and overfitting gap. The optimized Random Forest model achieved an accuracy of 70.00%, an F1-score of 70.10%, an area under the receiver operating characteristic curve of 69.42%, and the lowest overfitting gap of 4.20%. Feature engineering improved model generalization, while aggressive feature selection reduced model performance. The proposed model can support objective and systematic evaluation of training participant satisfaction.

**Keywords:** Participant Satisfaction, Random Forest, Hyperparameter Optimization, Feature Engineering, Classification

### 1. PENDAHULUAN

Peningkatan kualitas sumber daya manusia merupakan aspek fundamental dalam pengembangan sektor pendidikan dan pelatihan, khususnya pada pelatihan guru. Lembaga pelatihan memiliki peran strategis dalam membentuk kompetensi profesional yang relevan dengan kebutuhan peserta, sejalan dengan kewajiban pengembangan kompetensi sumber daya manusia yang diatur dalam kebijakan nasional [1]. Kualitas layanan menjadi indikator keberhasilan program yang tercermin melalui tingkat kepuasan peserta, yang selanjutnya berperan sebagai dasar evaluasi mutu layanan dan keberlanjutan program pelatihan [2]. Pengelolaan dan analisis data kepuasan yang tepat menjadi kebutuhan penting bagi lembaga pelatihan.

Seiring meningkatnya jumlah peserta dan variasi indikator penilaian, data kepuasan yang dihasilkan menjadi semakin kompleks dan beragam. Dalam praktik evaluasi, pendekatan statistik deskriptif sering digunakan untuk memperoleh gambaran umum tingkat kepuasan peserta. Evaluasi kepuasan dimanfaatkan untuk menilai kesesuaian antara kebutuhan peserta dan layanan pembelajaran yang diberikan sebagai bagian dari upaya peningkatan kinerja program [3]. Namun, pendekatan evaluatif yang bersifat umum tersebut memiliki keterbatasan dalam mengidentifikasi hubungan non-linear dan interaksi kompleks antarvariabel ketika data digunakan untuk tujuan klasifikasi tingkat kepuasan. Dalam konteks ini, penerapan metode pembelajaran mesin menjadi relevan sebagai pendekatan tambahan yang mampu memodelkan pola data secara lebih mendalam, terutama untuk mengidentifikasi faktor-faktor yang mempengaruhi kepuasan secara komprehensif [4].

Salah satu tantangan utama dalam klasifikasi data kepuasan adalah ketidakseimbangan kelas, di mana jumlah peserta yang menyatakan puas lebih dominan dibandingkan peserta yang tidak puas. Kondisi ini menyebabkan algoritma klasifikasi cenderung mempelajari karakteristik kelas mayoritas dan kurang mampu mengenali pola pada kelas minoritas [5]. Akibatnya, model dapat menghasilkan nilai akurasi yang terlihat baik secara keseluruhan, tetapi memiliki kemampuan terbatas dalam mengklasifikasikan peserta yang tidak puas [6]. Dalam evaluasi layanan, informasi dari kelas minoritas





memiliki nilai strategis karena mencerminkan aspek layanan yang perlu diperbaiki. Ketidakakuratan dalam mengklasifikasikan kelas ini dapat mengurangi kualitas hasil analisis dan berpotensi menghambat upaya peningkatan mutu layanan.

Random Forest merupakan salah satu algoritma klasifikasi yang banyak digunakan dalam analisis kepuasan karena kemampuannya menangani data berdimensi tinggi serta hubungan non-linear antarvariabel. Algoritma ini bekerja dengan mengombinasikan sejumlah pohon keputusan untuk menghasilkan keputusan klasifikasi yang lebih stabil dibandingkan model tunggal, dan dilaporkan memiliki kinerja yang lebih baik dibandingkan penggunaan model pohon keputusan secara individual [7]. Penelitian sebelumnya menunjukkan bahwa Random Forest memiliki kinerja yang baik pada berbagai permasalahan klasifikasi, termasuk pada data dengan distribusi kelas tidak seimbang. Sebagai metode non-parametrik, Random Forest lebih fleksibel karena tidak mengasumsikan bentuk distribusi data tertentu dan mampu mempelajari pola kompleks secara otomatis [8]. Algoritma ini juga menyediakan informasi tingkat kepentingan fitur yang membantu interpretasi kontribusi setiap variabel terhadap hasil klasifikasi. Namun, penerapan Random Forest tanpa pengaturan parameter yang tepat tetap berpotensi menghasilkan model yang bias terhadap kelas mayoritas, karena ketidakseimbangan distribusi kelas dapat menyebabkan bias prediksi pada model [9].

Kondisi tersebut teridentifikasi pada data kepuasan peserta pelatihan guru di Lembaga HAFECS. Pengujian awal menggunakan algoritma Random Forest dengan konfigurasi *default* menunjukkan penurunan kinerja ketika model diterapkan pada data pengujian. Kemampuan model dalam mengklasifikasikan peserta yang tidak puas masih terbatas, karena Random Forest memiliki keterbatasan ketika dihadapkan pada data yang tidak seimbang sehingga dapat menyebabkan bias dan menurunkan kemampuan generalisasi model [10]. Pada distribusi kelas yang tidak seimbang, model cenderung lebih mempelajari pola dari kelas dominan sehingga nilai *recall* dan *precision* pada kelas minoritas menjadi rendah [11].

Penelitian terdahulu telah mengkaji optimasi algoritma Random Forest melalui pengaturan *hyperparameter* untuk meningkatkan kinerja klasifikasi. Metode *hyperparameter tuning* yang umum digunakan meliputi *Grid Search* dan *Random Search*. Penggunaan *GridSearchCV* dalam penyetelan *hyperparameter* terbukti mampu mengoptimalkan performa model melalui eksplorasi sistematis terhadap ruang parameter yang tersedia [12], sementara *Randomized Search* merupakan teknik optimasi yang menggunakan kombinasi parameter secara acak untuk menemukan konfigurasi optimal pada model *machine learning* [13]. Teknik optimasi seperti *Grid Search* dan *Random Search* berperan dalam meningkatkan akurasi model Random Forest, termasuk pada data dengan *noise* dan anomali yang tinggi [14]. Penelitian-penelitian tersebut umumnya berfokus pada optimasi *hyperparameter* dan belum mengintegrasikan rekayasa fitur secara terstruktur untuk memperkaya representasi data, khususnya pada konteks data kepuasan peserta pelatihan.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan mengoptimasi algoritma Random Forest untuk klasifikasi kepuasan peserta pelatihan guru di lembaga HAFECS melalui pendekatan bertahap dan terintegrasi. Kontribusi penelitian ini meliputi penerapan *hyperparameter tuning* berbasis *Random Search* untuk menekan *overfitting* serta penerapan rekayasa fitur untuk merepresentasikan interaksi antarvariabel secara lebih baik. Pendekatan ini diharapkan menghasilkan model klasifikasi yang stabil dan seimbang sehingga dapat dimanfaatkan sebagai dasar sistem pendukung keputusan dalam evaluasi kepuasan peserta pelatihan.

## 2. METODOLOGI PENELITIAN

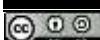
Penelitian ini menggunakan pendekatan *Cross-Industry Standard Process for Data Mining (CRISP-DM)* yang menyediakan kerangka kerja sistematis untuk proyek data mining melalui enam tahapan berurutan: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [15]. Adapun tahapan yang diterapkan mencakup lima tahapan awal CRISP-DM, yaitu *business understanding hingga evaluation*. Tahap *deployment* tidak disertakan karena penelitian berfokus pada evaluasi kinerja optimasi algoritma Random Forest dalam konteks akademis. Diagram alur penelitian disajikan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

### 2.1 Business Understanding

Tahap ini bertujuan memahami permasalahan Lembaga HAFECS dalam mengevaluasi kepuasan peserta pelatihan secara objektif dan sistematis. Tujuan penelitian adalah mengembangkan model klasifikasi kepuasan dengan akurasi optimal menggunakan algoritma Random Forest yang dioptimasi. Permasalahan utama adalah kebutuhan pendekatan pembelajaran mesin untuk mengekstraksi pola kompleks dari data survei yang tidak dapat ditangkap secara optimal melalui analisis deskriptif.





## 2.2 Data Understanding

Data penelitian diperoleh dari survei kepuasan peserta pelatihan Lembaga HAFECS periode Mei hingga Oktober 2025. Dataset awal terdiri dari 1.096 responden dengan 30 kolom yang mencakup informasi waktu pengisian, identitas kegiatan dan pengajar, karakteristik peserta, evaluasi dampak pelatihan, materi pembelajaran, keunggulan program, saran perbaikan, dan kepuasan sebagai variabel target.

Tahap pemilihan fitur dilakukan untuk mengidentifikasi subset fitur yang paling relevan dan informatif terhadap variabel target [16]. Proses ini bertujuan meningkatkan kinerja model melalui pengurangan dimensi data dan penurunan kompleksitas komputasi. Berdasarkan analisis relevansi dan kelengkapan data, 9 atribut terpilih dari 30 atribut awal.

Atribut yang dieliminasi meliputi atribut identitas seperti waktu pengisian dan nama peserta karena tidak memiliki hubungan dengan kepuasan serta atribut teks panjang yang diisi bebas karena banyak mengandung nilai kosong atau tanda strip (-) yang dapat menyulitkan pengolahan. Sembilan atribut terpilih adalah nama kegiatan, trainer, pelaksanaan kegiatan, usia, lama mengajar, asal jenjang, materi yang ingin dipelajari, keunggulan, dan kepuasan sebagai target klasifikasi dua kelas (puas/tidak puas). Distribusi kelas target ditunjukkan pada Tabel 1.

**Tabel 1.** Distribusi Kelas Target

Kelas	Jumlah Sampel	Persentase
Tidak Puas (0)	358	32,7%
Puas (1)	738	67,3%
Total	1.096	100%

Tabel 1 menunjukkan distribusi kelas dengan ketidakseimbangan rasio 2,06:1, dimana kelas puas berjumlah 738 sampel, 67,3% lebih dominan dibanding kelas tidak puas berjumlah 358 sampel, 32,7%.

## 2.3 Data Preparation

Sembilan atribut terpilih ditransformasi menjadi representasi numerik melalui pengkodean bertahap. Tahun lahir dikonversi menjadi usia, nama kegiatan menghasilkan 5 fitur biner, asal jenjang menjadi 4 fitur, *trainer* menjadi 11 fitur, lama mengajar diberi label ordinal menjadi *level\_mengajar*, materi yang ingin dipelajari dan keunggulan dihitung frekuensi kemunculannya menjadi *total\_materi* dan *total\_keunggulan*, serta pelaksanaan kegiatan dikodekan secara biner, yaitu *offline* (0) dan *online* (1). Proses ini menghasilkan ekspansi dari 9 atribut awal menjadi 26 variabel, yang terdiri atas 25 fitur input dan 1 variabel target, dengan jumlah data tetap sebanyak 1.096 baris.

Nilai ekstrem pada fitur kontinu ditangani menggunakan metode *Interquartile Range (IQR)*. Penentuan batas nilai ekstrem dirumuskan pada Persamaan (1) dan Persamaan (2):

$$\text{Batas Bawah} = Q_1 - 2,5 \times IQR \quad (1)$$

$$\text{Batas Atas} = Q_3 + 2,5 \times IQR \quad (2)$$

Keterangan variabel pada Persamaan (1) dan Persamaan (2) meliputi  $Q_1$  sebagai kuartil pertama dan  $Q_3$  sebagai kuartil ketiga. Nilai *Interquartile Range* dihitung menggunakan Persamaan (3):

$$IQR = Q_3 - Q_1 \quad (3)$$

Sebanyak 78 data yang berada di luar rentang IQR dilakukan pembatasan nilai (*clipping*) guna mempertahankan jumlah sampel dan menjaga kestabilan distribusi data.

Implementasi *feature selection* dilakukan menggunakan pendekatan univariat dengan mengevaluasi setiap fitur secara independen berdasarkan nilai *statistic* [16]. Seleksi awal menerapkan *Variance Threshold* untuk menghapus fitur bervarians rendah dan penghapusan fitur berkorelasi tinggi ( $>0,95$ ), yang mengeliminasi satu fitur (*trainer\_f*) sehingga tersisa 24 fitur. Selanjutnya, empat fitur turunan dibentuk, yaitu *usia\_level\_interaction* (interaksi usia dan pengalaman mengajar), *materi\_keunggulan\_ratio* (rasio materi dan keunggulan), *total\_engagement* (penjumlahan *total\_materi* dan *total\_keunggulan*), serta *materi\_per\_level* (rata-rata materi per level pengalaman) proses ini meningkatkan total menjadi 28 fitur.

Dataset dibagi menjadi data latih (876 sampel, 80%) dan data uji (220 sampel, 20%) menggunakan *stratified splitting* untuk mempertahankan proporsi kelas. Standardisasi fitur kontinu menggunakan *StandardScaler* dengan transformasi sebagai berikut:

$$z = \frac{\chi - \mu}{\sigma} \quad (4)$$

Dimana  $\chi$  adalah nilai asli,  $\mu$  adalah rata-rata, dan  $\sigma$  adalah simpangan baku.





## 2.4 Modeling

Penelitian ini menerapkan tiga skenario Random Forest dengan tingkat optimasi berbeda untuk mengevaluasi dampak optimasi terhadap kinerja model.

- Model 1 (RF *Default*) menggunakan *hyperparameter default* ( $n\_estimators=100$ ) tanpa teknik penyeimbangan data sebagai pembanding;
- Model 2 (RF + *RandomizedSearchCV*) menerapkan *hyperparameter tuning* untuk memperoleh kombinasi parameter optimal yang memaksimalkan kinerja model [17]. Optimasi dilakukan menggunakan *RandomizedSearchCV* dengan 100 iterasi dan *10-fold Stratified Cross-Validation*. Ruang pencarian parameter mencakup  $n\_estimators$  (200–350),  $max\_depth$  (18–30),  $min\_samples\_split$  (10–22),  $min\_samples\_leaf$  (6–12),  $max\_features$  (sqrt, log2, 0,5–0,65),  $max\_samples$  (0,7–0,9), serta  $min\_impurity\_decrease$  (0,0–0,002);
- Model 3 (RF + *RandomizedSearchCV* + *Feature Selection*) mengombinasikan optimasi *hyperparameter* dengan seleksi fitur lanjutan. Pemilihan fitur dilakukan menggunakan tiga pendekatan, yaitu *Feature Importance*, *Mutual Information*, dan *SelectFromModel*, dengan mekanisme *majority voting*, dimana fitur dipertahankan apabila terpilih oleh minimal dua metode. Seluruh proses pemodelan menggunakan  $random\_state = 42$  untuk memastikan reproduksibilitas hasil eksperimen.

## 2.5 Evaluation

Evaluasi kinerja model dilakukan menggunakan enam metrik, yaitu *accuracy*, *F1-score*, *recall* per kelas, *AUC-ROC*, *cross-validation F1*, dan *overfitting gap*. Pemilihan model terbaik ditentukan menggunakan skor gabungan terbobot yang dirumuskan sebagai berikut:

$$SG = 0,30 \times F1 + 0,30 \times Recall_0 + 0,20 \times AUC + 0,15 \times CV\_F1 - 0,05 \times Overfitting \quad (5)$$

Perancangan bobot pada Skor Gabungan (SG) di rumus (5) menempatkan *F1-score* dan *recall* kelas tidak puas sebagai metrik utama untuk memastikan kemampuan deteksi kelas minoritas, diikuti AUC untuk menilai kemampuan pemisahan kelas, CV F1 untuk kestabilan model, serta penalti *overfitting* guna menghindari model yang terlalu spesifik terhadap data latih.

## 3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dan pembahasan evaluasi tiga konfigurasi model Random Forest, yaitu model *default*, model dengan optimasi *hyperparameter*, dan model dengan kombinasi optimasi *hyperparameter* serta *feature selection*, untuk menentukan model terbaik dalam klasifikasi kepuasan peserta pelatihan.

### 3.1 Karakteristik dan Persiapan Dataset

#### 3.1.1 Distribusi dan Transformasi Data

Dataset penelitian terdiri dari 1.096 sampel dengan distribusi kelas tidak seimbang, dimana 738 sampel (67,3%) termasuk kategori puas dan 358 sampel (32,7%) kategori tidak puas, menghasilkan rasio ketidakseimbangan 2,06:1. Kondisi ini menyebabkan algoritma pembelajaran mesin cenderung mendominasi pembelajaran pada kelas mayoritas, mengakibatkan kemampuan klasifikasi kelas minoritas menurun karena model lebih fokus mempelajari pola kelas mayoritas [5].

Proses transformasi data menghasilkan ekspansi dari 9 atribut terpilih menjadi 26 fitur yang terdiri dari 25 fitur input dan 1 variabel target (kepuasan) melalui pengkodean kategorikal, kemudian dikurangi menjadi 24 fitur setelah eliminasi multikolinearitas, dan akhirnya diperluas menjadi 28 fitur setelah *feature engineering*. Transformasi atribut kategorikal dilakukan secara sistematis: nama kegiatan dengan 5 kategori menghasilkan 5 fitur biner, asal jenjang menghasilkan 4 fitur, dan trainer menghasilkan 11 fitur menggunakan *one-hot encoding*. Atribut lama mengajar ditransformasi menjadi level\_mengajar dengan label ordinal untuk menangkap hierarki pengalaman. Atribut materi yang ingin dipelajari dan keunggulan program dihitung frekuensi kemunculannya menjadi total\_materi dan total\_keunggulan untuk mengukur tingkat keterlibatan peserta.

#### 3.1.2 Penanganan Nilai Ekstrem dan Feature Engineering

Pendeteksian nilai ekstrem menggunakan metode IQR dengan faktor 2,5 mengidentifikasi 78 data anomali pada fitur kontinu yang tersebar pada usia (23 data), level\_mengajar (19 data), total\_materi (18 data), dan total\_keunggulan (18 data). Strategi *clipping* dipilih untuk mempertahankan jumlah sampel karena analisis menunjukkan sebagian data ekstrem berasal dari peserta dengan karakteristik unik namun valid. Penghapusan fitur *trainer\_f* yang berkorelasi  $>0,95$  dengan fitur lain mengurangi masalah multikolinearitas yang dapat mengganggu stabilitas model.

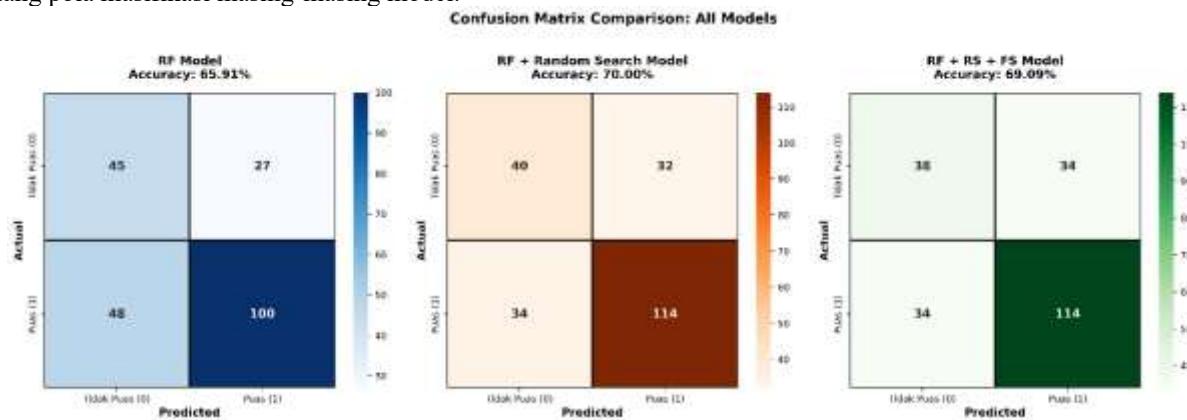




*Feature engineering* menghasilkan 4 fitur turunan yang terbukti informatif: *total\_engagement* (penjumlahan *total\_materi* dan *total\_keunggulan*), *usia\_level\_interaction* (perkalian *usia* dan *level pengalaman*), *materi\_keunggulan\_ratio* (perbandingan materi dengan keunggulan), dan *materi\_per\_level* (rata-rata materi per level). Standardisasi menggunakan *StandardScaler* diterapkan setelah *train-test split* untuk mencegah *data leakage*, dimana parameter dihitung dari data latih dan diterapkan pada data uji.

### 3.2 Evaluasi Performa Model

Evaluasi performa dilakukan terhadap tiga model Random Forest dengan konfigurasi berbeda berikut perbandingan *confusion matrix* ketiga model disajikan pada Gambar 3 untuk memberikan gambaran komprehensif tentang pola klasifikasi masing-masing model.



Gambar 2. Perbandingan Confusion Matrix dan Akurasi Ketiga Model

#### 3.2.1 Random Forest (Model 1)

Model 1 dengan parameter *default* menghasilkan akurasi uji 65,91%, F1-score 66,78%, *recall* kelas tidak puas 62,50%, *recall* kelas puas 67,57%, dan AUC-ROC 69,36%. Selisih *overfitting* mencapai 10,35% mengindikasikan model terlalu menyesuaikan dengan pola spesifik data latih. Random Forest memiliki keunggulan dalam mengurangi risiko *overfitting* melalui mekanisme agregasi dari *multiple* pohon keputusan, yang meningkatkan kemampuan generalisasi dan akurasi prediksi model [18], namun keunggulan ini belum optimal karena parameter belum teroptimasi.

*Confusion matrix* pada Gambar 2 menunjukkan dari 72 sampel tidak puas: 45 diprediksi benar (*true negative*) dan 27 diprediksi salah sebagai puas (*false negative*), berarti 37,5% peserta tidak puas terlewatkan. Dari 148 sampel puas: 100 diprediksi benar (*true positive*) dan 48 diprediksi salah sebagai tidak puas (*false positive*). Presisi kelas tidak puas 48,39% menunjukkan dari 93 prediksi tidak puas, hanya 45 yang benar, mengindikasikan tingginya *false positive rate*. OOB score 63,01% memberikan referensi konsistensi internal model tanpa memerlukan data validasi terpisah.

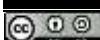
#### 3.2.2 Optimasi Hyperparameter (Model 2)

*RandomizedSearchCV* dengan 100 iterasi dan *10-fold stratified cross-validation* mengidentifikasi parameter optimal: *n\_estimators*=250, *max\_depth*=22, *min\_samples\_split*=10, *min\_samples\_leaf*=8, *max\_features*='log2', *max\_samples*=0,8, dan *min\_impurity\_decrease*=0,0. Proses sistematis untuk menemukan kombinasi nilai *hyperparameter* terbaik ini terbukti mengoptimalkan kinerja model *machine learning* [17].

Model 2 mencapai akurasi uji 70,00%, meningkat 4,09 poin persentase dari Model 1 (65,91%). *F1-score* meningkat dari 66,78% menjadi 70,10%, peningkatan sebesar 3,32 poin persentase. AUC-ROC sedikit meningkat dari 69,36% menjadi 69,42%. Yang paling signifikan adalah penurunan drastis *overfitting gap* dari 10,35% menjadi 4,20%, penurunan sebesar 6,15 poin persentase, menunjukkan model lebih seimbang antara pembelajaran pola dan kemampuan generalisasi.

*Confusion matrix* pada Gambar 2 menunjukkan dari 72 sampel tidak puas: 40 diprediksi benar (*true negative*) dan 32 diprediksi salah sebagai puas (*false negative*). Dari 148 sampel puas: 114 diprediksi benar (*true positive*) dan 34 diprediksi salah sebagai tidak puas (*false positive*). *Recall* kelas tidak puas turun menjadi 55,56%, penurunan 6,94 poin persentase dari Model 1, namun *recall* kelas puas meningkat signifikan menjadi 77,03%, peningkatan 9,46 poin persentase. *Trade-off* ini dapat diterima secara metodologis mengingat peningkatan substantial pada generalisasi model secara keseluruhan.

Parameter *max\_features*='log2' membatasi sekitar 5 fitur per pemisahan dari 28 fitur total, meningkatkan keragaman pohon sehingga setiap pohon dalam *ensemble* belajar dari perspektif berbeda dan mengurangi korelasi antar pohon. Parameter *max\_depth*=22 mencegah pohon terlalu kompleks, sementara *min\_samples\_leaf*=8 memastikan





prediksi lebih *reliable*. *CV F1-score* meningkat dari  $64,82\% \pm 4,09\%$  pada Model 1 menjadi  $65,11\% \pm 3,95\%$  pada Model 2, peningkatan 0,29 poin persentase dengan variasi yang lebih rendah, menunjukkan kestabilan lebih baik.

### 3.2.3 Hyperparameter Tuning dengan Feature Selection (Model 3)

Model 3 menerapkan *feature selection* berbasis *ensemble voting* menggunakan tiga metode: *Feature Importance*, *Mutual Information*, dan *SelectFromModel*. Sistem voting mayoritas memilih 16 fitur dari 28 fitur yang lolos minimal 2 dari 3 metode. Namun, reduksi fitur justru menurunkan performa yang mengakibatkan akurasi turun menjadi 69,09% (0,91 poin persentase lebih rendah dari Model 2), *recall* kelas tidak puas turun menjadi 52,78% (turun 2,78 poin persentase), AUC-ROC turun menjadi 68,25% (turun 1,17 poin persentase). Variasi CV F1 meningkat menjadi  $\pm 4,86\%$ , menunjukkan kestabilan model berkurang dibandingkan Model 2 yang memiliki variasi  $\pm 3,95\%$ .

*Confusion matrix* pada Gambar 3 menunjukkan dari 72 sampel tidak puas: 38 diprediksi benar (*true negative*) dan 34 diprediksi salah sebagai puas (*false negative*). Dari 148 sampel puas: 114 diprediksi benar (*true positive*, sama dengan Model 2) dan 34 diprediksi salah sebagai tidak puas (*false positive*, sama dengan Model 2). Penurunan *true negative* dari 40 (Model 2) menjadi 38 (Model 3) dan peningkatan *false negative* dari 32 menjadi 34 menunjukkan kemampuan model dalam mengidentifikasi kelas tidak puas menurun setelah *feature selection*.

Analisis mendalam terhadap 12 fitur tereliminasi menunjukkan bahwa meskipun memiliki *importance score* individual rendah, fitur-fitur tersebut berkontribusi kolektif dalam membedakan pola kelas minoritas. Fenomena ini mengonfirmasi bahwa dalam konteks klasifikasi dengan *data imbalance*, fitur dengan *importance* rendah secara individual dapat memiliki kontribusi penting dalam mengidentifikasi pola kelas minoritas [5], sehingga *feature selection* yang terlalu agresif kontraproduktif.

### 3.3 Faktor-Faktor yang Mempengaruhi Performa Model

Akurasi 70,00% yang dicapai Model 2 dipengaruhi beberapa faktor berdasarkan hasil eksperimen. Pertama, ketidakseimbangan kelas dengan perbandingan 2,06:1 membatasi pembelajaran pola kelas minoritas, tercermin dari presisi kelas tidak puas (54,05%) yang lebih rendah dari kelas puas (78,08%). Kedua, parameter optimal yang ditemukan *RandomizedSearchCV* terbukti mengurangi selisih *overfitting* dari 10,35% (Model 1) menjadi 4,20% (Model 2), sejalan dengan keunggulan Random Forest dalam mengurangi risiko *overfitting* melalui mekanisme agregasi pohon keputusan [18]. Parameter *max\_features='log2'* (sekitar 5 fitur per pemisahan dari 28 fitur total) meningkatkan keragaman pohon, terlihat dari penurunan variasi *CV F1-score* dari  $\pm 4,09\%$  menjadi  $\pm 3,95\%$ .

Ketiga, analisis *feature importance* menunjukkan *total\_keunggulan* (0,1326) dan *usia* (0,1256) sebagai faktor dominan. Penerapan Random Forest *Feature Importance* berhasil mengidentifikasi fitur-fitur yang memiliki pengaruh dominan terhadap klasifikasi [19]. *Total\_keunggulan* menangkap persepsi peserta terhadap nilai tambah program, sementara *usia* merepresentasikan karakteristik demografi yang memengaruhi pola kepuasan. Keempat, batas atas akurasi sekitar 70% pada ketiga model mengindikasikan ada dimensi kepuasan yang tidak tertangkap oleh 28 fitur final yang tersedia. Penurunan performa Model 3 meskipun menggunakan *feature selection* menunjukkan 12 fitur yang dieliminasi membawa informasi yang diperlukan untuk klasifikasi, terutama untuk kelas minoritas yang *recall*-nya turun dari 55,56% menjadi 52,78%.

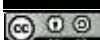
### 3.4 Analisis Komparatif dan Pemilihan Model Terbaik

Evaluasi menggunakan skor gabungan berdasarkan rumus (5) menghasilkan: Model 1 = 0,6186, Model 2 = 0,6124, dan Model 3 = 0,5957. Meskipun Model 1 memiliki skor gabungan tertinggi, selisih *overfitting* yang sangat besar (10,35%) mengindikasikan model tidak *reliable* untuk data baru. Dengan demikian, skor gabungan digunakan sebagai indikator kuantitatif awal, sedangkan pemilihan akhir model mempertimbangkan kemampuan generalisasi dan stabilitas, yang lebih krusial untuk penerapan pada data baru. Model 2 dipilih sebagai model terbaik berdasarkan evaluasi holistik yang mempertimbangkan keseimbangan seluruh aspek: akurasi tertinggi (70,00%), *overfitting* terendah (4,20%), stabilitas terbaik (CV F1  $65,11\% \pm 3,95\%$ ), dan AUC-ROC tertinggi (69,42%).

Tabel 2. Perbandingan Performa Ketiga Model

Model	Accuracy	F1-Score	Recall-0	Overfitting Gap
RF	0,659	0,668	0,625	0,103
RF + <i>RandomizedSearchCV</i>	0,700	0,701	0,556	0,042
RF + <i>RandomizedSearchCV</i> + Feature Selection	0,691	0,691	0,528	0,066

Perbandingan performa pada Tabel 2 menunjukkan keunggulan Model 2 dengan akurasi tertinggi dan *overfitting* minimal. Model 1 memiliki *recall* kelas tidak puas tertinggi (62,50%) namun selisih *overfitting* yang besar (10,35%) mengindikasikan model terlalu menyesuaikan dengan data latih. Model 3 mengalami penurunan di semua metrik: akurasi turun menjadi 69,09%, *recall* kelas tidak puas turun menjadi 52,78%, dan AUC-ROC turun menjadi 68,25%, menunjukkan bahwa pengurangan fitur dari 28 menjadi 16 justru mengurangi kemampuan klasifikasi, terutama pada kelas minoritas yang paling memerlukan informasi dari berbagai fitur untuk dapat diklasifikasikan dengan akurat.

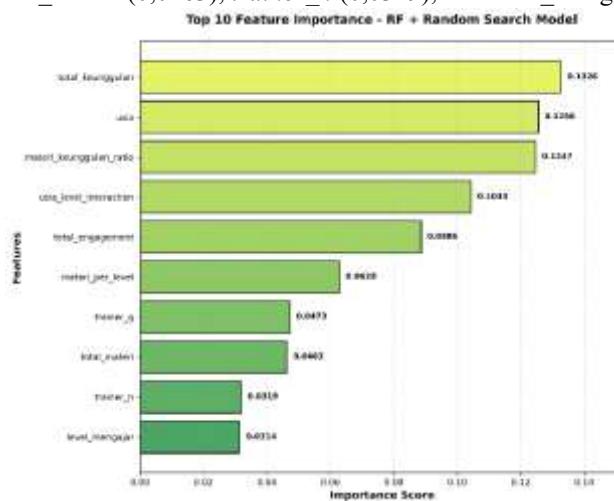




Perbandingan dengan penelitian terdahulu menunjukkan konsistensi dengan temuan bahwa teknik optimasi seperti *Grid Search* dan *RandomizedSearchCV* dilaporkan berperan dalam meningkatkan akurasi model Random Forest, termasuk pada data dengan *noise* dan anomali yang tinggi [14]. Namun, penelitian ini berkontribusi dengan menunjukkan bahwa pada klasifikasi kepuasan berbasis data *imbalanced*, *feature engineering* lebih efektif dibanding *feature selection* agresif dalam meningkatkan generalisasi model Random Forest.

### 3.5 Analisis Feature Importance dan Faktor Kepuasan

Penerapan Random Forest *Feature Importance* berhasil mengidentifikasi fitur-fitur yang memiliki pengaruh dominan terhadap klasifikasi, memungkinkan fokus pada variabel yang paling berkontribusi terhadap performa model [19]. Analisis Model 2 mengidentifikasi total\_keunggulan (0,1326) sebagai fitur terpenting, diikuti usia (0,1256), materi\_keunggulan\_ratio (0,1247), usia\_level\_interaction (0,1043), total\_engagement (0,0986), materi\_per\_level (0,0628), trainer\_g (0,0473), total\_materi (0,0463), trainer\_h (0,0319), dan level\_mengajar (0,0314).



Gambar 3. Top 10 Fitur Terpenting Model 2

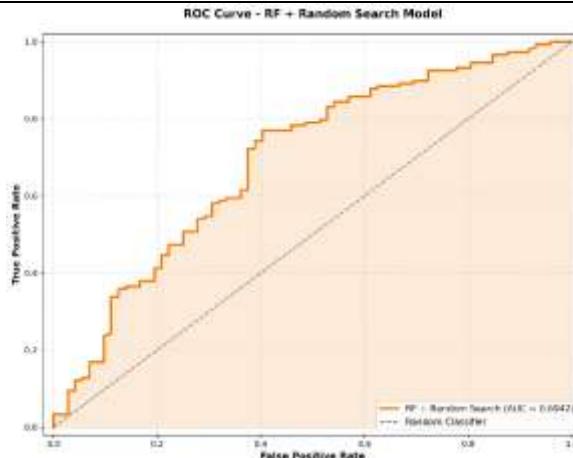
Total\_keunggulan sebagai faktor paling dominan pada Gambar 3 menunjukkan nilai tambah program berpengaruh besar terhadap kepuasan peserta. Fitur ini memiliki *importance* tertinggi (0,1326), mengindikasikan bahwa persepsi peserta terhadap keunggulan program pelatihan menjadi faktor utama dalam menentukan kepuasan. Usia menempati posisi kedua (0,1256), menunjukkan karakteristik demografi peserta memiliki pengaruh signifikan terhadap pola kepuasan.

Empat dari 10 fitur teratas merupakan hasil *feature engineering* (materi\_keunggulan\_ratio, usia\_level\_interaction, total\_engagement, dan materi\_per\_level), membuktikan transformasi fitur berhasil menangkap pola kompleks dalam data. Materi\_keunggulan\_ratio (0,1247) di posisi ketiga menunjukkan rasio antara materi yang disampaikan dengan keunggulan program yang dipersepsikan memberikan informasi penting untuk klasifikasi. *Usia\_level\_interaction* (0,1043) di posisi keempat mengonfirmasi bahwa kepuasan tidak hanya dipengaruhi oleh faktor tunggal, tetapi juga oleh kombinasi karakteristik peserta seperti interaksi antara usia dan pengalaman mengajar.

### 3.6 Kurva ROC dan Diskriminasi Model

Kurva ROC Model 2 menunjukkan AUC 69,42%, berada dalam kategori *fair-to-good discrimination* menurut klasifikasi *Hosmer-Lemeshow* yang mendefinisikan AUC 0,60-0,70 sebagai *acceptable*. Nilai ini menunjukkan bahwa jika dipilih secara acak satu sampel dari setiap kelas, model memiliki probabilitas 69,42% memberikan skor lebih tinggi pada sampel puas. Pada *threshold default* 0,50, model mencapai TPR 77,03% (114 dari 148 sampel puas teridentifikasi benar) dan FPR 44,44% (32 dari 72 sampel tidak puas salah diklasifikasikan).





Gambar 4. Kurva ROC Model 2

Kurva ROC pada Gambar 6 mengungkap *trade-off fundamental* antara sensitivitas dan spesifisitas yang dapat dikontrol melalui penyesuaian *threshold*. Pada *threshold* rendah (0,30-0,40), TPR meningkat hingga 85% namun FPR juga meningkat hingga 60%, menghasilkan banyak *false positive* dimana peserta tidak puas salah diklasifikasikan sebagai puas. Sebaliknya, *threshold* tinggi (0,60-0,70) menurunkan FPR hingga 30% namun TPR juga menurun hingga 65%, menghasilkan banyak *false negative* dimana peserta puas salah diklasifikasikan sebagai tidak puas. Dalam konteks evaluasi kepuasan peserta, *false positive* lebih dapat diterima secara operasional dibanding *false negative*, sehingga *threshold* dapat disesuaikan untuk menekan kesalahan klasifikasi peserta tidak puas dan memungkinkan lembaga melakukan intervensi preventif terhadap peserta yang berisiko tidak puas.

Model 2 memiliki AUC tertinggi (69,42%) dibanding Model 1 (69,36%) dan Model 3 (68,25%), menunjukkan kemampuan diskriminasi terbaik dalam membedakan kedua kelas. Peningkatan AUC sebesar 0,06 poin dari Model 1 ke Model 2, meskipun tampak kecil, menunjukkan perbaikan konsisten dalam kemampuan model membedakan kelas pada berbagai *threshold*, yang merupakan indikator penting untuk aplikasi praktis dimana *threshold* optimal dapat disesuaikan berdasarkan prioritas bisnis lembaga.

### 3.7 Implikasi untuk Pengembangan Program Pelatihan

Penelitian ini membuktikan Random Forest dengan optimasi parameter menggunakan *RandomizedSearchCV* mencapai akurasi 70,00%, meningkat 4,09 poin persentase dari model *default*. Peningkatan ini didorong oleh karakteristik Random Forest yang mampu meminimalisir risiko overfitting melalui penggabungan hasil klasifikasi dari berbagai pohon keputusan, sehingga menghasilkan model dengan kemampuan generalisasi yang lebih kuat [18]. Analisis *feature importance* mengidentifikasi total\_keunggulan (0,1326), usia (0,1256), dan materi\_keunggulan\_ratio (0,1247) sebagai faktor penentu utama kepuasan, memberikan panduan praktis bagi lembaga untuk fokus pada peningkatan nilai tambah program dan mempertimbangkan karakteristik demografi peserta [19]. Model klasifikasi yang dikembangkan berpotensi dimanfaatkan sebagai alat bantu evaluasi awal yang dikombinasikan dengan metode evaluasi kualitatif untuk pemahaman yang lebih komprehensif tentang kepuasan peserta pelatihan.

## 4. KESIMPULAN

Penelitian ini berhasil mengoptimasi algoritma Random Forest untuk klasifikasi kepuasan peserta pelatihan melalui integrasi *hyperparameter tuning* dan *feature engineering*. Model 2 dengan *RandomizedSearchCV* terbukti optimal dengan akurasi 70,00% (peningkatan 4,09 poin dari *default*), F1-Score 70,10%, AUC-ROC 69,42%, dan *overfitting gap* terendah 4,20%. Optimasi *hyperparameter* menemukan kombinasi optimal (*n\_estimators*=250, *max\_depth*=22, *min\_samples\_leaf*=8, *max\_features*='log2'), sementara *feature engineering* menghasilkan fitur informatif dengan total\_keunggulan (*importance* 0,1326), usia (0,1256), dan materi\_keunggulan\_ratio (0,1247) sebagai faktor dominan. Temuan penting menunjukkan *feature selection* agresif kontraproduktif pada data *imbalanced*, dimana Model 3 yang mengurangi fitur dari 28 menjadi 16 mengalami penurunan performa, mengonfirmasi bahwa fitur dengan *importance* rendah tetap berkontribusi kolektif dalam mengidentifikasi pola kelas minoritas.

Penelitian selanjutnya disarankan menerapkan teknik penanganan *imbalanced* data seperti *SMOTE* (*Synthetic Minority Over-sampling Technique*) atau kombinasi *SMOTE-Tomek Links* untuk meningkatkan performa pada kelas minoritas. Eksplorasi algoritma *ensemble* lain dan metode optimasi alternatif dapat memberikan perspektif berbeda dalam klasifikasi kepuasan, sementara penambahan fitur kualitatif seperti *sentiment analysis* dari feedback peserta berpotensi



menangkap dimensi kepuasan yang belum teridentifikasi. Identifikasi total\_keunggulan, usia, dan materi\_keunggulan\_ratio sebagai faktor dominan memberikan panduan praktis bagi lembaga untuk fokus pada peningkatan nilai tambah program dan mempertimbangkan karakteristik demografi peserta, dengan model klasifikasi sebagai alat bantu evaluasi awal yang dikombinasikan dengan metode kualitatif untuk pemahaman komprehensif kepuasan peserta pelatihan.

## UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada lembaga HAFECS atas dukungan dan akses data yang diberikan untuk kelancaran penelitian ini. Terima kasih juga disampaikan kepada dosen pembimbing yang telah memberikan arahan dan masukan konstruktif, serta seluruh peserta pelatihan yang telah berpartisipasi dalam pengisian kuesioner evaluasi yang menjadi data penting dalam penelitian ini.

## REFERENCES

- [1] W. Wajiyono, W. T. Astuti, and W. Wisnalmawati, “Pengaruh Kualitas Pelayanan terhadap Loyalitas yang dimediasi Kepuasan Peserta Pelatihan pada Badan Pendidikan dan Pelatihan DIY,” *J-MAS (Jurnal Manajemen dan Sains)*, vol. 9, no. 2, p. 872, Oct. 2024, doi: 10.33087/jmas.v9i2.1833.
- [2] Dekanawati V, Astriawati N, Setiyantara Y, and Subekti J, “ANALISIS PENGARUH KUALITAS PELAYANAN DIKLAT KEPABEANAN TERHADAP KEPUASAN PESERTA PELATIHAN,” *Jurnal Saintek Maritim*, vol. 23, no. 2, pp. 159–176, Mar. 2023, doi: 10.33556/jstm.v23i2.344.
- [3] E. Purwaningsih *et al.*, “Efektifitas Dan Kepuasan Mahasiswa Terhadap Media E-learning INFORMASI ARTIKEL ABSTRAK,” vol. 5, no. 1, pp. 126–130, 2024, doi: 10.55338/jpkmn.v5i1.
- [4] M. F. A. M. Fariz A and M. Muharrom, “Analisis Kepuasan Pengguna Terhadap Keberhasilan Pembelajaran Bahasa Pada Duolingo Menggunakan Metode PIECES Framework,” *Software Development, Digital Business Intelligence, and Computer Engineering*, vol. 3, no. 1, pp. 33–40, Sep. 2024, doi: 10.57203/session.v3i1.2024.33-40.
- [5] C. Magnolia, A. Nurhopipah, D. Bagus, and A. Kusuma, “Edu Komputika Journal Penanganan Imbalanced Dataset untuk Klasifikasi Komentar Program Kampus Merdeka Pada Aplikasi Twitter,” 2022. [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/edukom>
- [6] Z. Darojah, R. Susetyoko, N. Ramadjanti, P. Elektronika Negeri Surabaya, and P. Korespondensi, “STRATEGI PENANGANAN IMBALANCE CLASS PADA MODEL KLASIFIKASI PENERIMA KARTU INDONESIA PINTAR KULIAH BERBASIS NEURAL NETWORK MENGGUNAKAN KOMBINASI SMOTE DAN ENN,” vol. 10, no. 2, pp. 457–466, 2023, doi: 10.25126/jtiik.2023106480.
- [7] F. Dona Marleny, “PREDIKSI PEMILIHAN PROGRAM STUDI PADA MAHASISWA KIP FAKULTAS TEKNIK MENGGUNAKAN ALGORITMA RANDOM FOREST,” *Sains dan Teknologi*, vol. 12, no. 3, pp. 2025–1448, 2025, doi: 10.47668/edusaintek.v12i3.1800.
- [8] H. Argyretya, “Implementasi Random Oversampling pada Algoritma Random Forest untuk Klasifikasi Indeks Standar Udara di DKI Jakarta 2024,” *Jurnal Informatika: Jurnal pengembangan IT*, vol. 10, no. 4, pp. 2477–5126, 2025, doi: 10.30591/jpit.v10i4.9466.
- [9] H. Akbar and W. K. Sanjaya, “Kajian Performa Metode Class Weight Random Forest pada Klasifikasi Imbalance Data Kelas Curah Hujan,” *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, vol. 3, no. 1, Dec. 2023, doi: 10.20885/snati.v3i1.30.
- [10] A. N. Pratiwi and E. Utami, “Sistemasi: Jurnal Sistem Informasi Prediksi Kinerja Akademik Matematika Siswa berdasarkan Kepribadian Big Five menggunakan Random Forest dengan Teknik Synthetic Minority Over-Sampling Predicting Students’ Academic Performance in Mathematics based on Big Five Personality Traits using Random Forest with Synthetic Minority Over-Sampling Technique,” 2025. [Online]. Available: <http://sistemasii.ftik.unisi.ac.id>
- [11] A. Agustina, T. Tukino, B. Huda, and E. Novalia, “Prediksi Volume Penjualan Gadget Berdasarkan Promo dan Channel Penjualan Menggunakan Random Forest,” *JUSIFOR : Jurnal Sistem Informasi dan Informatika*, vol. 4, no. 1, pp. 85–91, Jun. 2025, doi: 10.70609/jusifor.v4i1.6962.
- [12] E. R. Susanto and A. Eka Pranajaya, “Optimasi Random Forest untuk Prediksi Penyakit Jantung Menggunakan SMOTEEENN dan Grid Search,” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 7, pp. 1965–1979, Jul. 2025, doi: 10.52436/1.jpti.855.
- [13] L. Oriana, A. Dwi, S. Wati, A. P. Ramahdani, N. N. Safira, and E. Ismanto, “Peningkatan Kinerja Model Random Forest untuk Deteksi Kecurangan Kartu Kredit Menggunakan RandomizedSearchCV”.





- [14] N. P. Nur Fauzi, S. Khomsah, and A. D. Putra Wicaksono, “Penerapan Feature Engineering dan Hyperparameter Tuning untuk Meningkatkan Akurasi Model Random Forest pada Klasifikasi Risiko Kredit,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 2, pp. 251–262, Apr. 2025, doi: 10.25126/jtiik.2025128472.
- [15] D. B. Saputra, V. Atina, F. E. Nastiti, and F. I. Komputer, “PENERAPAN MODEL CRISP-DM PADA PREDIKSI NASABAH KREDIT MENGGUNAKAN ALGORITMA RANDOM FOREST,” 2024. [Online]. Available: <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexDwiBagusSaputra>|<http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexex>
- [16] F. Fakhriza *et al.*, “OPTIMALISASI ALGORITMA RANDOM FOREST FEATURE SELECTION DAN HYPERPARAMETER TUNING KLASIFIKASI GENRE MUSIK,” 2025. [Online]. Available: <https://www.kaggle.com/datasets/maharshipa>
- [17] D. Sandhi Bhakti, A. Prasetyo, and P. Arsi, “IMPLEMENTATION OF HYPERPARAMETER TUNING IN RANDOM FOREST ALGORITHM FOR LOAN APPROVAL PREDICTION,” *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 4, pp. 63–69, Jul. 2024, doi: 10.52436/1.jutif.2024.5.4.2032.
- [18] P. H. Gunawan and I. V. Paputungan, “Sistemasi: Jurnal Sistem Informasi Deteksi Tingkat Potensi Kelulusan Calon Mahasiswa menggunakan Algoritma Random Forest Detection of Graduation Potential in Prospective Students using the Random Forest Algorithm.” [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [19] S. NURADILLA, K. SADIK, C. SUHAENI, and A. M. SOLEH, “Klasifikasi Halaman SEO Berbasis Machine Learning Melalui Mutual Information dan Random Forest Feature Importance,” *MIND Journal*, vol. 10, no. 1, pp. 114–129, Jun. 2025, doi: 10.26760/mindjournal.v10i1.114-129.

