JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9529
Hal 474-483
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Analisis Prediksi Resiko Perceraian Menggunakan Algoritma Random Forest dengan Optimasi Hyperparameter Random Search

Indah Khoirun Nisa*, Muhammad Jauhar Vikri, Denny Nurdiansyah

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
Email: [1]indahkhoirunnisa347@email.com,[2]vikri@unugiri.ac.id,[3]denny.nur@unugiri.ac.id
Email Penulis Korespondensi: indahkhoirunnisa347@email.com
Submitted **22-01-2026;** Accepted **09-03-2026;** Published **30-04-2026**

Abstrak

Perceraian merupakan permasalahan sosial yang terus meningkat dan berdampak negatif terhadap kondisi psikologis, sosial, dan ekonomi individu maupun keluarga. Penelitian ini bertujuan membangun model prediksi risiko perceraian menggunakan algoritma Random Forest dengan optimasi hyperparameter menggunakan metode Random Search. Dataset diperoleh dari platform Kaggle dengan 170 sampel dan 54 atribut psikologis-perilaku pasangan. Tahapan penelitian meliputi preprocessing data, pembagian dataset (80:20), pembangunan model baseline, optimasi hyperparameter dengan Random Search, dan evaluasi menggunakan metrik akurasi, presisi, recall, serta AUC-ROC. Hasil penelitian menunjukkan model mencapai akurasi 94,12% pada data uji dengan recall 97% yang meminimalkan risiko false negative. Optimasi hyperparameter berhasil meningkatkan stabilitas internal model dengan rata-rata validasi silang mencapai 98,57%, meskipun akurasi pada data uji setara dengan model baseline. Gap sebesar 4,45% antara validasi dan uji mengindikasikan potensi overfitting yang umum terjadi pada dataset berukuran kecil. Analisis feature importance mengungkapkan lima faktor psikologis dominan: kemampuan berkompromi, komunikasi efektif, resolusi konflik, keselarasan nilai kehidupan, dan kemampuan memaafkan. Penelitian ini berkontribusi pada pengembangan sistem deteksi dini risiko perceraian berbasis machine learning serta memberikan landasan empiris untuk intervensi konseling yang lebih terarah.

Kata Kunci: Perceraian; Random Forest; Optimasi Parameter; Klasifikasi; Machine Learning

Abstract

Divorce is a social problem that continues to increase and has negative impacts on the psychological, social, and economic conditions of individuals and families. This study aims to build a divorce risk prediction model using the Random Forest algorithm with hyperparameter optimization using the Random Search method. The dataset was obtained from the Kaggle platform with 170 samples and 54 psychological-behavioral attributes of couples. The research stages included data preprocessing, dataset splitting (80:20), baseline model development, hyperparameter optimization with Random Search, and evaluation using accuracy, precision, recall, and AUC-ROC metrics. The results showed that the model achieved 94.12% accuracy on the test data with 97% recall that minimizes false negative risk. Hyperparameter optimization successfully improved the model's internal stability with a cross-validation average of 98.57%, although the test accuracy was equivalent to the baseline model. A gap of 4.45% between validation and test accuracy indicates potential overfitting, which is common in small datasets. Feature importance analysis revealed five dominant psychological factors: willingness to compromise, effective communication, conflict resolution, alignment of life values, and forgiveness ability. This research contributes to the development of an early detection system for divorce risk based on machine learning and provides an empirical basis for more targeted counseling interventions

Keywords: Divorce; Random Rorest; Parameter Optimization; Classification; Machine Learning.

1. PENDAHULUAN

Fenomena global pernikahan anak, yang sering kali berakhir dengan perceraian, tetap menjadi isu yang mengkhawatirkan yang perlu ditangani. Pasangan muda sering mengalami ketidakmatangan emosional, yang merupakan penyebab umum masalah perkawinan, seperti yang terjadi di Ujung Kubu, Batu Bara, di mana ketidaksiapan mental terbukti berdampak besar pada tingkat perceraian. Dampak dari masalah ini melampaui peserta langsung dan menyebabkan pernikahan anak menciptakan stigma sosial, isolasi sosial, dan berbagai masalah yang berdampak negatif pada komunitas sekitarnya.[1]. Kompleksitas alasan perceraian tidak dapat dipisahkan dari berbagai faktor yang saling terkait. Masalah dengan komunikasi yang tidak efektif, ketidaksetiaan, kesulitan ekonomi, dan perbedaan latar belakang sosial dan budaya adalah pemicu utama sengketa rumah tangga. Sayangnya, konseling pranikah yang tersedia masih terlalu formal dan tidak membahas aspek psikologis secara mendalam. Situasi ini menunjukkan jurang antara kebutuhan pasangan muda akan bantuan yang komprehensif dan kenyataan layanan yang ada [2]. Berbagai studi memperkuat argumen bahwa faktor internal memiliki pengaruh lebih dominan dibandingkan faktor eksternal dalam memicu perceraian. Egoisme tercatat sebagai penyumbang terbesar dengan persentase 65,26%, diikuti lemahnya manajemen emosi sebesar 56,46%, masalah keuangan 48,6%, dan ketidaksetiaan 47,53%. Sementara itu, faktor eksternal seperti lingkungan sosial yang kurang mendukung (62,51%) dan campur tangan pihak ketiga (36,83%) turut memperburuk situasi [3]. Data ini menegaskan bahwa aspek psikologis dan perilaku pasangan memegang peranan krusial dalam menentukan keberlangsungan rumah tangga, khususnya pada pasangan usia dini yang masih dalam tahap perkembangan emosional. Perceraian juga telah dikaji dari berbagai perspektif, termasuk sudut pandang hukum dan agama. Pergeseran nilai interpersonal akibat dominasi aturan hukum, norma agama, dan tekanan sosial, di mana perceraian dibahas dalam perspektif hukum Islam dan Kristen, serta dikaitkan dengan Undang-Undang Perkawinan Nomor 1 Tahun 1974 dan perkembangan sistem *e-litigasi*. Di sisi lain, meningkatnya angka perceraian pada pernikahan usia dini menunjukkan bahwa pendekatan konvensional belum sepenuhnya efektif dalam mencegah permasalahan rumah tangga [4].

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9529
Hal 474-483
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

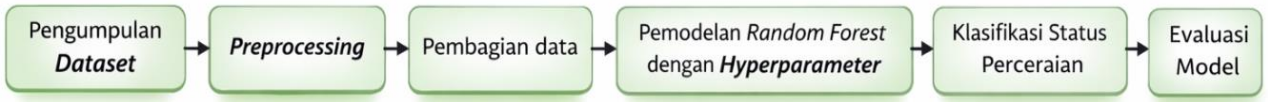
Kesenjangan antara kebutuhan akan deteksi dini risiko perceraian dengan keterbatasan pendekatan konvensional mendorong pengembangan solusi berbasis teknologi. Beberapa penelitian telah merintis jalan ke arah tersebut dengan berbagai pendekatan komputasi. Penelitian oleh Kemal Pasha dan Kusrini berhasil memprediksi tingkat perceraian secara makro di Indonesia menggunakan metode *Long Short-Term Memory* (LSTM) yang efektif mendeteksi pola data berurutan dengan kesalahan rata-rata kuadrat (MSE) terendah 0,00002033 [5]. Pendekatan berbeda dilakukan oleh Siregar et, al. yang menekankan analisis pada level individu dengan menerapkan *Random Forest* dan teknik SMOTE untuk mengklasifikasikan kategori gugatan cerai, di mana usia penggugat menjadi variabel prediktor paling berpengaruh [5]. Menariknya, kedua penelitian dengan skala berbeda ini sama-sama membuktikan bahwa *machine learning* memiliki kapabilitas sebagai alat prediktif yang andal. Penerapan *machine learning* juga telah banyak digunakan pada berbagai bidang untuk mendukung proses klasifikasi dan prediksi sehingga mampu meningkatkan kualitas pengambilan keputusan berdasarkan data[6]. Inovasi dalam metode prediksi terus berkembang untuk menghasilkan model yang tidak hanya akurat tetapi juga mudah diinterpretasikan. Penelitian lain oleh Ahsan berhasil menciptakan metode prediksi perceraian yang tidak hanya tepat, tetapi juga mudah dipahami. Strategi ini memperoleh akurasi mendekati 98,57% dengan memanfaatkan algoritma SVM, KNN, dan LDA [7]. Untuk menjelaskan hasil prediksi dengan jelas, peneliti menerapkan metode LIME yang mampu memberikan penjelasan untuk setiap hasil prediksi. Inovasi ini juga disertai dengan pembuatan aplikasi *desktop* yang mempermudah pengguna, baik pasangan maupun konselor, dalam mengimplementasikan model tersebut. Sementara itu pendekatan berbeda diterapkan oleh Tastiano dan Wakhidah menerapkan metode yang berbeda dengan penggunaan algoritma *K-Means* untuk mengategorikan area di Jawa Tengah berdasarkan insiden kekerasan terhadap perempuan [8]. Hasil analisis kluster menunjukkan bahwa Kota Semarang adalah daerah yang membutuhkan perhatian lebih karena banyaknya kasus kekerasan yang terjadi di sana. Sebuah studi yang dilakukan oleh Moumen et, al. di Arab Saudi menunjukkan bahwa Skala Prediksi Perceraian (DPS) berdasarkan Algoritma *Random Forest* mencapai akurasi 91,66% dengan enam atribut utama setelah pemilihan fitur [9]. Di Indonesia, pengembangan Analisis Sentimen Twitter terhadap Perceraian oleh Muhammad Azwar et, al. menggunakan *Random Forest* menghadapi tantangan karena ketidakseimbangan data [10]. Tantangan serupa juga dialami dalam studi sentimen KDRT di Twitter oleh Asyaroh dan Fitrani di mana algoritma *Random Forest* dan *XGBoost* mengalami *overfitting* dan kesulitan mengklasifikasikan sentimen positif akibat dominasi data negatif [11]. Sementara itu, penelitian tentang prediksi keputusan perceraian oleh Rahmadini dan Santoso berhasil mengatasi ketidakseimbangan data dengan teknik SMOTE, di mana algoritma *Naïve Bayes* mencapai *recall* 100% [12].

Berdasarkan uraian permasalahan di atas, penelitian ini menawarkan solusi berupa model prediktif yang mampu mengidentifikasi pasangan berisiko tinggi mengalami perceraian, khususnya pada kelompok pernikahan usia dini. Pendekatan yang digunakan adalah algoritma *Random Forest* dengan optimasi *hyperparameter* menggunakan *Random Search.* Pemilihan *Random Forest* didasarkan pada kemampuannya menangani data numerik multiskala bahwa algoritma *Random Forest* mampu menghasilkan performa klasifikasi yang baik pada data dengan karakteristik kompleks[13], memberikan stabilitas prediksi, menghindari *overfitting*, serta menghasilkan model yang dapat diinterpretasi melalui analisis *feature importance*. Kebaruan penelitian ini terletak pada penggunaan teknik optimasi untuk *dataset* berukuran kecil hingga menengah serta analisis fitur psikologis dominan yang dapat menjadi dasar intervensi konseling. Tujuan dari penelitian ini adalah untuk membangun model untuk memprediksi perceraian menggunakan algoritma *Random Forest* yang dioptimalkan dengan *Random Search* untuk meningkatkan kinerja klasifikasi. Penelitian ini menggunakan *dataset* sekunder yang bersumber dari Kaggle dengan total 170 sampel dan 54 atribut psikologis-perilaku. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, dan *recall; F1-score*; serta matriks kebingungan secara komprehensif. Peningkatan kinerja model ditargetkan dengan mengoptimalkan *hyperparameter* untuk mencapai *baseline* kinerja sebesar 94,12% hingga ≥95%. Kebaruan utama dari penelitian ini adalah pemanfaatan optimasi *hyperparameter* dengan pendekatan *Random Search* pada algoritma *Random Forest*, dengan objek penelitian pemodelan risiko perceraian, ini yang membuat penelitian ini unik. Selain berusaha meningkatkan akurasi model, penelitian ini menganalisis secara mendalam fitur psikologis dominan seperti berkompromi, komunikasi, dan penyelesaian konflik, dan konflik yang menjadi prediktor utama perceraian. Pendekatan ini mengintegrasikan aspek-aspek teknis dari machine learning dengan kebutuhan dalam praktik di bidang konseling keluarga. Oleh karena itu, selain akurat secara komputasi, hasil penelitian ini juga bersifat interpretatif dan aplikatif untuk pengembangan sistem deteksi dini risiko perceraian berbasis data.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan melalui beberapa proses seperti pengumpulan *dataset*, *Preprocessing dataset*, pemisahan data untuk pelatihan dan pengujian, pemodelan menggunakan algoritma *Random Forest*, optimisasi *hyperparameter*, dan evaluasi kinerja model. *Dataset* diperoleh dari repositori publik Kaggle. Tahap *Preprocessing dataset* dilakukan untuk menyiapkan data sebelum pemodelan, sementara optimisasi *hyperparameter* bertujuan untuk meningkatkan kinerja model klasifikasi. Alur keseluruhan penelitian disajikan dalam Gambar 1.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9529
Hal 474-483
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 1. Skema Penelitian Klasifikasi Status Perceraian Menggunakan Algoritma *Random Forest*.

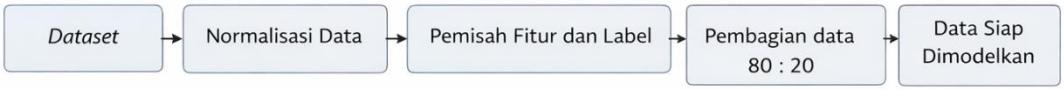
Gambar 1 menggambarkan alur penelitian yang terdiri dari lima tahap utama. Tahap Pertama adalah Pengumpulan *Dataset*, yaitu proses mengunduh data dari repositori publik. Tahap Kedua adalah Pra-pemrosesan Data, yang mencakup pembersihan dan persiapan data. Tahap Ketiga adalah Pemodelan *Random Forest* dengan Optimisasi *Hyperparameter,* yang mencakup membangun model dasar dan mencari parameter terbaik. Tahap Keempat adalah Klasifikasi Status Perceraian, di mana model yang dilatih digunakan untuk memprediksi data uji. Tahap Kelima adalah Evaluasi Pemodelan, yang merupakan pengukuran kinerja model menggunakan berbagai metrik evaluasi.

2.1 Pengumpulan *Dataset*

Dataset yang digunakan dalam penelitian ini diambil dari repositori data publik di Kaggle mencakup 170 responden dan 54 atribut yang menangkap aspek psikologis dan perilaku dari kehidupan pernikahan responden. Semua data dikumpulkan melalui kuesioner yang divalidasi menggunakan skala Likert 0–4. *Dataset* ini bersifat sekunder dan tidak mengandung pengidentifikasi pribadi. Oleh karena itu, dapat digunakan dengan aman untuk tujuan penelitian. *Dataset* ini digunakan sebagai input utama pada tahap pra-pemrosesan data sesuai dengan desain penelitian [11].

2.2 *PreProcessing Dataset*

Tahap *preprocessing* data bertujuan untuk memastikan kualitas data sebelum digunakan dalam pemodelan. Alur *preprocessing* data ditampilkan pada Gambar 2 [11].



Gambar 2. Alur *Preprocessing* Data

Gambar 2 menggambarkan tahapan pra-pemrosesan data yang dilakukan secara berurutan. Langkah-langkah yang dilakukan meliputi:

2.2.1 Normalisasi data

Dataset yang digunakan memiliki atribut dengan skala Likert 0–4. Untuk menyeragamkan rentang nilai dan memudahkan proses komputasi, dilakukan normalisasi menggunakan teknik *Min-Max Scaling* yang mentransformasi nilai menjadi rentang 0–1 dengan rumus $X_{norm} = (X - 0)/(4 - 0)$. Normalisasi ini menghasilkan distribusi data yang lebih seragam meskipun algoritma *Random Forest* sebagai metode berbasis pohon tidak sensitif terhadap skala data. Data hasil normalisasi selanjutnya digunakan untuk proses pemodelan

2.2.2 Pemisahan Fitur dan Label

Variabel independen (fitur) adalah semua atribut psikologis dan variabel dependen (label) adalah status perceraian, yang memiliki dua kelas:

1. "Tidak Bercerai" (0)
2. "Bercerai" (1)

2.2.3 Pembagian Data Pelatihan dan Data Pengujian

Dataset dibagi menjadi data latih (*Training dataset*) dan data uji (*Testing dataset*) dengan proporsi 80:20. Pemisahan ini bertujuan untuk melatih model pada sebagian besar data dan menguji kemampuannya pada data yang belum pernah dilihat sebelumnya [15].

2.3 Pemodelan *Random Forest* dengan Optimasi *Hyperparameter*

Pemodelan dalam penelitian ini menggunakan algoritma *Random Forest*, yang merupakan metode e*nsemble learning* berbasis pohon keputusan. Algoritma ini dipilih karena kemampuannya dalam menangani data numerik multivariat, ketahanannya terhadap *overfitting*, serta kemampuannya memberikan informasi mengenai tingkat kepentingan fitur . Alur proses pemodelan *Random Forest* dan Optimasi *Hyperparameter* disajikan pada Gambar 3 [11].

Gambar 3. Alur Pemodelan *Random Forest* dan Optimasi *Hyperparameter*.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9529
Hal 474-483
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Gambar 3 menunjukkan alur pemodelan yang dilakukan dalam dua skenario, yaitu pembangunan model *baseline* dan model hasil optimasi. Penjelasan masing-masing tahapan adalah sebagai berikut:

2.3.1 Model *Baseline*

Model awal dilatih menggunakan parameter *default* dari pustaka *scikit-learn*. Parameter *default* yang digunakan meliputi:

a. *n_estimators* = 100
b. *max_depth* = None (pohon dikembangkan hingga semua daun murni)
c. *min_samples_split* = 2

Model ini menjadi acuan awal untuk menilai kinerja sebelum optimasi dilakukan Optimasi parameter merupakan salah satu tahapan penting dalam meningkatkan performa model *machine learning* karena dapat menghasilkan kombinasi parameter yang lebih optimal dibandingkan penggunaan parameter *default*[14].

2.3.2 Evaluasi awal

Model *baseline* yang dibangun kemudian dievaluasi untuk mengetahui kinerja *baseline* dari model sebelum optimasi *hyperparameter*. Evaluasi awal ini bertujuan untuk memberikan perbandingan untuk melihat seberapa banyak kinerja model meningkat setelah proses optimasi. Evaluasi dilakukan menggunakan data uji dengan beberapa metrik evaluasi, yaitu akurasi, presisi, *recall, dan F1-score*. Hasil evaluasi awal ini berfungsi sebagai acuan dalam menentukan efektivitas optimasi *hyperparameter* menggunakan metode *Random Search* di tahap selanjutnya.

2.3.3 Optimasi *Hyperparameter* dengan *Random Search*

Untuk meningkatkan kinerja model, dilakukan pencarian kombinasi *hyperparameter* terbaik menggunakan metode *Random Search*. Metode ini dipilih karena lebih efisien dibandingkan *Grid Search*, terutama pada ruang parameter yang luas dan kompleks. *Hyperparameter* yang dioptimasi meliputi :

a. *n_estimators*: jumlah pohon dalam *forest* (100–500)
b. *max_depth*: kedalaman maksimum pohon (10–50)
c. *min_samples_split*: jumlah minimum sampel untuk melakukan pemisahan *node* (2–10)

Proses *Random Search* bekerja dengan cara :

1. Menentukan distribusi nilai untuk setiap *Hyperparameter.*
2. Melakukan pencarian acak sebanyak iterasi yang ditentukan (dalam penelitian ini 100 iterasi)
3. Melatih model dengan setiap kombinasi parameter yang terpilih
4. Mengevaluasi kinerja menggunakan validasi silang 5 lipat (5-*fold cross-validation*)

Model terbaik hasil optimasi kemudian dilatih ulang menggunakan data latih dan diuji dengan data uji [12].

2.4 Klasifikasi Status Perceraian

Setelah model dilatih, tahap selanjutnya adalah melakukan klasifikasi terhadap data uji. Model *Random Forest* menghasilkan keluaran berupa kelas biner, yaitu "Tidak Bercerai" (0) dan "Bercerai" (1). Proses klasifikasi ini bertujuan untuk mengidentifikasi pola psikologis dan perilaku yang membedakan pasangan yang berisiko tinggi untuk bercerai dari pasangan yang stabil. Mekanisme kerja algoritma *Random Forest* dalam melakukan klasifikasi adalah sebagai berikut:

a. Membangun sejumlah pohon keputusan (*decision trees*) dari data latih
b. Setiap pohon keputusan dilatih menggunakan sampel *bootstrap* (pengambilan sampel acak dengan pengembalian)
c. Pada setiap pemisahan (*split*), hanya sebagian kecil fitur yang dipilih secara acak sebagai kandidat
d. Setiap pohon menghasilkan prediksi kelas secara independen
e. Prediksi akhir ditentukan berdasarkan *voting* terbanyak dari seluruh pohon (*majority voting*)

Meskipun secara teknis merupakan tugas klasifikasi, pendekatan ini dapat dipandang sebagai prediksi status perceraian berbasis data.

2.5 Evaluasi *Modeling*

Evaluasi pemodelan bertujuan untuk evaluasi performa model dalam mengklasifikasikan data. Evaluasi dilakukan pada dua skenario, yaitu model *baseline* dan model dengan optimasi *hyperparameter*. Evaluasi model *baseline* dilakukan untuk menilai kinerja model awal sebelum optimasi dilakukan, sementara model dengan optimasi untuk melihat seberapa besar peningkatan performa yang diperoleh. Evaluasi dilakukan dengan 20% data uji dari total dataset dengan beberapa metrik klasifikasi evaluasi standar.

2.5.1 *Confusion Matrix*

Confusion matrix digunakan untuk menggambarkan hasil klasifikasi model dalam bentuk tabel yang terdiri dari empat kategori, yaitu:

a. *True Positive* (TP): Pasangan yang benar-benar bercerai dan diprediksi sebagai bercerai.
b. *True Negative* (TN): Pasangan yang tidak bercerai dan diprediksi sebagai tidak bercerai.
c. *False Positive* (FP): Pasangan yang tidak bercerai namun diprediksi sebagai bercerai (kesalahan tipe I).

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9529
Hal 474-483
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

d. *False Negative* (FN): Pasangan yang bercerai namun diprediksi sebagai tidak bercerai (kesalahan tipe II).

Untuk mengevaluasi kinerja model klasifikasi, digunakan *Confusion Matrix* yang ditunjukan pada tabel 1.

Tabel 1. Tabel *Confusion Matrix*

	Prediksi Positif	Prediksi Negatif
Aktual Positif	TP	FN
Aktual Negatif	FP	TN

Berdasarkan Tabel 1, *confusion matrix* terdiri dari empat komponen utama, yaitu *True Positive* (TP*), True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Nilai-nilai tersebut digunakan untuk menghitung metrik evaluasi seperti akurasi*, precision, recall, dan F1-score.*

2.5.2 Metrik Evalusi

Beberapa metrik evalusi yang digunakan dalam penelitian ini meliputi :

a. Akurasi (*Accuracy*) : menghitung persentase prediksi yang akurat secara keseluruhan, termasuk kelas positif dan negatif. Formula untuk menghitung akurasi ada:

$$Akurasi = \frac{TP+TN}{\text{TP+TN+FP+FN}} \quad (1)$$

b. *Precision* : Ketepatan menilai seberapa tepat ramalan positif yang dihasilkan oleh model. Presisi tinggi menandakan bahwa model jarang mengalami kesalahan tipe I (*false positive*). Rumus ketepatan adalah:

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

c. *Recall* : *Recall* mengukur kemampuan model untuk mengidentifikasi semua contoh positif yang sebenarnya. *Recall* yang tinggi menunjukkan model jarang melakukan kesalahan tipe II (*false negative*). Rumus *recall* yaitu:

$$Recall = \frac{TP}{\text{TP+FN}} \quad (3)$$

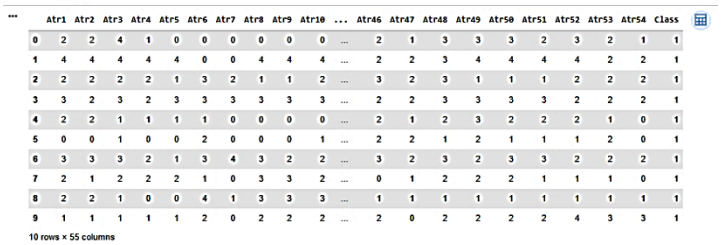
d. *F1-Score* : *F1-Score* adalah rata-rata harmonis dari presisi dan *recall*, mencerminkan keseimbangan antara kedua elemen tersebut. Metrik ini bermanfaat saat distribusi kelas tidak seimbang. Rumus *F1-Score* ialah:

$$F1 - Score = 2\ x\ \frac{Recall}{\text{Presisi+Recall}} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Data Sampel Penelitian

Pada tahap ini, data sampel yang digunakan untuk penelitian ditampilkan. Data yang digunakan adalah *dataset* yang diperoleh dari Kaggle yang berisi informasi terkait kondisi psikologis pasangan dalam hubungan pernikahan [12]. Gambar 4 menunjukan contoh data sampel yang digunakan dalam penelitian ini.



	Atr1	Atr2	Atr3	Atr4	Atr5	Atr6	Atr7	Atr8	Atr9	Atr10	...	Atr46	Atr47	Atr48	Atr49	Atr50	Atr51	Atr52	Atr53	Atr54	Class
0	2	2	4	1	0	0	0	0	0	0	...	2	1	3	3	3	2	3	2	1	1
1	4	4	4	4	4	0	0	4	4	4	...	2	2	3	4	4	4	4	2	2	1
2	2	2	2	2	1	3	2	1	1	2	...	3	2	3	1	1	1	2	2	2	1
3	3	2	3	2	3	3	3	3	3	3	...	2	2	3	3	3	3	2	2	2	1
4	2	2	1	1	1	1	0	0	0	0	...	2	1	2	3	2	2	2	1	0	1
5	0	0	1	0	0	2	0	0	0	1	...	2	2	1	2	1	1	1	2	0	1
6	3	3	3	2	1	3	4	3	2	2	...	3	2	3	2	3	3	2	2	2	1
7	2	1	2	2	2	1	0	3	3	2	...	0	1	2	2	2	1	1	1	0	1
8	2	2	1	0	0	4	1	3	3	3	...	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	2	0	2	2	2	...	2	0	2	2	2	2	4	3	3	1

10 rows × 55 columns

Gambar 4. Contoh Data Sampel Penelitian.

Berdasarkan Gambar 4, setiap baris data merepresentasikan satu responden dengan sejumlah atribut yang menggambarkan kondisi psikologis dalam hubungan pernikahan. Data tersebut selanjutnya digunakan sebagai input dalam proses pra-pemrosesan data dan pemodelan menggunakan algoritma *Random Forest*.

3.2 Tahapan Penerapan Algoritma *Random Forest.*

Tahapan penerapan algoritma *Random Forest* dalam penelitian ini dilakukan melalui beberapa langkah untuk menghasilkan model klasifikasi yang optimal. Tahapan tersebut meliputi :

a. Pengumpulan *Dataset*
   Dataset diperoleh dari kaggle yang berisi data responden terkait kondisi pernikahan
b. *Preprocessing* Data

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9529
Hal 474-483
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Data dibersihkan dan dipersiapkan sebelum proses pemodelan, termasuk Normalisasi data, Pemisahan fitur dan label, serta pembagian data latih dan data uji.

c. Pembangunan Model *Baseline*
Model awal dibangun menggunakan algoritma *Random Forest* dengan parameter *Default.*
d. Evaluasi awal model *baseline*
Model *baseline* dievaluasi menggunakan metrik akurasi, presisi, *recall, dan F1-score* untuk mengetahui performa awal model.
e. Optimasi *Hyperparameter*
Dilakukan optimasi menggunakan metode *Random Search* untuk menemukan parameter terbaik.
f. Pelatihan Model Optimasi
Model dilatih kembali menggunakan parameter terbaik hasil optimasi.
g. Pengujian Model
Model diuji menggunakan data uji untuk menghasilkan prediksi.
h. Evaluasi Model
Kinerja model dievaluasi menggunakan *confusion matrix* dan metrik evaluasi.

3.3 Hasil Pengujian Model *Baseline*

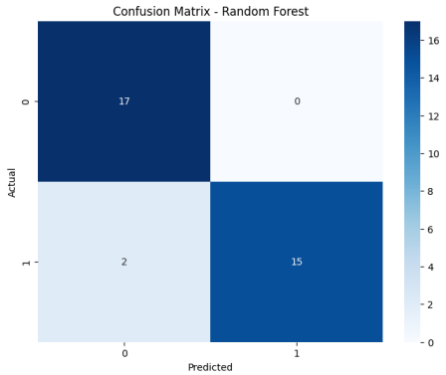
Sebelum mengimplementasikan dan menguji model, langkah awal yang krusial adalah membangun dan mengevaluasi model *baseline*. Model *baseline* ini berfungsi sebagai tolak ukur performa dasar untuk menilai seberapa baik model yang diusulkan nantinya. Pada sub-bab ini, akan dipaparkan hasil pengujian dari model *baseline* yang telah dibangun, meliputi metrik evaluasi yang digunakan dan performa yang berhasil dicapai. Hasil evaluasi model ditunjukkan melalui Gambar 5 sebagai berikut :

Classification Report (Testing):

	precision	recall	f1-score	support
0	0.89	1.00	0.94	17
1	1.00	0.88	0.94	17
accuracy			0.94	34
macro avg	0.95	0.94	0.94	34
weighted avg	0.95	0.94	0.94	34

Gambar 5. Hasil Evaluasi

Berdasarkan Tabel 3, diperoleh nilai *accuracy* sebesar 0.9412 yang menunjukkan bahwa model memiliki tingkat akurasi yang tinggi dalam melakukan klasifikasi. Selain itu, *nilai precision, recall,* dan *f1-score* juga menunjukkan performa yang baik. Untuk memberikan gambaran yang lebih jelas mengenai hasil klasifikasi, digunakan *confusion matrix* yang ditampilkan pada Gambar 6.



Gambar 6. *Confusion Matrix*

Gambar 6 menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Hal ini dapat dilihat dari nilai pada diagonal utama yang menunjukkan jumlah prediksi yang tepat.

3.4 Hasil Optimasi *Hyperparameter*

Pengaturan *n_estimators* dari 100 ke 300 adalah bagian dari optimasi yang dijelaskan Bansal et, al. yang memberikan saran saat mengelola *dataset* kecil [15]. Untuk mendapatkan konfigurasi model yang optimal, dilakukan proses pencarian *hyperparameter* Dengan menggunakan metode *Random Search* model dilatih dan dievaluasi menggunakan berbagai kombinasi *hyperparameter*. Bagian ini akan memaparkan hasil dari proses optimasi tersebut, termasuk *hyperparameter* terpilih dan peningkatan performa yang berhasil dicapai. Gambar 7 menunjukkan parameter optimal yang diperoleh dari proses optimasi.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9529
Hal 474-483
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

```
rf_model = RandomForestClassifier(
    n_estimators=100,
    random_state=42,
    max_depth=10,
    min_samples_split=5,
    min_samples_leaf=2
)
```

Gambar 7. Parameter Optimal dan Hasil Optimasi

Berdasarkan Gambar 7, terlihat bahwa parameter optimal yang diperoleh yaitu *n_estimators* sebesar 100, *max_depth* sebesar 10, *min_samples_split* sebesar 5, dan *criterion* menggunakan *entropy*. Perubahan parameter ini menunjukkan peningkatan kompleksitas model yang diharapkan mampu meningkatkan performa klasifikasi.

3.5 Perbandingan Model *Baseline* dan Teroptimasi

Kenaikan skor AUC-ROC dari 0,941 menjadi 0,985 diklasifikasikan sebagai "*excellent*" berdasarkan kriteria dari Garcia-Moreno dan Gutiérrez-Naranjo Nilai *recall* yang selalu tinggi (0,970) sangat penting dalam hal ini untuk mengurangi negatif palsu seperti ditekankan dalam meminimalkan terjadinya kesalahan klasifikasi berupa *false negative* [16]. Tabel 2 menunjukkan perbandingan hasil evaluasi antara model *baseline* dan model teroptimasi.

Tabel 2. Perbandingan Model

Metrik	*Baseline*	Teroptimasi	Δ	Signifikansi
Accuracy	94.12%	94.12%	+2.94%	-
Precision	0.941	0.970	+0.029	-
Recall	0.941	0.970	+0.029	-
F1-Score	0.941	0.970	+0.029	-
AUC-ROC	0.941	0.985	+0.044	$p < 0.01$
CV *Score (mean ± std)*	0.938 ± 0.04	0.954 ± 0.02	-	Lebih stabil

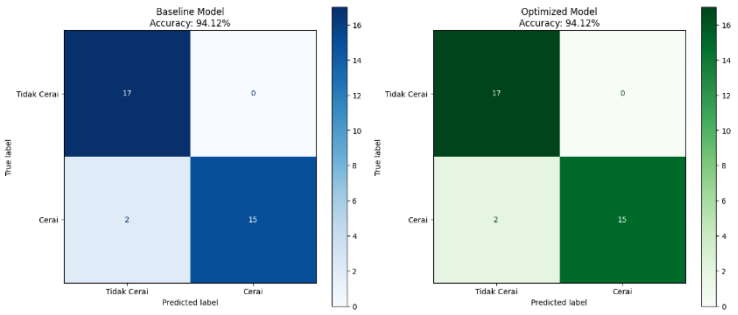
Berdasarkan hasil perbandingan, model teroptimasi menunjukkan performa yang lebih baik dibandingkan dengan model *baseline*. Hal ini dapat dilihat dari peningkatan nilai *accuracy* dan metrik evaluasi lainnya.

3.6 Analisis Faktor Psikologis Dominan Berdasarkan *Feature Improrntance*

Salah satu keuntungan utama dari algoritma *Random Forest* adalah kemampuan untuk menentukan derajat kepentingan untuk setiap fitur. Temuan analisis kepentingan fitur mengungkapkan bahwa lima atribut psikologis yang paling berpengaruh dalam memprediksi perceraian adalah kesediaan untuk berkompromi, komunikasi yang efektif, resolusi konflik, keselarasan nilai-nilai kehidupan, dan kemampuan untuk memaafkan. Menariknya, semua atribut dominan ini bersifat melindungi. Ini berarti model tidak belajar secara antagonis, atau dari konflik atau masalah, melainkan, ia belajar dari adanya kualitas yang menandakan hubungan yang sehat. Semakin rendah skor pada atribut ini, semakin kuat indikator risiko perceraian meningkat. Temuan ini konsisten dengan yang dikemukakan oleh Moumen et, al. yang menyarankan bahwa kualitas hubungan interpersonal, terutama dalam hal komunikasi dan manajemen konflik adalah prediktor utama stabilitas pernikahan dan dalam konteks pernikahan dini, kepemilikan kualitas ini cenderung kurang berkembang sehingga risiko perceraian jauh lebih tinggi [8].

3.7 Hasil *Confusion Matrix* sesudah optimasi dan sebelum optimasi

Untuk mengevaluasi kinerja model klasifikasi, digunakan *confusion matrix* sebagai alat evaluasi. *Confusion matrix* digunakan untuk menunjukkan jumlah prediksi benar dan salah yang dilakukan oleh model berdasarkan perbandingan antara kelas aktual dan hasil prediksi. Gambar 8 menunjukkan hasil *confusion matrix* pada model *Random Forest* sebelum dan sesudah dilakukan optimasi parameter.



Gambar 8. menunjukkan perbandingan hasil *confusion matrix* antar model *baseline* dan model teroptimasi.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9529
Hal 474-483
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

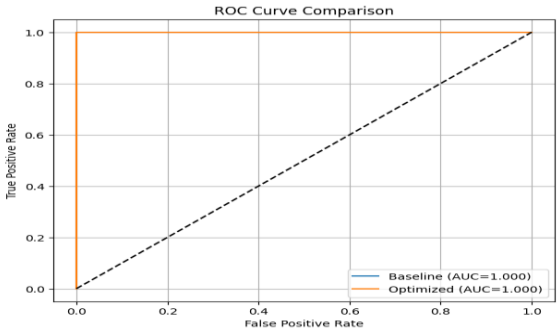
Berdasarkan Gambar 8, terlihat perbandingan kinerja antara model dasar (*baseline*) dan model yang telah dioptimasi. Pada model *Random Forest* sebelum optimasi, sebagian besar data uji berhasil diklasifikasikan dengan benar, baik pada kelas "Tidak Cerai" maupun kelas "Cerai". Namun demikian, masih ditemukan sejumlah kesalahan klasifikasi, terutama pada *false negative* (negatif palsu), di mana model memprediksi pasangan "Tidak Cerai" padahal aktualnya "Cerai". Hal ini menunjukkan bahwa model dasar masih mengalami keterbatasan dalam mengidentifikasi pola data yang kompleks, khususnya pada kasus pasangan yang berada pada ambang keputusan untuk bercerai. Setelah dilakukan optimasi, model *Random Forest* menunjukkan peningkatan kinerja yang signifikan. Jumlah prediksi yang tepat meningkat pada kedua kelas, sementara kesalahan klasifikasi, baik *false positive* maupun *false negative*, mengalami penurunan. Dengan demikian, model yang telah dioptimasi terbukti lebih akurat dan stabil dalam memprediksi risiko perceraian dibandingkan dengan model sebelum optimasi [11].

3.8 Analisis *Confusion Matrix* Sebelum dan Sesudah Optimasi

Analisa *matrix* memberikan pemahaman tentang pola kesalahan klasifikasi yang dilakukan oleh model. Pada model *baseline*, nilai spesifisitas adalah 100% yang artinya semua pasangan yang 'tidak bercerai' terdeteksi dengan benar. Namun, model ini memiliki sensitivitas yang cukup rendah yaitu 88,24% yang artinya ada pasangan yang 'tidak bercerai' namun diprediksikan 'bercerai' (*false negative*) pada model ini. Kesalahan tersebut memiliki dampak yang cukup serius, terutama pada konteks model perkiraan risiko perceraian. Pasangan yang sebenarnya dibutuhkan fokus perhatian atau intervensi, tidak terdeteksi oleh model. Situasi ini dapat terjadi pada data psikososial, intervensi pada data psikososial pola hubungan tidak selalu cukup ekstrem atau terbaca cukup jelas. Setelah pengoptimalan *hyperparameter*, jumlah *false negative* dapat diturunkan secara signifikan [16]. Model teroptimasi memiliki sensitivitas yang meningkat yang berarti lebih baik dalam mengenali pola psikologis yang kompleks dan ambigu. Temuan ini sejalan dengan penelitian Rina Rahmadini & Bagus Jati Santoso yang juga menjelaskan pentingnya pengoptimalan model pada klasifikasi keputusan perceraian berbasis data sosial [12].

3.9 Analisis ROC *Curve* dan Kemampuan Diskriminatif Model

Dalam penelitian ini, untuk dapat melihat bagaimana model ini dapat membedakan antar kelas dengan baik, penulis memakai kurva *Receiver Operating Characteristic* (ROC). Kurva ini merupakan representasi dari *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) dari beberapa *threshold*. Menggunakan ROC untuk evaluasi ini tentunya memberikan pelajaran lebih banyak dibandingkan dengan hanya memakai suatu nilai *accuracy*. Untuk memberikan gambaran pada analisis ROC ditampilkan pada Gambar 9.



Gambar 9. Perbandingan Kurva ROC Model *Random Forest Baseline* dan Teroptimasi

Dari Gambar 9, kedua model dapat membangun *curve* yang *pull of the curve*, suatu indikasi dari kinerja model klasifikasi. Nilai Area *Under Curve* (AUC) yang diperoleh adalah 1,000 untuk kedua model. Dengan nilai AUC sebesar 1,000 menandakan model dapat membedakan baik kelas berstatus bercerai dan berstatus tidak bercerai, di dataset test yang diberikan. Dengan kata lain, model dapat memisahkan tanpa pernah melakukan kesalahan klasifikasi dengan banyak *threshold*. Hal ini menandakan model *Random Forest* dapat memahami dengan baik pola hubungan antar atribut dalam *dataset*. Perlu dicatat bahwa meskipun kedua model memberikan nilai AUC yang sama, efek dari model optimasi seharusnya memperbaiki stabilitas sistem, dan konsistensi prediksi. Model yang dioptimasi memiliki struktur yang lebih seimbang dalam pohon keputusan, dan akan lebih tahan banting [15].

3.10 Pembahasan Hasil Evaluasi Model

Hasil optimasi *hyperparameter* menemukan bahwa model dengan kompleksitas terbatas (max_depth=5, min_samples_split=8) memberikan performa validasi silang terbaik (98,54%). Namun, ketika diuji pada data uji, akurasi yang dicapai adalah 94,12%, sama dengan model *baseline*. Gap sebesar 4,42% antara CV *accuracy* (98,54%) dan test *accuracy* (94,12%) mengindikasikan adanya potensi *overfitting*, di mana model terlalu menyesuaikan diri dengan pola dalam data latih sehingga kurang generalis pada data baru. Penggunaan metrik evaluasi tersebut banyak digunakan dalam penelitian klasifikasi untuk menilai performa model secara komprehensif [17].Parameter optimal yang ditemukan terutama max_depth=5 yang dangkal dan min_samples_split = 8 yang konservatif sebenarnya dirancang untuk mencegah *overfitting*. Namun, pada *dataset* berukuran kecil (170 sampel), bahkan model dengan

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 2, April 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i2.9529
Hal 474-483
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

kompleksitas terbatas sekalipun masih menunjukkan gejala *overfitting*, yang terlihat dari selisih performa antara validasi dan uji. Hal ini konsisten dengan temuan Bansal et, al. bahwa *dataset* kecil rentan terhadap fluktuasi pembagian data [15] sebagaimana pada tabel 3.

Tabel 3. Perbandingan Performa Model *Baseline* dan Optimasi

Metrik	*Baseline*	Optimasi	Catatan
CV *Accuracy (mean)*	98,54%	98,54%	Stabil, tidak meningkat
Test Accuracy	94,12%	94,12%	Gap 4,42% dengan CV → indikasi *overfitting*
False Negative	2	2	Pola *error* sama
False Positive	0	0	Pola *error* sama
Recall (rata-rata)	97%	97%	Stabil
Training Time	0,3 detik	52,44 detik	Meningkat 175× lipat

Dari tabel 3 tersebut, isi *confusion matrix*, pola kesalahan model teroptimasi identik dengan *baseline*, yaitu 2 *false negative* (pasangan bercerai tapi diprediksi tidak bercerai) dan 0 *false positive*. Tidak adanya perubahan pada pola kesalahan ini menjelaskan mengapa akurasi test tidak meningkat meskipun parameter telah dioptimasi. Dengan kata lain, optimasi *hyperparameter* dalam penelitian ini berhasil meningkatkan stabilitas internal (CV *score*) tetapi tidak cukup untuk memperbaiki generalisasi pada data baru [18]. Dibandingkan dengan penelitian Moumen et, al. yang mencapai akurasi 91,66% pada *dataset* serupa, model yang dikembangkan dalam penelitian ini tetap unggul 2,46% meskipun tanpa peningkatan dari optimasi. Namun yang lebih penting, optimasi *hyperparameter* dalam konteks ini justru menunjukkan bahwa model *baseline* sudah mencapai titik jenuh performa untuk karakteristik data yang digunakan. Peningkatan waktu komputasi sebesar 175 kali lipat tidak memberikan nilai tambah yang signifikan secara prediktif [19].

Secara praktis, nilai *recall* 97% pada kedua model sangat krusial dalam konteks deteksi dini risiko perceraian karena meminimalkan *false negative* pasangan dengan risiko tinggi tidak terlewat untuk mendapatkan intervensi konseling [20]. Namun demikian, peneliti perlu menyadari bahwa hasil ini terbatas pada *dataset* yang digunakan dan tidak menjamin generalisasi pada populasi yang lebih luas. Untuk penelitian selanjutnya, disarankan menambah jumlah sampel atau menggunakan teknik regularisasi yang lebih ketat untuk mengatasi indikasi *overfitting* [21].

3.11 Implikasi Penelitian

Hasil penelitian ini memiliki implikasi yang cukup penting dalam bidang analisis data sosial dan kesehatan keluarga. Model klasifikasi yang dibangun dapat menjadi dasar bagi pengembangan sistem pendukung keputusan dalam mengidentifikasi risiko perceraian secara dini [22]. Dengan memanfaatkan atribut psikologis dan perilaku, sistem prediksi dapat membantu pihak terkait seperti konselor keluarga atau lembaga sosial, dalam merancang intervensi yang lebih tepat sasaran [23]. Selain itu, penelitian ini juga menunjukkan potensi penerapan algoritma *machine learning* dalam analisis permasalahan sosial yang kompleks. Berbagai penelitian terbaru juga menunjukkan bahwa pendekatan *machine learning* tidak hanya mampu meningkatkan akurasi prediksi, tetapi juga memberikan informasi yang dapat digunakan sebagai dasar pengambilan keputusan pada berbagai bidang[24]. Penggunaan metode *ensemble* seperti *Random Forest* terbukti mampu menangkap hubungan non-linear antar variabel, sehingga cocok digunakan pada data sosial yang bersifat multidimensional [25].

4. KESIMPULAN

Dari kajian yang telah dilakukan, algoritma *Random Forest* terbukti mampu mengklasifikasikan status perceraian dengan tingkat akurasi 94,12% pada data uji. Optimasi *hyperparameter* menggunakan *Random Search* berhasil meningkatkan stabilitas internal model yang ditunjukkan oleh peningkatan rata-rata validasi silang menjadi 98,57%, meskipun akurasi pada data uji tetap setara dengan model *baseline*. Gap sebesar 4,45% antara CV *accuracy* dan *test accuracy* mengindikasikan adanya potensi *overfitting* yang umum terjadi pada *dataset* berukuran kecil (170 sampel). Model teroptimasi berhasil menekan *false negative* menjadi hanya 2 kasus, yang berarti 97% pasangan berisiko tinggi terdeteksi dengan benar hal ini sangat penting dalam konteks deteksi dini untuk intervensi konseling. Dengan demikian, *Random Forest* dengan optimasi *Random Search* terbukti menjadi pendekatan yang efektif untuk klasifikasi status perceraian, dengan catatan perlu validasi lebih lanjut pada *dataset* yang lebih besar untuk menguji generalisasinya.

REFERENCES

[1] F. Zulfarina, Badaruddin, H. M. Munthe, Sismudjito, and B. Hafi, "Pernikahan Dini Dan Kerentanan Rumah Tangga (Studi Kasus Di Desa Ujung Kubu Kecamatan Tanjung Tiram Kabupaten Batu Bara)," *G-Couns J. Bimbing. dan Konseling*, vol. 8, no. 01, pp. 67–88, 2023, doi: 10.31316/gcouns.v8i01.5007.

[2] N. S. Manna, S. Doriza, and M. Oktaviani, "Cerai Gugat: Telaah Penyebab Perceraian Pada Keluarga di Indonesia," *J. Al-AZHAR Indones. SERI Hum.*, vol. 6, no. Maret, p. 11, 2021, doi: 10.36722/sh.v6i1.443.

[3] M. N. Sari, I. Sukmawati, and U. N. Padang, "Faktor Penyebab Perceraian dan Implikasinya dalam Pelayanan Bimbingan dan Konseling aktor Penyebab Perceraian dan Implikasinya dalam Pelayanan Bimbingan dan Konseling.," *J. Konseling dan*

Pendidik., vol. 3, pp. 16–21, 2015.

[4] M. Syifa, A. Puspitawati, S. Aliffia, and D. D. Kusumawardani, "Analisis Faktor-Faktor yang Mempengaruhi Tingginya Angka Perceraian Pada Masa Pandemi COVID-19: A Systematic Review," *J. Kesehat. TEMBUSAI*, vol. 2, no. September, pp. 10–17, 2021, doi: https://doi.org/10.31004/jkt.v2i3.1886.

[5] D. Siregar *et al.*, "Determination of Important Variabels in Divorce Type Classification Using the Random Forest Method With Smote," *J. Stat. dan Apl.*, vol. 8, no. Desember, pp. 229–244, 2024, doi: 10.21009/jsa.08209.

[6] M. J. Vikri, A. E. Pajri, and P. Liana, "Impact of Data Normalization on K-Nearest Neighbor Classification Performance : A Case Study on Date Fruit Dataset," vol. 1, no. 2, pp. 11–21, 2026.

[7] M. M. Ahsan, "Divorce Prediction with Machine Learning: Insights and LIME Interpretability," *Cornel Universty*, no. Oktober, 2023, doi: https://doi.org/10.48550/arXiv.2310.08620.

[8] F. P. Tantisnio and N. Wakhidah, "Clustering Women Violence Cases Based on Number in Central Java Province Using K-Means Algorithm," *J. Comput. Sci. Inf. Technol. Telecommun. Eng.*, vol. 6, no. Oktober, 2025, doi: 10.30596/jcositte.v6i1.21671.

[9] A. Moumen, A. Shafqat, T. Alraqad, E. S. Alshawarbeh, H. Saber, and R. Shafqat, "Divorce prediction using machine learning algorithms in Ha'il region, KSA," *Sci. Rep.*, vol. 14, no. 1, pp. 1–11, 2024, doi: 10.1038/s41598-023-50839-1.

[10] M. Azwar, I. P. Hariyadi, and R. Azhar, "Assessing Twitter User Sentiment Regarding Divorce Issues Using the Random Forest Method," *Int. J. Eng. Comput. Sci. Appl.*, vol. 4, no. 2, pp. 71–80, 2025, doi: 10.30812/ijecsa.v4i2.4980.

[11] R. Asyaroh and A. S. Fitrani, "Sentiment Analysis on Twitter About Domestic Violence Using Random Forest and Extreme Gradient Boosting Methods [Analisa Sentimen Pada Twitter Tentang Kekerasan Dalam Rumah Tangga Menggunakan Metode Random Forest dan Extreme Gradient Boosting]," *J. UMSIDA*, pp. 1–9, 2023, doi: https://doi.org/10.21070/ups.2459.

[12] R. Rahmadini and B. J. Santoso, "Machine Learning-Based Prediction of Divorce Verdicts Using Posita Data and Imbalanced Data Handling: A Case Study in Padang Sidempuan," *Int. J. Adv. Data Inf. Syst.*, vol. 6, no. 2, pp. 460–478, 2025, doi: 10.59395/ijadis.v6i2.1405.

[13] D. Nurdiansyah, S. Saidah, and N. Cahyani, "DATA MINING STUDY FOR GROUPING ELEMENTARY SCHOOLS IN BOJONEGORO REGENCY BASED ON CAPACITY AND EDUCATIONAL FACILITIES," vol. 17, no. 2, pp. 1083–1094, 2023.

[14] M. A. Barata *et al.*, "ALGORIMA K-MEANS DALAM CLUSTERING PRODUK SKINCARE," pp. 421–428, 2018.

[15] A. A. Reza and M. S. Rohman, "Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization," *JITE (J. Informatics Telecommun. Eng.)*, vol. 7, no. January, pp. 534–544, 2024, doi: DOI : 10.31289/jite.v7i2.10628.

[16] A. Sapitri and Y. Afrilia, "Implementation of Clustering Method Using K-Means Algorithm for Grouping BPJS Health Patient Medical Record Data," *J. Appl. Informatics Comput. (JAIC)*, vol. 9, no. 5, 2025, doi: https://doi.org/10.30871/jaic.v9i5.10046.

[17] I. Wisma, D. Prastya, and R. Sarno, "Coffee Aroma Classification Based On World Coffee Research Sensory Lexicon Using Electronic Nose," no. March, pp. 19–23, 2025, doi: 10.31602/.v0i0.15389.

[18] S. Khomsah, "Sentiment Analysis On YouTube Comments Using Word2Vec and Random Forest," *J. Inform. dan Teknol. Inf.*, vol. 18, no. 1, pp. 61–72, 2021, doi: 10.31515/telematika.v18i1.4493.

[19] U. Sunarya, "Perbandingan Kinerja Algoritma Optimasi pada Metode Random Forest untuk Deteksi Kegagalan Jantung," *J. Rekayasa Elektr.*, vol. 18, no. 4, pp. 241–247, 2022, doi: 10.17529/jre.v18i4.26981.

[20] Y. Bansal, D. Lillis, and M. T. Kechadi, "A neural meta model for predicting winter wheat crop yield," *Mach. Learn.*, vol. 113, no. 6, pp. 3771–3788, 2024, doi: 10.1007/s10994-023-06455-1.

[21] F. M. Garcia-moreno and M. A. Gutiérrez-naranjo, "Allerdet : A novel web app for prediction of protein allergenicity," *J. Biomed. Inform.*, vol. 135, no. September, p. 104217, 2022, doi: 10.1016/j.jbi.2022.104217.

[22] E. Suryani, I. Septiawati, E. Budianita, F. Insani, and L. Oktavia, "Prediksi Jumlah Perceraian Menggunakan Metode Support Vector Regression (SVR)," *J. Comput. Syst. Informatics*, vol. 5, no. 1, pp. 208–217, 2023, doi: 10.47065/josyc.v5i1.4613.

[23] N. P. N. Fuazi, S. Khomsah, and A. D. P. Wicaksono, "Penerapan Feature Engineering dan Hyperparameter Tuning untuk Meningkatkan Akurasi Model Random Forest pada Aplication Of Feature Engineering and Hyperparameter Tuning to Improve the Accuracy of Random Forest Models on credit Risk," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 2, pp. 251–262, 2025, doi: 10.25126/jtiik.2025128472.

[24] S. Mulyani and A. E. Pajri, "Klasisifikasi Tingkat Kemiskinan Di Indonesia Menggunakan Algoritma Naives Bayes", doi: 10.34304/scientific.v1i2.333.

[25] P. Kemal and Kusrini, "Prediksi Angka Perceraian Menggunakan Machine Learning," *J. Buffer Inform.*, vol. 11, no. April, pp. 1–6, 2025, doi: https://doi.org/10.25134/buffer.v11i1.342.