Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

KLASIFIKASI BINER TINDAK PIDANA MENGGUNAKAN *DECISION TREE* DAN *RANDOM FOREST* DENGAN *SMOTE* PADA DATA SISTEM INFORMASI PENELUSURAN PERKARA PENGADILAN NEGERI KEBUMEN

Rahmat Hidayat[1], Weri Sirait [2], M. Syafrizal Zain [3], Firdaus[4], David Nugroho[5]
1,2,3,4,5) Rekayasa Perangkat Lunak, Informatika dan Bisnis, Politeknik Manufaktur Negeri Bangka Belitung

Article Info

Article history:

Received: 02 Juli 2026
Revised: 16 Juli 2026
Accepted: 21 Juli 2026

ABSTRACT

Abstrak

Kriminalitas merupakan salah satu permasalahan yang memengaruhi keamanan dan ketertiban masyarakat. Pemanfaatan data historis perkara melalui pendekatan machine learning dapat mendukung penyusunan strategi pencegahan kejahatan yang lebih terarah. Penelitian ini bertujuan mengembangkan model klasifikasi tindak pidana menggunakan data Sistem Informasi Penelusuran Perkara (SIPP) Pengadilan Negeri Kebumen periode 2021–2025. Permasalahan utama yang dihadapi adalah ketidakseimbangan distribusi kelas dan rendahnya performa klasifikasi pada data dengan banyak kategori tindak pidana. Untuk mengatasinya, sembilan rumpun tindak pidana direkonstruksi menjadi dua kategori, yaitu Kejahatan Kekerasan dan Kejahatan Non-Kekerasan. Penyeimbangan data dilakukan menggunakan Synthetic Minority Over-sampling Technique (SMOTE), sedangkan proses klasifikasi menggunakan algoritma *Decision Tree* dan Random Forest dengan skema stratified split 80:20. Hasil pengujian menunjukkan bahwa *Decision Tree* memberikan performa terbaik dengan akurasi 73,81%, presisi 79,08%, recall 73,81%, dan F1-score 75,45%, sedangkan Random Forest memperoleh akurasi 65,48%. Analisis feature importance menunjukkan bahwa tingkat pendidikan, pekerjaan, dan lokasi kejadian merupakan faktor yang paling berkontribusi dalam membedakan kategori tindak pidana. Hasil penelitian menunjukkan bahwa pendekatan yang diusulkan berpotensi mendukung pengambilan keputusan dalam penyusunan strategi pencegahan kejahatan berbasis data.

Kata Kunci: Klasifikasi Biner, Kriminalitas, *Decision Tree*, Random Forest, SMOTE.

Abstract

Crime remains a significant challenge to maintaining public security and social order. The utilization of historical case records through machine learning approaches can support the development of more effective crime prevention strategies. This study aims to develop a crime classification model using data obtained from the Case Tracking Information System (SIPP) of the Kebumen District Court for the period 2021–2025. The main challenges addressed are class imbalance and the low classification performance associated with multiple crime categories. To overcome these issues, nine crime categories were reconstructed into two classes: Violent Crime and Non-Violent Crime. The Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the training data, while Decision Tree and Random Forest algorithms were employed using an 80:20 stratified split scheme. The results indicate that Decision Tree achieved the best performance, with an accuracy of 73.81%, precision of 79.08%, recall of 73.81%, and an F1-score of 75.45%, whereas Random Forest obtained an accuracy of 65.48%. Feature importance analysis identified education level, occupation, and crime location as the most influential variables. These findings demonstrate the potential of the proposed approach to support data-driven decision-making in crime prevention strategies.

Keywords: Binary Classification, Criminology, Decision Tree, Random Forest, SMOTE.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Corresponding Author:
E-mail : rahmathidayat@polman-babel.ac.id

1. PENDAHULUAN

Keamanan dan ketertiban masyarakat merupakan faktor penting yang mendukung stabilitas sosial serta keberlangsungan pembangunan di suatu daerah. Perubahan kondisi sosial, ekonomi, dan demografi masyarakat dapat memengaruhi pola terjadinya tindak pidana, baik dari segi jenis kejahatan maupun karakteristik pelakunya [1]. Dalam praktiknya, upaya penanganan kriminalitas masih banyak dilakukan secara reaktif setelah suatu peristiwa terjadi. Pendekatan tersebut sering menghadapi berbagai keterbatasan, terutama terkait ketersediaan sumber daya dan efektivitas pengawasan di lapangan [2]. Oleh karena itu, diperlukan pendekatan yang mampu mendukung upaya pencegahan melalui pemanfaatan data historis sebagai dasar pengambilan keputusan yang lebih terukur.

Perkembangan *machine learning* membuka peluang untuk mengidentifikasi pola-pola tersembunyi pada data kriminalitas yang sulit diamati melalui analisis konvensional [3]. Berbagai algoritma klasifikasi telah digunakan untuk mempelajari hubungan antara karakteristik pelaku, lokasi kejadian, dan jenis tindak pidana. Di antara berbagai metode yang tersedia, *Decision Tree* dan *Random Forest* menjadi dua algoritma yang banyak digunakan karena mampu menangani data kategorikal, mudah diinterpretasikan, serta memiliki kemampuan klasifikasi yang relatif baik pada data *tabular*.

Meskipun demikian, penerapan *machine learning* pada data kriminalitas lokal masih menghadapi sejumlah tantangan yang menjadi celah penelitian (*research gap*) saat ini. Beberapa penelitian terdahulu umumnya memaksakan penggunaan klasifikasi multikelas dengan jumlah kategori tindak pidana yang banyak [4]. Pendekatan tersebut sering kali menghasilkan performa pemodelan yang rendah karena batas keputusan antar-kelas menjadi sangat kompleks, terutama ketika karakteristik demografis pelaku dan lokasi kejadian memiliki kemiripan yang tinggi. Selain itu, distribusi data kriminalitas di tingkat daerah cenderung tidak seimbang (*imbalanced dataset*) karena beberapa jenis tindak pidana minoritas hanya memiliki sedikit sampel. Upaya penanganan ketidakseimbangan kelas menggunakan metode *oversampling* seperti *Synthetic Minority*

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Over-sampling Technique (SMOTE) memang kerap digunakan untuk mendongkrak performa model linear. Namun, dampak integrasi SMOTE pada algoritma berbasis pohon—khususnya komparasi antara model pohon tunggal (*Decision Tree*) dan model ensemble (*Random Forest*) pada data kriminalitas *tabular* lokal yang berukuran terbatas—masih jarang dieksplorasi secara kritis. Algoritma *ensemble* yang biasanya unggul sering kali rentan mengalami overfitting ketika mempelajari pola data sintetis hasil interpolasi SMOTE pada sampel berskala kecil, yang berpotensi menurunkan akurasi generalisasinya pada data uji riil [5].

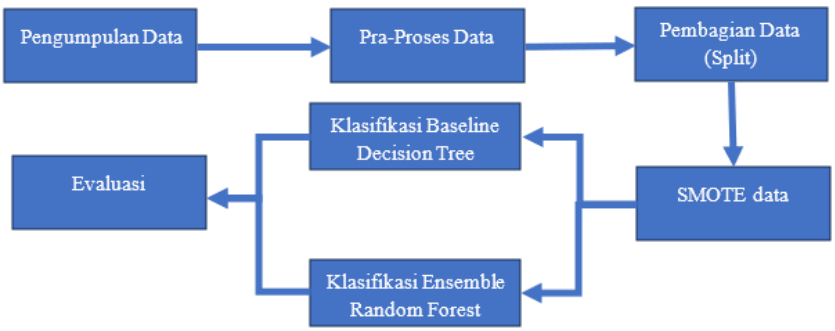
Berdasarkan gap tersebut, penelitian ini menerapkan pendekatan klasifikasi biner dengan mengelompokkan sembilan rumpun tindak pidana asal ke dalam dua kategori utama, yaitu Kejahatan Kekerasan dan Kejahatan Non-Kekerasan. Pendekatan ini dipilih untuk menyederhanakan ruang klasifikasi sekaligus meningkatkan kemampuan model dalam membedakan karakteristik inti kedua kelompok kejahatan tersebut. Untuk mengatasi permasalahan ketidakseimbangan data, teknik SMOTE diterapkan pada data latih sehingga distribusi kelas menjadi lebih proporsional selama proses pembelajaran model. Eksperimen ini dirancang secara khusus untuk mengevaluasi secara kritis sensitivitas arsitektur Decision Tree tunggal dan Random Forest ensemble terhadap intervensi data sintetis tersebut.

Penelitian ini menggunakan data perkara pidana yang diperoleh dari Sistem Informasi Penelusuran Perkara (SIPP) Pengadilan Negeri Kebumen periode 2021–2025. Selanjutnya dilakukan evaluasi komparatif terhadap algoritma Decision Tree dan Random Forest guna mengetahui metode yang memberikan performa terbaik dalam mengklasifikasikan tindak pidana ke dalam kategori kekerasan dan non-kekerasan. Selain membandingkan kinerja kedua algoritma, penelitian ini juga menganalisis tingkat kontribusi setiap variabel prediktor untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap karakteristik tindak pidana. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan model prediksi kriminalitas berbasis data serta menjadi informasi pendukung bagi penyusunan strategi pencegahan dan pengawasan kriminalitas di tingkat daerah.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

2. METODE PENELITIAN

Tahapan penelitian disusun secara berurutan untuk memastikan proses pengembangan model klasifikasi berlangsung secara sistematis dan menghasilkan evaluasi yang objektif. Proses diawali dengan pengumpulan data perkara dari Sistem Informasi Penelusuran Perkara (SIPP) Pengadilan Negeri Kebumen, kemudian dilanjutkan dengan tahap prapemrosesan data untuk memastikan kualitas dan konsistensi data yang akan digunakan. Dataset yang telah dipersiapkan selanjutnya dipisahkan menjadi data latih dan data uji menggunakan metode *stratified split* guna mempertahankan proporsi kelas pada kedua subset data [6]. Ketidakseimbangan distribusi kelas pada data latih ditangani menggunakan teknik *Synthetic Minority Over-sampling Technique (SMOTE)* sebelum proses pelatihan model dilakukan [7]. Data hasil penyeimbangan kemudian digunakan untuk membangun dua model klasifikasi, yaitu *Decision Tree* sebagai model pembanding (*baseline*) dan *Random Forest* sebagai model *ensemble*. Tahap akhir penelitian berupa evaluasi performa kedua model menggunakan data uji yang tidak terlibat selama proses pelatihan sehingga kemampuan generalisasi model dapat diukur secara lebih objektif. Alur keseluruhan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

2.1 Pengumpulan Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Sistem Informasi Penelusuran Perkara (SIPP) Pengadilan Negeri Kebumen. Data yang digunakan berupa catatan perkara tindak pidana yang telah berkekuatan hukum dan tercatat dalam rentang waktu tahun 2021 hingga 2025. Pemanfaatan data SIPP dipilih karena menyediakan informasi perkara yang terdokumentasi secara sistematis sehingga dapat digunakan

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

untuk mengidentifikasi pola kriminalitas berdasarkan karakteristik pelaku dan lokasi kejadian.

Jumlah data yang berhasil dihimpun sebanyak 417 catatan perkara yang selanjutnya digunakan sebagai unit analisis dalam penelitian. Setiap catatan perkara memuat sejumlah atribut yang berpotensi memengaruhi karakteristik tindak pidana. Variabel prediktor yang digunakan terdiri atas lokasi kejadian (kecamatan), jenis kelamin, usia, pekerjaan, dan tingkat pendidikan terakhir pelaku. Sementara itu, variabel target ditentukan berdasarkan kelompok tindak pidana yang tercantum dalam data perkara. Variabel-variabel tersebut dipilih karena mewakili aspek demografis dan spasial yang dinilai relevan dalam proses pembentukan model klasifikasi kriminalitas.

2.2 Pra-Proses Data

Tahap prapemrosesan dilakukan untuk memastikan kualitas data sebelum digunakan dalam proses pembentukan model klasifikasi. Pada tahap ini dilakukan pemeriksaan terhadap kelengkapan data serta identifikasi kemungkinan adanya data duplikat yang dapat memengaruhi hasil analisis [7]. Hasil pemeriksaan menunjukkan bahwa seluruh atribut pada 417 catatan perkara memiliki tingkat kelengkapan sebesar 100%, sehingga tidak ditemukan nilai hilang (*missing value*) pada dataset. Selain itu, proses identifikasi data duplikat juga menunjukkan bahwa tidak terdapat baris data yang memiliki kesamaan atribut secara keseluruhan (*exact duplicate*). Oleh karena itu, seluruh data dapat digunakan pada tahap analisis tanpa memerlukan proses imputasi maupun penghapusan data[8].

Setelah kualitas data dipastikan, dilakukan transformasi pada variabel target. Kategori tindak pidana yang semula terdiri atas sembilan rumpun perkara disederhanakan menjadi dua kelompok utama, yaitu Kejahatan Kekerasan dan Kejahatan Non-Kekerasan. Transformasi ini dilakukan untuk mengurangi kompleksitas klasifikasi sekaligus mengatasi keterbatasan jumlah sampel pada beberapa kategori tindak pidana yang memiliki frekuensi kemunculan relatif rendah. Dengan jumlah kelas yang lebih sedikit, model diharapkan mampu membentuk batas keputusan yang lebih jelas dan menghasilkan performa klasifikasi yang lebih stabil. Pemetaan kategori tindak pidana ke dalam dua kelas baru ditunjukkan pada Tabel 1.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Tabel 1. Pengkategorian Tindak Pidana

Rumpun Tindak Pidana Asal	Domain Kelas Baru	Label Indeks	Jumlah Sampel
Tindak Pidana Kekerasan Tindak Pidana Pembunuhan	Kejahatan Kekerasan	1	93 data
Tindak Pidana Ekonomi Tindak Pidana Narkotika Tindak Pidana Khusus Tindak Pidana Perlindungan Anak Tindak Pidana Pencurian Tindak Pidana Kesusilaan Tindak Pidana Umum	Kejahatan Non-Kekerasan	0	324 data

Hasil transformasi menunjukkan bahwa dataset terdiri atas 93 data pada kelas Kejahatan Kekerasan dan 324 data pada kelas Kejahatan Non-Kekerasan. Dataset yang telah melalui tahap prapemrosesan selanjutnya digunakan pada proses pembagian data dan penyeimbangan kelas menggunakan metode *Synthetic Minority Over-sampling Technique (SMOTE).*

2.3 Pembagian Data (*Split Data*)

Setelah melalui tahap prapemrosesan, dataset dibagi menjadi dua subset, yaitu data latih (*training set*) dan data uji (*testing set*). Pembagian dilakukan menggunakan rasio 80:20, di mana 80% data digunakan untuk proses pelatihan model dan 20% sisanya digunakan untuk menguji kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya [9].

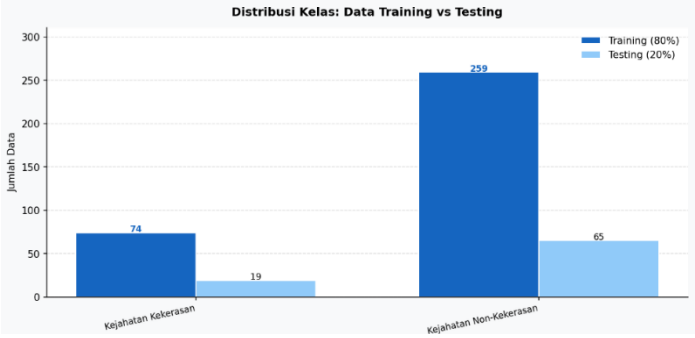
Proses pembagian data menerapkan metode *Stratified Sampling* untuk mempertahankan proporsi distribusi kelas pada data latih dan data uji agar tetap merepresentasikan distribusi kelas pada dataset asli. Pendekatan ini dipilih untuk mengurangi potensi bias yang dapat muncul akibat ketidakseimbangan jumlah sampel antar kelas [10].

Hasil pembagian menghasilkan 333 data pada subset pelatihan dan 84 data pada subset pengujian. Pada data latih terdapat 74 sampel Kejahatan Kekerasan dan 259 sampel Kejahatan Non-Kekerasan. Sementara itu, data uji terdiri atas 19 sampel

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Kejahatan Kekerasan dan 65 sampel Kejahatan Non-Kekerasan. Dengan demikian, proporsi distribusi kelas pada kedua subset tetap terjaga dan sesuai dengan distribusi data awal.

Distribusi jumlah sampel pada data latih dan data uji untuk masing-masing kelas ditunjukkan pada Gambar 2.



Gambar 2. Distribusi Frekuensi Sampel Kelas Biner pada Data Latik dan Data Uji

2.4 Penyeimbangan Kelas (*SMOTE*)

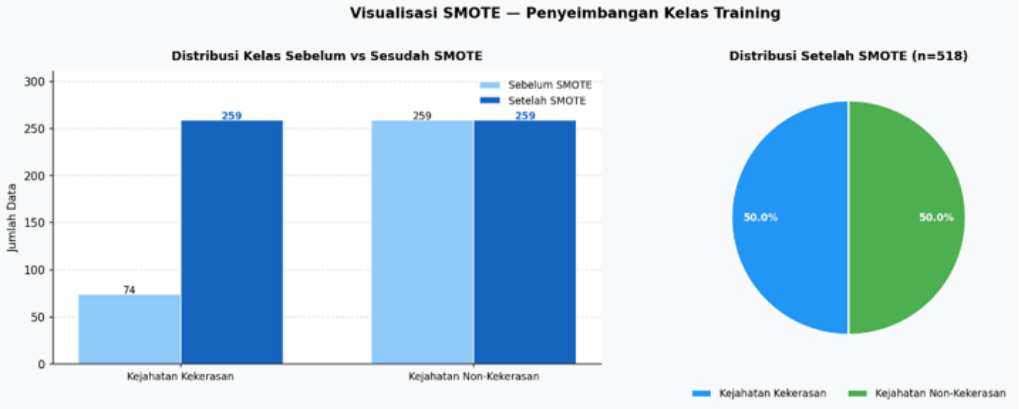
Distribusi kelas pada data latih menunjukkan adanya ketidakseimbangan yang cukup besar. Dari total 333 data latih, kelas Kejahatan Kekerasan hanya berjumlah 74 sampel, sedangkan kelas Kejahatan Non-Kekerasan berjumlah 259 sampel. Kondisi ini berpotensi menyebabkan model lebih mudah mempelajari karakteristik kelas mayoritas sehingga performa klasifikasi terhadap kelas minoritas menjadi kurang optimal.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan metode *Synthetic Minority Over-sampling Technique (SMOTE)* pada data latih. *SMOTE* merupakan teknik *oversampling* yang menghasilkan sampel sintetis pada kelas minoritas berdasarkan kedekatan antar data dalam ruang fitur [11]. Proses pembentukan sampel baru dilakukan melalui interpolasi antara suatu sampel minoritas dengan salah satu tetangga terdekatnya (*k-nearest neighbors*), sehingga distribusi kelas dapat ditingkatkan tanpa melakukan duplikasi data secara langsung.

Penerapan *SMOTE* dilakukan setelah proses pembagian data dan hanya pada subset data latih. Pendekatan ini bertujuan untuk mencegah terjadinya data *leakage* yang dapat menyebabkan estimasi performa model menjadi terlalu optimistis. Pada penelitian ini, jumlah sampel kelas minoritas ditingkatkan hingga setara dengan jumlah sampel kelas mayoritas. Setelah proses penyeimbangan dilakukan, kedua kelas masing-masing memiliki 259 sampel sehingga total data latih bertambah menjadi 518 sampel dengan distribusi kelas yang seimbang.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Perubahan distribusi kelas sebelum dan sesudah penerapan *SMOTE* ditunjukkan pada Gambar 3. Hasil tersebut menunjukkan bahwa ketimpangan kelas yang sebelumnya cukup besar berhasil dikurangi sehingga proses pelatihan model dapat berlangsung pada distribusi data yang lebih proporsional.



Gambar 3. Visualisasi Penyeimbangan Kelas Data Latih Menggunakan SMOTE

2.5 Algoritma *Decision Tree*

Decision Tree digunakan sebagai model pembanding (*baseline model*) dalam penelitian ini. Algoritma ini membentuk struktur pohon keputusan yang terdiri atas simpul akar (*root node*), simpul percabangan (*internal node*), dan simpul daun (*leaf node*). Setiap simpul percabangan merepresentasikan proses pengujian terhadap suatu atribut prediktor, sedangkan simpul daun menunjukkan hasil klasifikasi akhir. Melalui mekanisme tersebut, *Decision Tree* mampu menghasilkan aturan keputusan yang mudah dipahami dan diinterpretasikan[12].

Proses pembentukan pohon dilakukan secara rekursif dengan memilih atribut yang menghasilkan pemisahan data terbaik pada setiap simpul. Dalam penelitian ini, kualitas pemisahan diukur menggunakan kriteria *Gini Impurity*, yaitu ukuran ketidakmurnian suatu simpul berdasarkan distribusi kelas yang terdapat di dalamnya. *Nilai Gini Impurity* dihitung menggunakan Persamaan (1) [13].

$$G = 1 - \sum_{i=1}^{c} (P_i)^2$$

dengan (G) menyatakan nilai Gini Impurity, (c) merupakan jumlah kelas target, dan (P_i) adalah proporsi sampel yang termasuk ke dalam kelas ke-(i). Nilai Gini Impurity yang semakin kecil menunjukkan bahwa data pada suatu simpul semakin homogen, sehingga lebih baik digunakan sebagai dasar pemisahan data [6].

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Decision Tree dipilih sebagai model pembanding karena memiliki struktur yang sederhana, mudah diinterpretasikan, serta mampu menangani data kategorikal maupun numerik tanpa memerlukan asumsi distribusi tertentu. Karakteristik tersebut menjadikan *Decision Tree* sesuai untuk mengidentifikasi pola hubungan antara faktor demografis dan spasial dengan kategori tindak pidana yang dipelajari.

2.6 Algoritma *Random Forest*

Random Forest merupakan metode *ensemble learning* yang mengombinasikan sejumlah pohon keputusan (*Decision Tree*) untuk menghasilkan model klasifikasi yang lebih stabil [14]. Berbeda dengan *Decision Tree* yang membangun satu struktur pohon tunggal, *Random Forest* membentuk banyak pohon keputusan secara independen berdasarkan variasi sampel data dan atribut yang dipilih secara acak. Pendekatan ini bertujuan untuk mengurangi variabilitas model serta meningkatkan kemampuan generalisasi terhadap data yang belum pernah diamati sebelumnya [15].

Proses pembentukan *Random Forest* diawali dengan penerapan teknik *Bootstrap Aggregating (bagging),* yaitu pengambilan sampel data latih secara acak dengan pengembalian (*sampling with replacement*) untuk membangun setiap pohon keputusan [16]. Selanjutnya, pada setiap proses percabangan hanya sebagian atribut yang dipilih secara acak untuk dipertimbangkan sebagai kandidat pemisah (*split candidate*). Kombinasi kedua mekanisme tersebut menghasilkan keragaman antar pohon sehingga prediksi yang dihasilkan menjadi lebih *robust* dibandingkan penggunaan satu pohon keputusan tunggal [17], [18].

Pada tahap klasifikasi, setiap pohon keputusan menghasilkan prediksi kelas secara independen. Hasil prediksi dari seluruh pohon kemudian digabungkan menggunakan mekanisme *majority voting*, di mana kelas yang memperoleh jumlah suara terbanyak ditetapkan sebagai hasil prediksi akhir. Secara matematis, proses tersebut dinyatakan pada Persamaan [2], [19], [20].

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_N(x)\}$$

dengan (G) menyatakan nilai *Gini Impurity*, (c) merupakan jumlah kelas target, dan (P_i) adalah proporsi sampel yang termasuk ke dalam kelas ke-(i). Nilai *Gini Impurity* yang semakin kecil menunjukkan bahwa data pada suatu simpul semakin homogen, sehingga lebih baik digunakan sebagai dasar pemisahan data [6].

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Decision Tree dipilih sebagai model pembanding karena memiliki struktur yang sederhana, mudah diinterpretasikan, serta mampu menangani data kategorikal maupun numerik tanpa memerlukan asumsi distribusi tertentu. Karakteristik tersebut menjadikan *Decision Tree* sesuai untuk mengidentifikasi pola hubungan antara faktor demografis dan spasial dengan kategori tindak pidana yang dipelajari.

3. HASIL DAN PEMBAHASAN

3.1 Pengujian Validasi Silang (*Cross-Validation*)

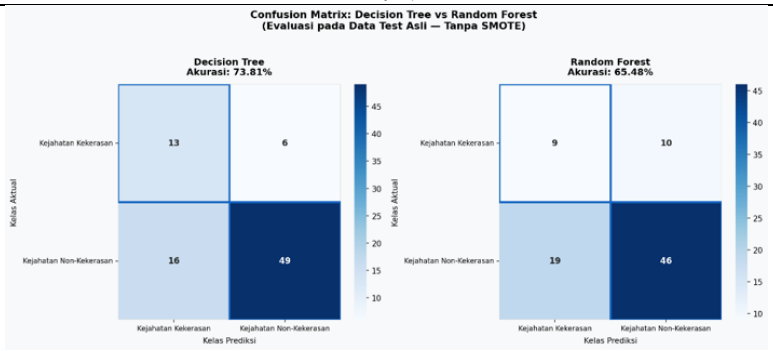
Sebelum dilakukan pengujian pada data uji, performa awal model dievaluasi menggunakan metode *Stratified 5-Fold Cross-Validation* yang terintegrasi dengan SMOTE pada setiap lipatan (*fold*). Pendekatan ini digunakan untuk memperoleh estimasi performa model yang lebih stabil sekaligus meminimalkan pengaruh variasi pembagian data latih.

Hasil pengujian menunjukkan bahwa *Random Forest* memperoleh rata-rata akurasi sebesar 73,29% dengan standar deviasi 3,04%, sedangkan *Decision Tree* menghasilkan rata-rata akurasi sebesar 71,48% dengan standar deviasi 3,83%. Nilai standar deviasi yang relatif rendah pada kedua model menunjukkan bahwa performa klasifikasi cenderung konsisten pada berbagai skenario pembagian data. Hasil ini mengindikasikan bahwa proses penyeimbangan kelas menggunakan SMOTE mampu membantu menjaga stabilitas model selama proses pelatihan.

3.2 Evaluasi Komparatif pada Data Uji

Evaluasi akhir dilakukan menggunakan 84 data uji yang tidak terlibat dalam proses pelatihan maupun penyeimbangan data. Pengujian ini bertujuan untuk mengukur kemampuan generalisasi model dalam mengklasifikasikan data perkara yang belum pernah diamati sebelumnya. Hasil klasifikasi kedua model ditunjukkan melalui confusion matrix pada Gambar 4.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno



Gambar 4. Perbandingan *Confusion Matrix Decision Tree* dan *Random Forest*

Berdasarkan Gambar 4, *Decision Tree* menghasilkan performa yang lebih baik dibandingkan *Random Forest* pada data uji. Model *Decision Tree* mampu mengklasifikasikan 13 kasus Kejahatan Kekerasan dan 49 kasus Kejahatan Non-Kekerasan dengan benar, sehingga memperoleh akurasi sebesar 73,81%. Sementara itu, *Random Forest* berhasil mengklasifikasikan 9 kasus Kejahatan Kekerasan dan 46 kasus Kejahatan Non-Kekerasan dengan benar, dengan akurasi sebesar 65,48%.

Perbedaan hasil tersebut menunjukkan bahwa model yang memiliki performa validasi silang lebih tinggi belum tentu menghasilkan performa terbaik pada data uji. Pada penelitian ini, karakteristik dataset yang relatif kecil serta dominasi atribut kategorikal diduga memengaruhi efektivitas mekanisme pembentukan banyak pohon pada *Random Forest*. Sebaliknya, *Decision Tree* mampu membentuk aturan klasifikasi yang lebih sesuai terhadap pola data yang tersedia sehingga menghasilkan performa pengujian yang lebih baik.

Rangkuman hasil evaluasi menggunakan metrik akurasi, *presisi, recall, dan F1-score* disajikan pada Tabel 2.

Tabel 2. Tabel Perbandingan

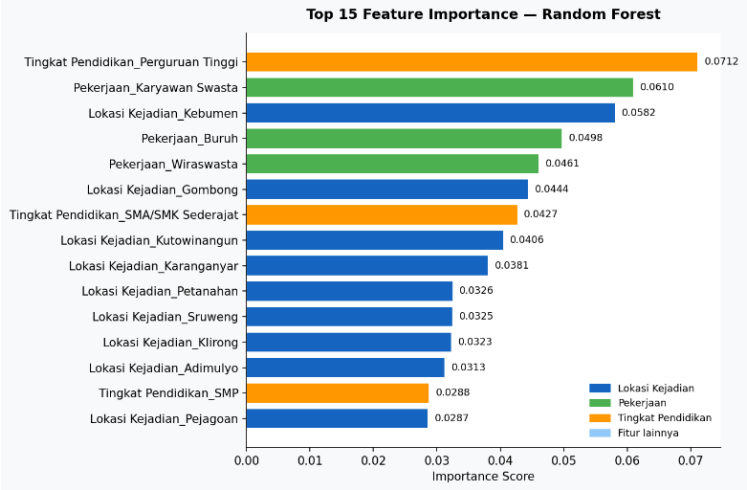
Algoritma Klasifikasi	Akurasi	Presisi	Recall	F1-Score
Decision Tree	73,81%	79,08%	73,81%	75,45%
Random Forest	65,48%	70,83%	65,48%	67,50%

Berdasarkan seluruh metrik evaluasi, *Decision Tree* menunjukkan performa yang lebih baik dibandingkan *Random Forest*. Selisih terbesar terlihat pada nilai F1-score, yang menunjukkan bahwa *Decision Tree* memiliki keseimbangan yang lebih baik antara presisi dan recall dalam mengklasifikasikan kedua kategori tindak pidana.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

3.3 Analisis Kentingan Fitur (*Feature Importance*)

Analisis *feature importance* dilakukan untuk mengidentifikasi variabel yang memberikan kontribusi terbesar dalam proses klasifikasi. Hasil peringkat 15 fitur teratas yang diperoleh dari model *Random Forest* ditunjukkan pada berikut:



Gambar 5. Top 15 Feature Importance

Berdasarkan gambar diatas, variabel Tingkat Pendidikan Perguruan Tinggi memiliki nilai kepentingan tertinggi dengan skor 0,0712, diikuti oleh Pekerjaan Karyawan Swasta sebesar 0,0610 dan Lokasi Kejadian Kecamatan Kebumen sebesar 0,0582. Selain itu, beberapa variabel lokasi lain seperti Kecamatan Gombong, Kutowinangun, dan Karanganyar juga termasuk dalam kelompok fitur dengan kontribusi yang relatif tinggi.

Temuan tersebut menunjukkan bahwa faktor pendidikan, pekerjaan, dan lokasi kejadian merupakan variabel yang paling berpengaruh dalam membedakan kategori Kejahatan Kekerasan dan Kejahatan Non-Kekerasan pada dataset yang digunakan. Dominannya variabel lokasi mengindikasikan bahwa aspek spasial memiliki kontribusi penting dalam proses klasifikasi, sementara variabel pendidikan dan pekerjaan menunjukkan bahwa karakteristik demografis pelaku juga berperan dalam membentuk pola tindak pidana yang tercatat dalam data perkara.

Perlu diperhatikan bahwa nilai *feature importance* menggambarkan kontribusi variabel terhadap proses prediksi model dan tidak dapat diinterpretasikan sebagai hubungan sebab-akibat. Oleh karena itu, hasil ini lebih tepat digunakan sebagai indikator faktor yang relevan dalam proses klasifikasi serta sebagai dasar untuk pengembangan

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

penelitian lanjutan yang melibatkan analisis sosial dan kriminalitas yang lebih mendalam.

4. SIMPULAN

Penelitian ini berhasil menerapkan pendekatan klasifikasi biner untuk mengelompokkan tindak pidana ke dalam kategori Kejahatan Kekerasan dan Kejahatan Non-Kekerasan menggunakan data Sistem Informasi Penelusuran Perkara (SIPP) Pengadilan Negeri Kebumen. Penerapan SMOTE pada data latih mampu mengatasi ketidakseimbangan kelas sehingga proses pelatihan model dapat dilakukan pada distribusi data yang lebih seimbang.

Berdasarkan hasil pengujian, algoritma *Decision Tree* menunjukkan performa yang lebih baik dibandingkan *Random Forest* dengan akurasi sebesar 73,81%, presisi 79,08%, recall 73,81%, dan F1-score 75,45%. Selain itu, analisis *feature importance* menunjukkan bahwa variabel pendidikan, pekerjaan, dan lokasi kejadian merupakan faktor yang paling berkontribusi dalam membedakan kategori tindak pidana. Temuan ini menunjukkan bahwa pendekatan *machine learning* dapat dimanfaatkan sebagai pendukung analisis kriminalitas berbasis data untuk membantu perumusan strategi pencegahan kejahatan di tingkat daerah.

REFERENCES

[1] J. S. Humaniora, D. Pendidikan, I. K. Kurniawan, D. Budimansyah, and R. Sartika, "Tinjauan Literatur terhadap Faktor Sosial Penyebab Kriminalitas," *Jurnal Sosial Mumaniora dan Pendidikan*, vol. 6, no. 1, pp. 1–6, Mar. 2026, doi: 10.51903/65dsd686.

[2] B. D. Kurniawan, R. Heriansyah, and Z. R. Mair, "ANALISIS PREDIKSI TERHADAP PENINGKATAN TINDAK PIDANA DENGAN METODE NAIVE BAYES BERDASARKAN LAPORAN KRIMINALITAS," 2025.

[3] R. D. Yuniarsyih R.A, R. A. Muhadi, A. Fitrianto, and P. Silvianti, "Analisis Regresi Logistik Biner dan Random Forest untuk Prediksi Faktor-Faktor Stunting di Pulau Jawa," *Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 13, no. 2, pp. 147–156, Jul. 2025, doi: 10.37905/euler.v13i2.31680.

[4] M. R. Kurniawan, O. N. Pratiwi, and F. Hamami, "KLASIFIKASI SOAL MENGGUNAKAN MULTI-LABEL PROBLEM TRANSFORMATION DENGAN METODE RANDOM FOREST DAN K-NEAREST NEIGHBOR," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 1, pp. 367–380, Jan. 2025, doi: 10.29100/jipi.v10i1.5910.

[5] A. K. Putri and H. Suparwito, "Uji Algoritma Stacking Ensemble Classifier pada Kemampuan Adaptasi Mahasiswa Baru dalam Pembelajaran Online," Jun. 2023.

[6] F. Puspitasari Budiana and H. Shelviani, "KOMBINASI ALGORITMA DECISION TREE DAN NAIVE BAYES UNTUK MENINGKATKAN AKURASI DAN KECEPATAN WAKTU DETEKSI MALWARE," 2025.

[7] A. Fauziah and J. Hernadi, "Klasifikasi Data Tak Seimbang Menggunakan Algoritma Random Forest dengan SMOTE dan SMOTE-ENN (Studi Kasus pada Data Stunting)," *Jurnal Riset Sistem dan Teknologi Informasi (RESTIA*, vol. 3, no. 2, pp. 112–121, 2025, [Online]. Available: https://bit.ly/DataSintetis.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2, Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

[8] J. J. Hidayat, M. R. Saputra, A. R. Sigand, A. L. N. Fadilah, M. D. I. Amin, and A. R. Ramadhan, "Evaluasi Kinerja Algoritma Ensemble Learning pada Klasifikasi Penyakit Diabetes Berbasis Boosting Method," *Jurnal Surya Informatika*, vol. 16, no. 1, pp. 71–80, May 2026, doi: https://doi.org/10.48144/suryainformatika.v16i1.2424.

[9] A. F. Anjani, D. Anggraeni, and I. M. Tirta, "Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 163–172, Sep. 2023, doi: 10.25077/teknosi.v9i2.2023.163-172.

[10] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 677–690, Jul. 2022, doi: 10.30812/matrik.v21i3.1726.

[11] T. Muthia, N. Nurrahma, and N. J. Putri, "Penerapan teknik oversampling pada dataset indikator kesehatan diabetes menggunakan SMOTE untuk klasifikasi penderita diabetes," 2025.

[12] H. Mulyo and N. A. Maori, "PENINGKATAN AKURASI PREDIKSI PEMILIHAN PROGRAM STUDI CALON MAHASISWA BARU MELALUI OPTIMASI ALGORITMA DECISION TREE DENGAN TEKNIK PRUNING DAN ENSEMBLE ENHACING PREDICTION ACCURACY OF NEW STUDENT PROGRAM SELECTION THROUGH DECISION TREE ALGORITHM OPTIMIZATION WITH PRUNING TECHNIQUE AND ENSEMBLE," vol. 15, no. 1, pp. 15–25, 2024, doi: 10.34001/jdpt.

[13] M. Khadafi and A. Yudhistira, "PENERAPAN METODE DECISION TREE DAN NAÏVE BAYES PADA KASUS KRIMINALITAS DI LAMPUNG," *DINAMIK*, vol. 31, no. 1, pp. 256–266, 2026.

[14] W. Amanda and A. Voutama, "Klasifikasi Pendapatan Menggunakan Algoritma Random Forest: Studi Kasus Dataset Adult Income," *Jurnal Ilmiah Informatika Global*, vol. 2, no. 16, pp. 79–84, Jul. 2025.

[15] Suci Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng," *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 4, no. 3, pp. 121–127, Dec. 2022, doi: 10.35580/variansiunm31.

[16] A. Khairunnisa, "PERBANDINGAN MODEL RANDOM FOREST DAN XGBOOST UNTUK PREDIKSI KEJAHATAN KESUSILAAN DI PROVINSI JAWA BARAT," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 2, p. 202, Sep. 2023, doi: 10.26798/jiko.v7i2.799.

[17] F. Santoso and E. Hartati, "PENGGUNAAN ALGORITMA RANDOM FOREST DALAM KLASIFIKASI BUAH SEGAR DAN BUSUK," *Jurnal Algoritme*, vol. 3, no. 1, pp. 133–140, Oct. 2022, doi: 10.35957/algoritme.xxxx.

[18] C. A. Bahri *et al.*, "ANALISIS FAKTOR RISIKO PEMICU SERANGAN JANTUNG DI INDONESIA, MENGGUNAKAN METODE KLASIFIKASI (DECISION TREE, NAIVE BAYES, DAN RANDOM FOREST)," 2025.

[19] I. A. Rahmi, F. M. Afendi, and A. Kurnia, "Metode AdaBoost dan Random Forest untuk Prediksi Peserta JKN-KIS yang Menunggak," *Jambura Journal of Mathematics*, vol. 5, no. 1, pp. 83–94, Jan. 2023, doi: 10.34312/jjom.v5i1.15869.

[20] A. Suaib and I. I. Tritosmoro, "Perbandingan Performa Metode Local Binary Pattern dan Random Forest dalam Identifikasi COVID-19 pada Citra X-ray Paru-paru," Mar. 2023.