

KLASIFIKASI DAN PREDIKSI PENANGANAN ANAK TERLANTAR BERBASIS CART DAN RANDOM FOREST

Dila Siti Nurfadhilah¹⁾, Gina Indah Permata Nastia²⁾, Raden Sukmana³⁾

^{1,3)} Teknik Informatika Universitas Langlangbuana

²⁾ Pendidikan Kesejahteraan Keluarga Universitas Pendidikan Indonesia

email : dilastnrf@unla.ac.id¹⁾, gina.nastia@upi.edu²⁾, radensukmana2111@gmail.com³⁾

Abstraksi

Permasalahan anak terlantar memerlukan penanganan berbasis data yang sistematis. Penelitian ini membangun model klasifikasi status penanganan anak terlantar menggunakan Decision Tree (CART) dan Random Forest yang diimplementasikan dalam dashboard interaktif. Data sintetis dihasilkan dari distribusi agregat UPTD PPSGRA Dinas Sosial Jawa Barat periode 2020-2025 menggunakan Distribution-based Synthetic Data Generator. Dataset terdiri dari 5 variabel input (lembaga, jenis kelamin, tingkat pendidikan, kondisi ekonomi, usia) untuk memprediksi status penanganan (Berhasil, Proses, Gagal). Metode supervised learning dengan pembagian 80% training dan 20% testing menghasilkan Random Forest unggul dengan accuracy 74.6%, lebih baik dari Decision Tree (71.5%). Analisis feature importance mengidentifikasi usia sebagai faktor paling dominan (48.53%), diikuti oleh lembaga penanganan (21.46%) dan tingkat pendidikan (17.86%), sementara kondisi ekonomi (6.70%) dan jenis kelamin (5.46%) menunjukkan kontribusi yang lebih rendah. Temuan bahwa usia mendominasi feature importance mengindikasikan pentingnya pendekatan age-specific dalam program intervensi anak terlantar. Dashboard berbasis Streamlit menyediakan visualisasi perbandingan model, confusion matrix, feature importance, dan prediksi real-time dengan interpretasi otomatis. Penelitian ini berkontribusi pada penerapan machine learning untuk kesejahteraan sosial dengan pendekatan privacy-preserving, menyediakan decision support system untuk Dinas Sosial dalam pengambilan keputusan berbasis data.

Kata Kunci :

Anak Terlantar, Machine Learning, CART, Random Forest, Dashboard Interaktif

Abstract

The problem of neglected children requires systematic data-based management. This study developed a classification model for the status of neglected children's treatment using Decision Tree (CART) and Random Forest, implemented in an interactive dashboard. Synthetic data was generated from the aggregate distribution of UPTD PPSGRA of the West Java Social Service for the 2020-2025 period using a Distribution-based Synthetic Data Generator. The dataset consisted of 5 input variables (institution, gender, education level, economic condition, age) to predict the treatment status (Successful, In Progress, Failed). The supervised learning method with a division of 80% training and 20% testing resulted in a superior Random Forest with an accuracy of 74.6%, better than Decision Tree (71.5%). Feature importance analysis identified age as the most dominant factor (48.53%), followed by the treatment institution (21.46%) and education level (17.86%), while economic condition (6.70%) and gender (5.46%) showed lower contributions. The finding that age dominates feature importance indicates the importance of an age-specific approach in intervention programs for neglected children. The Streamlit-based dashboard provides visualizations of model comparisons, confusion matrices, feature importance, and real-time predictions with automated interpretation. This research contributes to the application of machine learning to social welfare with a privacy-preserving approach, providing a decision support system for the Social Services Agency in data-driven decision-making.

Keywords :

Abandoned Children, Machine Learning, CART, Random Forest, Interactive Dashboard

Pendahuluan

Masalah anak terlantar merupakan salah satu permasalahan sosial yang masih menjadi perhatian serius di Indonesia. Anak terlantar didefinisikan sebagai anak yang tidak mendapatkan hak dasar seperti pengasuhan, perlindungan, pendidikan, dan kesehatan secara layak, serta tidak memiliki dukungan keluarga yang memadai. Berdasarkan data Badan Pusat Statistik (BPS), pada tahun 2023 tercatat sekitar 4,5 juta anak di Indonesia masuk

dalam kategori anak yang memerlukan perlindungan khusus, termasuk anak terlantar. Di Provinsi Jawa Barat sendiri, berdasarkan laporan dari Dinas Sosial tahun 2023, terdapat lebih dari 65.000 anak terlantar yang tersebar di berbagai kabupaten/kota, dan angka ini cenderung meningkat setiap tahunnya.

Pemerintah daerah melalui Dinas Sosial telah melakukan berbagai upaya dalam menangani permasalahan anak terlantar. Pola penanganan yang

diterapkan meliputi proses identifikasi dan pendataan, asesmen kondisi sosial dan ekonomi anak, penyaluran bantuan sosial, penempatan anak di lembaga perlindungan (panti asuhan, rumah singgah), hingga reunifikasi dengan keluarga atau komunitas yang suportif. Selain itu, dilakukan juga monitoring dan evaluasi terhadap anak dan keluarganya [1][2][3]. Namun, implementasi dari pola-pola ini kerap menemui hambatan, terutama terkait keterbatasan data yang valid, tumpang tindih program, dan kurangnya sistem informasi yang dapat digunakan secara integratif antar lembaga terkait [4]. Akibatnya, penanganan sering kali bersifat reaktif dan kurang berbasis pada analisis yang mendalam dan sistematis.

Dalam pelaksanaannya, penanganan anak terlantar membutuhkan proses asesmen yang menyeluruh untuk menentukan strategi intervensi yang tepat. Namun, proses asesmen dan evaluasi keberhasilan penanganan selama ini masih banyak dilakukan secara manual, mengandalkan pengalaman dan pertimbangan subjektif pekerja sosial. Pendekatan ini memiliki beberapa keterbatasan, antara lain potensi subjektivitas dalam pengambilan keputusan, waktu asesmen yang relatif lama, serta kesulitan dalam mengidentifikasi pola-pola tertentu dari data historis. Kondisi tersebut dapat berdampak pada kurang optimalnya penentuan prioritas penanganan dan alokasi sumber daya.

Perkembangan teknologi informasi, khususnya di bidang data science, membuka peluang untuk mengatasi permasalahan tersebut. Machine learning sebagai bagian dari kecerdasan buatan telah banyak digunakan untuk menganalisis data dalam jumlah besar dan menghasilkan prediksi yang akurat di berbagai bidang. Dalam konteks penanganan anak terlantar, machine learning berpotensi membantu mengidentifikasi pola-pola penting dari data historis serta mendukung prediksi status penanganan secara lebih objektif dan berbasis data.

Algoritma Decision Tree atau Classification and Regression Tree (CART) merupakan salah satu metode klasifikasi yang banyak digunakan karena hasilnya mudah dipahami dan diinterpretasikan. Algoritma ini membentuk struktur pohon keputusan berdasarkan atribut yang paling berpengaruh terhadap variabel target. Meskipun demikian, Decision Tree memiliki kelemahan berupa kecenderungan mengalami overfitting, terutama pada data yang kompleks. Oleh karena itu, penelitian ini juga menggunakan algoritma Random Forest sebagai model pembanding. Random Forest merupakan metode ensemble yang menggabungkan banyak Decision Tree untuk menghasilkan prediksi yang lebih stabil dan akurat. Berbagai penelitian menunjukkan bahwa Random Forest umumnya memiliki performa yang lebih baik dibandingkan Decision Tree tunggal pada dataset yang kompleks.

Sejumlah penelitian sebelumnya telah menerapkan machine learning dalam bidang kesejahteraan sosial. Namun, penelitian yang secara khusus membahas prediksi status penanganan anak terlantar di Indonesia, khususnya di Provinsi Jawa Barat, masih terbatas. Selain itu, sebagian besar penelitian hanya berfokus pada aspek pemodelan tanpa memperhatikan bagaimana hasil analisis dapat digunakan secara langsung oleh praktisi di lapangan. Kesenjangan inilah yang menjadi dasar dilakukannya penelitian ini.

Penelitian ini tidak hanya berfokus pada pengembangan dan perbandingan model machine learning, tetapi juga pada implementasi hasil penelitian dalam bentuk dashboard interaktif berbasis web. Dashboard dikembangkan menggunakan framework Streamlit untuk memudahkan visualisasi data, evaluasi model, serta prediksi status penanganan kasus baru secara interaktif. Dengan pendekatan ini, hasil penelitian diharapkan dapat lebih mudah dipahami dan dimanfaatkan oleh pekerja sosial tanpa memerlukan keahlian teknis di bidang machine learning.

Dataset yang digunakan dalam penelitian ini merupakan data sintesis yang dihasilkan berdasarkan distribusi data agregat riil UPTD PPSGRA periode 2020–2025. Penggunaan data sintesis dilakukan untuk menjaga kerahasiaan dan privasi data anak, sekaligus memastikan bahwa karakteristik dan pola data tetap merepresentasikan kondisi riil. Dataset terdiri dari lima variabel prediktor, yaitu lembaga, jenis kelamin, tingkat pendidikan, kondisi ekonomi, dan usia, serta satu variabel target berupa status penanganan yang terdiri dari kategori berhasil, proses, dan gagal.

Berdasarkan latar belakang tersebut, dibutuhkan pendekatan berbasis data (data-driven decision making) yang mampu mengintegrasikan teknologi informasi dengan analisis data yang sistematis [5]. Penelitian ini bertujuan untuk mengembangkan model klasifikasi status penanganan anak terlantar menggunakan algoritma Decision Tree (CART) dan Random Forest, membandingkan performa kedua model, mengidentifikasi faktor-faktor yang paling berpengaruh terhadap keberhasilan penanganan, serta mengimplementasikan hasil penelitian dalam bentuk dashboard interaktif.

Tinjauan Pustaka

Anak Terlantar

Anak terlantar merupakan anak yang mengalami kegagalan pemenuhan kebutuhan dasar akibat ketidakmampuan atau kelalaian sistem pengasuhan di sekitarnya. Anak terlantar adalah anak yang tidak memperoleh pemenuhan kebutuhan fisik, psikologis, sosial, dan pendidikan secara layak sehingga memerlukan perlindungan serta pelayanan dari lembaga kesejahteraan sosial anak [6]. Definisi ini menegaskan bahwa keterlantaran anak tidak semata-

mata disebabkan oleh kondisi ekonomi, tetapi juga oleh lemahnya fungsi keluarga dan lingkungan dalam memberikan pengasuhan yang memadai. Definisi anak terlantar juga dapat dipahami dari perspektif perlindungan anak secara lebih luas. Anak dalam kondisi rentan adalah anak yang mengalami deprivasi hak dasar, termasuk perlindungan, pengasuhan, pendidikan, dan kesejahteraan sosial [7]. Anak yang tidak memperoleh hak-hak tersebut berpotensi mengalami keterlantaran dan membutuhkan intervensi sosial yang terstruktur. Perspektif ini menempatkan keterlantaran anak sebagai isu perlindungan hak anak yang memerlukan penanganan lintas sektor.

Machine Learning

Machine Learning (ML) merupakan bagian dari kecerdasan buatan yang berfokus pada pengembangan model/algorithm agar komputer dapat mempelajari pola dari data dan meningkatkan kinerja tanpa harus diprogram ulang secara terus-menerus untuk setiap kondisi. Dalam konteks penerapan, ML memungkinkan sistem membangun hubungan antara variabel input dan output berdasarkan data historis sehingga mampu menghasilkan prediksi atau keputusan pada data baru. ML dipandang sebagai pendekatan penting karena dapat digunakan pada beragam tugas, termasuk pengenalan pola dan pengambilan keputusan otomatis pada kasus klasifikasi [8]. Penerapan ML umumnya mengikuti alur umum berupa pengumpulan data, pra-pemrosesan, pelatihan model, dan evaluasi performa sebelum digunakan dalam sistem nyata. Pada penelitian terapan, ML juga sering dipakai karena dapat meningkatkan efisiensi dan ketepatan dibanding pendekatan konvensional, terutama ketika data cukup banyak dan bervariasi. Penerapan ML dapat membantu menghasilkan kontrol/prediksi yang lebih tepat dan efisien pada suatu sistem berbasis data, yang memperlihatkan relevansi ML untuk mendukung proses pengambilan keputusan berbasis data [9].

Classification and Regression Tree (CART)

CART (Classification and Regression Tree) merupakan salah satu algoritma Decision Tree yang membangun pohon keputusan menggunakan pembagian biner pada setiap simpul. Artinya, setiap node hanya memiliki dua cabang keputusan berdasarkan kondisi tertentu dari suatu atribut. Pendekatan biner ini bertujuan untuk menyederhanakan struktur pohon dan meningkatkan konsistensi dalam proses pemisahan data. Dalam praktiknya, algoritma CART digunakan untuk tugas klasifikasi maupun regresi, tergantung pada jenis variabel target yang dianalisis. Algoritma pohon keputusan seperti CART atau C4.5 digunakan untuk memilih atribut terbaik yang mampu memisahkan data ke dalam kelas yang berbeda secara optimal [10]. Konsep CART ini menjadi dasar dalam berbagai penelitian klasifikasi karena mampu

menangani data numerik maupun kategorikal dengan baik.

Random Forest

Algoritma ini dirancang untuk mengatasi kelemahan decision tree tunggal, khususnya overfitting. Random Forest mampu menghasilkan performa klasifikasi yang lebih stabil karena keputusan akhir tidak bergantung pada satu pohon, melainkan pada kumpulan pohon keputusan [11]. Hasil prediksi dari seluruh pohon kemudian digabungkan untuk menentukan kelas akhir menggunakan mekanisme voting mayoritas. Proses agregasi hasil prediksi ini membuat Random Forest lebih tahan terhadap kesalahan prediksi individu dan meningkatkan akurasi model secara keseluruhan [12].

Metrik Evaluasi

Metrik evaluasi digunakan untuk mengukur performa model klasifikasi secara kuantitatif berdasarkan nilai yang diperoleh dari confusion matrix. Metrik yang umum digunakan meliputi accuracy, precision, recall, dan F1-score. Penggunaan beberapa metrik evaluasi secara bersamaan diperlukan agar penilaian kinerja model klasifikasi menjadi lebih komprehensif dan tidak bias terhadap distribusi kelas tertentu [13].

Accuracy menunjukkan tingkat ketepatan prediksi model secara keseluruhan dan dirumuskan sebagai berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision mengukur ketepatan model dalam memprediksi kelas positif:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall mengukur kemampuan model dalam mendeteksi seluruh data positif yang sebenarnya:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score merupakan rata-rata harmonis antara precision dan recall:

$$F_1 \text{ - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Metrik F1-score sangat berguna ketika dataset memiliki distribusi kelas yang tidak seimbang karena mampu menyeimbangkan nilai precision dan recall.

Penelitian Terkait

Beberapa penelitian terdahulu telah mengkaji permasalahan sosial seperti kesejahteraan keluarga dan anak terlantar dengan berbagai pendekatan.

Menurut [14], keterlantaran anak dipengaruhi oleh faktor ekonomi, keluarga, dan lingkungan sosial, sehingga diperlukan pemetaan kondisi sosial secara sistematis sebagai dasar penanganan. Penelitian tersebut menekankan pentingnya pengelompokan data sosial untuk mendukung pengambilan keputusan yang tepat, namun masih menggunakan pendekatan analisis deskriptif.

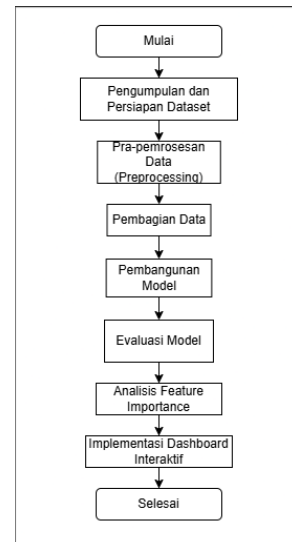
Pendekatan klasifikasi mulai banyak diterapkan dalam permasalahan sosial. Menurut [10], algoritma Decision Tree C4.5 mampu mengklasifikasikan tingkat kesejahteraan keluarga secara efektif dan menghasilkan aturan keputusan yang mudah dipahami. Hal ini menunjukkan bahwa metode pohon keputusan memiliki potensi besar untuk diterapkan pada data sosial. Namun, penelitian tersebut belum menggabungkan Decision Tree dengan metode ensemble untuk meningkatkan stabilitas model.

Dalam bidang machine learning, beberapa penelitian telah menerapkan algoritma klasifikasi pada berbagai domain. Menurut [15], Decision Tree efektif digunakan untuk klasifikasi data karena bersifat interpretable dan fleksibel terhadap berbagai tipe atribut. Sementara itu, menurut [11], Random Forest menunjukkan performa yang lebih baik dibandingkan algoritma klasifikasi tunggal seperti Naive Bayes, terutama dalam hal akurasi dan stabilitas model. Penelitian tersebut menegaskan bahwa Random Forest mampu mengurangi overfitting yang sering terjadi pada Decision Tree tunggal.

Penelitian lain oleh [12], menunjukkan bahwa Random Forest efektif digunakan pada kasus klasifikasi multi-class dan dataset dengan kompleksitas tinggi. Meskipun demikian, sebagian besar penelitian machine learning tersebut masih berfokus pada bidang kesehatan, ekonomi, atau lingkungan, dan belum banyak yang mengintegrasikan pendekatan klasifikasi machine learning dengan permasalahan sosial spesifik seperti penanganan anak terlantar.

Metode Penelitian

Rangkaian tahapan penelitian ini dirancang untuk memastikan setiap proses dilakukan secara terstruktur, mulai dari pengolahan data hingga implementasi sistem, sehingga hasil yang diperoleh dapat dimanfaatkan secara optimal. Penelitian ini dilakukan melalui beberapa tahapan sistematis sebagai berikut:



Gambar 1. Tahapan Penelitian

Pengumpulan dan Persiapan Dataset

Penelitian diawali dengan pengumpulan data agregat dari UPTD PPSGRA sebagai dasar penelitian. Data agregat ini kemudian dilakukan validasi untuk memastikan kelengkapan, konsistensi, dan kesesuaiannya dengan kebutuhan penelitian. Selanjutnya, data agregat digunakan sebagai acuan dalam proses generasi data sintetis menggunakan pendekatan *distribution-based generator*. Data sintetis yang dihasilkan kemudian divalidasi kembali dengan cara membandingkan distribusinya terhadap data agregat untuk memastikan bahwa pola statistik dan karakteristik data tetap terjaga.

Pra-pemrosesan Data (Preprocessing)

Tahap preprocessing meliputi:

1. Data Cleaning
 - a) Penanganan missing values
 - b) Penanganan outliers pada variabel numerik (usia)
 - c) Penanganan inkonsistensi data
2. Feature Engineering
Tidak dilakukan pembuatan fitur baru, menggunakan fitur yang tersedia
3. Encoding
 - a) Variabel kategorikal dikonversi menjadi numerik menggunakan Label Encoding
 - b) Variabel: Lembaga, Jenis_Kelamin, Tingkat, Kondisi_Ekonomi, Status
 - c) Label Encoding dipilih karena kompatibel dengan Decision Tree dan Random Forest
4. Feature Selection
 - a) Variabel tahun tidak digunakan sebagai fitur untuk menghindari temporal bias
 - b) Fitur yang digunakan: Lembaga, Jenis_Kelamin, Tingkat, Kondisi_Ekonomi, Usia

Pembagian Data

Dataset dibagi menjadi dua subset:

1. Data Training (80%): Digunakan untuk melatih model

2. Data Testing (20%): Digunakan untuk mengevaluasi performa model

Pembagian dilakukan secara random dengan `random_state=42` untuk reproducibility.

Pembangunan Model

Model 1: Decision Tree (CART)

```
dt_model =  
DecisionTreeClassifier(random_state=42)  
dt_model.fit(X_train, y_train)
```

Parameter yang digunakan:

1. Menggunakan parameter default scikit-learn
2. `random_state=42` untuk reproducibility
3. Kriteria split: Gini Index (default)

Model 2: Random Forest

```
rf_model =  
RandomForestClassifier(n_estimators=100,  
random_state=42)  
rf_model.fit(X_train, y_train)
```

Parameter yang digunakan:

1. `n_estimators=100` (jumlah trees)
2. `random_state=42` untuk reproducibility
3. Parameter lain menggunakan default scikit-learn

Evaluasi Model

Evaluasi dilakukan menggunakan data testing dengan metrik berikut:

1. Accuracy: Mengukur seberapa banyak prediksi yang benar secara keseluruhan.
2. Precision: Menilai tingkat ketepatan prediksi untuk setiap kelas (weighted average).
3. Recall: Mengukur kemampuan model dalam mengenali setiap kelas dengan benar (weighted average).
4. F1-Score: Nilai gabungan antara precision dan recall (weighted average).
5. Confusion Matrix: Menampilkan rincian kesalahan prediksi untuk setiap kelas secara visual.

Analisis Feature Importance

Analisis feature importance dilakukan pada model Random Forest untuk mengetahui fitur mana yang paling berpengaruh dalam menentukan status penanganan anak. Analisis ini menunjukkan kontribusi relatif setiap fitur, seperti usia, kondisi ekonomi, dan tingkat pendidikan, terhadap hasil prediksi model. Hasil feature importance digunakan sebagai bahan interpretasi model sehingga proses pengambilan keputusan tidak hanya bergantung pada hasil prediksi, tetapi juga pada pemahaman faktor-faktor yang memengaruhinya.

Implementasi Dashboard Interaktif

Dashboard dikembangkan menggunakan Streamlit dengan fitur-fitur:

1. Multi-Page Navigation
 - a) Home & Overview
 - b) Eksplorasi Data
 - c) Model & Evaluasi

- d) Prediksi
2. Fitur Home & Overview
 - a) Metrics cards untuk KPI utama
 - b) Quick statistics dataset
 - c) Visualisasi distribusi data
 3. Fitur Eksplorasi Data
 - a) Preview dataset
 - b) Statistik deskriptif
 - c) Visualisasi distribusi variabel kategorikal
 - d) Analisis bivariat dengan heatmap
 - e) Box plot distribusi usia per status
 4. Fitur Model & Evaluasi
 - a) Perbandingan performa CART vs Random Forest (tabel dan visualisasi)
 - b) Confusion matrix untuk kedua model
 - c) Feature importance analysis
 - d) Classification report detail
 5. Fitur Prediksi
 - a) Form input data kasus baru
 - b) Prediksi dari kedua model
 - c) Visualisasi probabilitas prediksi
 - d) Interpretasi otomatis dengan confidence level
 - e) Rekomendasi action berdasarkan tingkat keyakinan

Hasil dan Pembahasan

Pengumpulan dan Persiapan Dataset

Dataset yang digunakan dalam penelitian ini merupakan data sintesis yang dihasilkan berdasarkan distribusi data agregat riil dari UPTD PPSGRA (Unit Pelaksana Teknis Daerah Pelayanan dan Perlindungan Sosial Griya Remaja Asuh) Dinas Sosial Provinsi Jawa Barat. Dataset ini mencerminkan pola penanganan anak terlantar yang ditangani oleh UPTD PPSGRA dan Satpel GRA di wilayah Jawa Barat. Data agregat yang digunakan sebagai dasar pembentukan data sintesis mencakup periode tahun 2020 hingga 2025.

Proses pembentukan dataset dilakukan menggunakan pendekatan data generator berbasis Python, dengan memanfaatkan data agregat per lembaga dan per tahun yang meliputi jumlah anak, jenis kelamin, tingkat pendidikan, status penanganan, kondisi ekonomi, dan rentang usia. Data individu dihasilkan melalui proses pengacakan terkontrol (weighted random) sehingga distribusi setiap variabel tetap mengikuti proporsi data agregat aslinya. Variabel usia disesuaikan dengan karakteristik masing-masing lembaga, yaitu usia 0–6 tahun untuk lembaga balita dan 7–18 tahun untuk lembaga lainnya.

Pra-pemrosesan Data (Preprocessing)

Pembangunan Model

Model 1: Decision Tree (CART)

Model Decision Tree berhasil dibangun menggunakan algoritma CART dengan konfigurasi parameter default dari scikit-learn. Proses training dilakukan pada 80% data (data training) dengan menggunakan 5 fitur input: Lembaga, Jenis

Kelamin, Tingkat Pendidikan, Kondisi Ekonomi, dan Usia.

Karakteristik model Decision Tree yang dihasilkan:

1. Menggunakan Gini Index sebagai kriteria split
2. Parameter:
DecisionTreeClassifier(random_state=42)
3. Tree dapat tumbuh hingga kedalaman yang cukup dalam
4. Setiap path dari root ke leaf merepresentasikan aturan Keputusan

Model 2: Random Forest

Model Random Forest dibangun dengan konfigurasi: RandomForestClassifier(n_estimators=100, random_state=42).

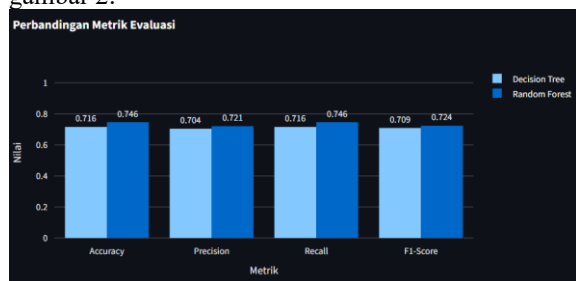
Model ini membentuk ensemble dari 100 decision trees dengan bootstrap sampling untuk meningkatkan stabilitas dan akurasi prediksi.

Proses training Random Forest:

1. Membuat 100 decision trees dengan bootstrap sampling
2. Setiap tree dilatih pada subset data yang berbeda
3. Setiap split mempertimbangkan subset fitur secara acak
4. Prediksi final menggunakan majority voting dari 100 trees

Evaluasi Model

Evaluasi dilakukan pada data testing (20% dari total data) menggunakan 4 metrik utama (accuracy, precision, recall dan f1-score) yang disajikan pada gambar 2.



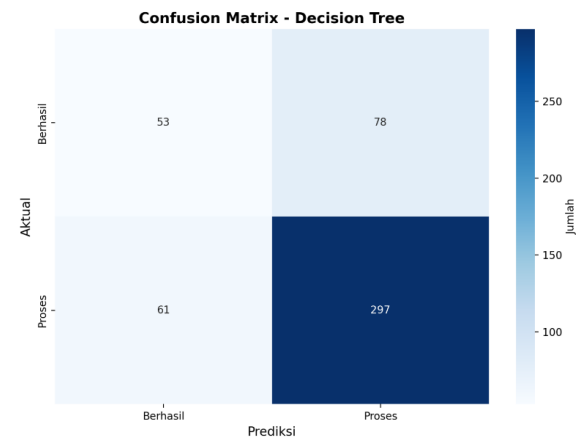
Gambar 2. Perbandingan Metrik Evaluasi

Analisis Hasil:

1. Random Forest Mengungguli Decision Tree: Random Forest memberikan performa yang lebih baik di semua metrik. Peningkatan accuracy sebesar 3.07% (dari 71.57% menjadi 74.64%) menunjukkan keunggulan ensemble method dalam menangani variasi data. Nilai accuracy 74.64% menunjukkan model dapat memprediksi status penanganan dengan benar pada sekitar 3 dari 4 kasus.
2. Konsistensi Metrik: Nilai accuracy dan recall yang identik (0.7157 untuk DT, 0.7464 untuk RF) mengindikasikan bahwa model tidak menggunakan threshold adjustment khusus. Perbedaan antara precision dan recall yang tidak terlalu besar menunjukkan performa yang cukup seimbang antar kelas.

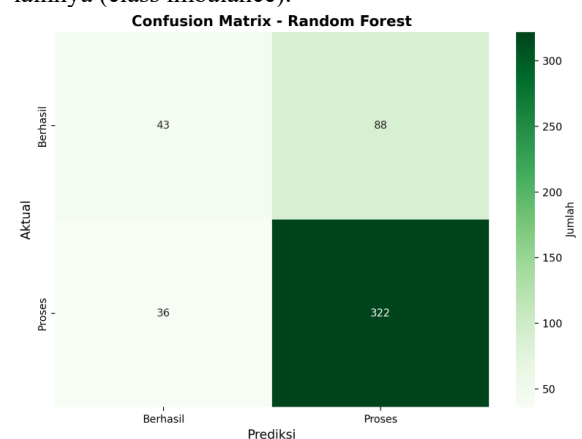
3. Peningkatan Precision: Random Forest menunjukkan peningkatan precision sebesar 1.64% (dari 70.44% menjadi 72.08%), yang berarti model lebih akurat dalam memprediksi kasus positif dengan mengurangi false positives.

4. F1-Score sebagai Metrik Balance: F1-Score Random Forest (72.36%) lebih tinggi 1.44% dibandingkan Decision Tree (70.92%), mengkonfirmasi bahwa ensemble method memberikan keseimbangan yang lebih baik antara precision dan recall.



Gambar 3. Confusion Matrix Decision Tree

Hasil analisis confusion matrix pada model Decision Tree menunjukkan bahwa model sudah cukup baik dalam memprediksi kelas “Proses”, yang merupakan kelas mayoritas dalam dataset. Namun, masih terdapat cukup banyak kesalahan prediksi, terutama antara kelas “Berhasil” dan “Proses”. Selain itu, model mengalami kesulitan dalam memprediksi kelas “Gagal”, yang kemungkinan disebabkan oleh jumlah sampel yang lebih sedikit dibandingkan kelas lainnya (class imbalance).



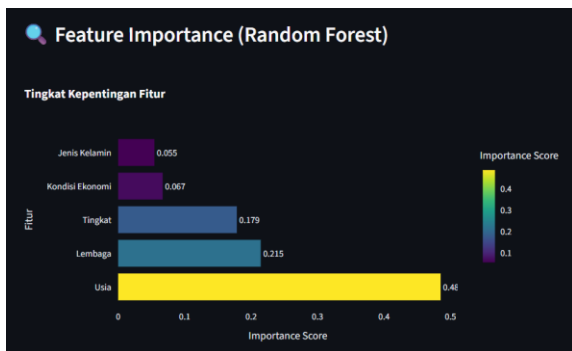
Gambar 4. Confusion Matrix Random Forest

Pada model Random Forest, terlihat adanya peningkatan performa dibandingkan Decision Tree. Model ini lebih akurat dalam memprediksi kelas mayoritas (“Proses”) dan jumlah kesalahan prediksi secara umum berkurang. Meskipun demikian,

tantangan dalam memprediksi kelas minoritas (“Gagal”) masih tetap ada.

Analisis Feature Importance

Analisis feature importance dari model Random Forest menghasilkan ranking fitur berdasarkan kontribusinya terhadap prediksi. Hasil ini dapat dilihat secara visual pada halaman Model & Evaluasi di dashboard. Analisis feature importance dari model Random Forest menghasilkan ranking sebagai berikut:



Gambar 5. Grafik Feature Importance (Random Forest)

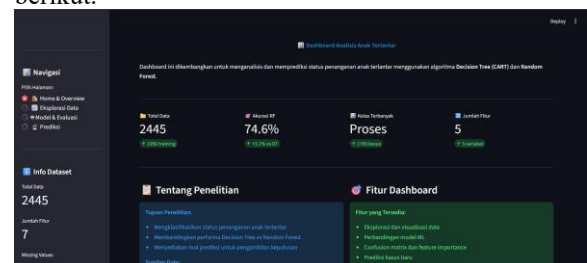
- Usia (Importance: 48.53% - Tertinggi)**
Usia anak menjadi faktor yang paling berpengaruh dalam menentukan status penanganan anak terlantar, dengan kontribusi hampir setengah dari total nilai importance. Hal ini menunjukkan bahwa tahap perkembangan anak sangat memengaruhi keberhasilan penanganan. Setiap kelompok usia memiliki kebutuhan dan tantangan yang berbeda. Anak usia dini relatif lebih mudah beradaptasi sehingga intervensi yang dilakukan sejak awal cenderung memberikan dampak jangka panjang. Anak usia sekolah membutuhkan kestabilan rutinitas dan dukungan pendidikan, sedangkan remaja menghadapi tantangan yang lebih kompleks terkait pencarian jati diri dan pengaruh lingkungan sosial.
- Lembaga (Importance: 21.46% - Sangat Penting)**
Lembaga yang menangani anak terlantar memiliki pengaruh yang cukup besar terhadap hasil penanganan. Perbedaan kualitas layanan antar lembaga dapat memengaruhi keberhasilan intervensi. Faktor seperti ketersediaan sumber daya, kompetensi pekerja sosial, kualitas program, serta budaya kerja lembaga berperan penting dalam menentukan outcome penanganan.
- Tingkat Pendidikan (Importance: 17.86% - Penting)**
Tingkat pendidikan anak juga berpengaruh cukup besar terhadap status penanganan. Anak yang masih bersekolah umumnya memiliki rutinitas yang lebih teratur, dukungan sosial dari

lingkungan sekolah, serta peluang yang lebih baik untuk masa depan. Sebaliknya, anak yang tidak bersekolah atau putus sekolah sering kali menghadapi permasalahan yang lebih kompleks, seperti kondisi ekonomi keluarga atau kurangnya dukungan dari lingkungan sekitar.

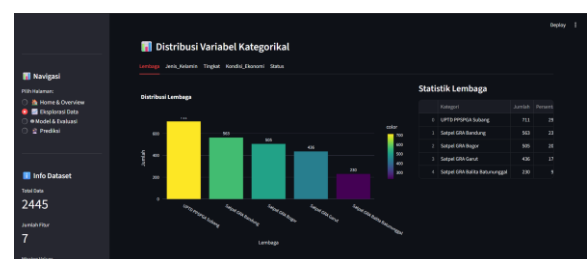
- Kondisi Ekonomi (Importance: 6.70% - Cukup Penting)**
Kondisi ekonomi memiliki pengaruh yang relatif lebih kecil dibandingkan variabel lainnya. Meskipun demikian, faktor ekonomi tetap berperan dalam kasus anak terlantar. Rendahnya nilai importance kemungkinan disebabkan oleh keterkaitan dengan variabel lain, seperti usia dan pendidikan, serta adanya program bantuan yang membantu mengurangi dampak langsung kondisi ekonomi terhadap anak.
- Jenis Kelamin (Importance: 5.46% - Kurang Penting)**
Jenis kelamin memiliki pengaruh paling rendah terhadap status penanganan anak terlantar. Hal ini menunjukkan bahwa permasalahan anak terlantar dapat terjadi baik pada anak laki-laki maupun perempuan dengan tingkat yang relatif serupa. Meskipun demikian, tetap perlu memperhatikan risiko tertentu yang mungkin berbeda antara anak laki-laki dan perempuan.

Implementasi Dashboard Interaktif

Tampilan antarmuka dashboard yang digunakan dalam penelitian ini ditunjukkan pada gambar berikut.



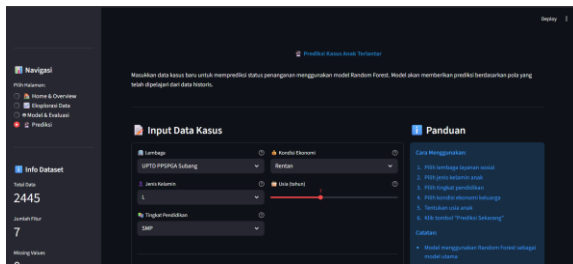
Gambar 6. Tampilan Halaman Home & Overview



Gambar 7. Tampilan Halaman Eksplorasi Data



Gambar 8. Tampilan Halaman Model & Evaluasi



Gambar 9. Tampilan Halaman Prediksi

Dashboard yang dikembangkan memiliki beberapa fitur utama yang saling terintegrasi untuk mendukung proses analisis dan pengambilan keputusan. Fitur tersebut meliputi halaman Home & Overview yang menyajikan ringkasan informasi penelitian serta indikator kinerja utama dalam bentuk metrik dan visualisasi singkat, halaman Eksplorasi Data yang memungkinkan pengguna melihat pratinjau dataset, statistik deskriptif, serta pola distribusi dan hubungan antar variabel melalui grafik interaktif, halaman Model & Evaluasi yang menampilkan perbandingan performa model klasifikasi beserta confusion matrix, feature importance, dan classification report, serta halaman Prediksi yang menyediakan form input data untuk menghasilkan hasil prediksi secara langsung lengkap dengan tingkat kepercayaan dan interpretasi otomatis. Secara keseluruhan, dashboard ini dirancang untuk memudahkan pengguna dalam memahami data, mengevaluasi model, dan memperoleh insight secara cepat dan informatif.

Kesimpulan dan Saran

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut:

1. Model Random Forest menunjukkan performa yang lebih baik dibandingkan Decision Tree dalam memprediksi status penanganan anak terlantar. Random Forest mencapai accuracy sebesar 74.64%, precision 72.08%, recall 74.64%, dan F1-score 72.36%. Sementara Decision Tree mencapai accuracy 71.57%, precision 70.44%, recall 71.57%, dan F1-score 70.92%. Peningkatan accuracy sebesar 3.07 poin persentase menunjukkan bahwa ensemble method Random Forest lebih efektif dalam mengurangi overfitting dan menangani variasi data.
2. Usia anak merupakan faktor paling dominan yang mempengaruhi keberhasilan penanganan

anak terlantar dengan kontribusi feature importance sebesar 48.53%. Hal ini mengindikasikan bahwa tahap perkembangan anak sangat menentukan efektivitas intervensi. Faktor lain yang berpengaruh secara berurutan adalah lembaga penanganan (21.46%), tingkat pendidikan (17.86%), kondisi ekonomi (6.70%), dan jenis kelamin (5.46%).

3. Dashboard interaktif berbasis Streamlit berhasil dikembangkan dengan empat halaman utama: Home & Overview, Eksplorasi Data, Model & Evaluasi, dan Prediksi. Dashboard ini memungkinkan pengguna untuk melakukan eksplorasi data, membandingkan performa model, dan melakukan prediksi kasus baru secara real-time dengan interface yang user-friendly.
4. Sistem yang dikembangkan dapat mendukung pengambilan keputusan dalam penanganan anak terlantar dengan menyediakan prediksi berbasis data, analisis faktor-faktor penting, dan visualisasi yang memudahkan pemahaman pola data. Model dengan accuracy 74.64% berarti dapat memprediksi status penanganan dengan benar pada sekitar 3 dari 4 kasus.

Untuk penelitian selanjutnya, variabel yang digunakan sebaiknya diperluas, tidak hanya mencakup aspek demografis dan ekonomi, tetapi juga kondisi psikologis anak, kesehatan, dukungan sosial, jenis layanan yang diterima, serta lingkungan tempat tinggal. Data yang digunakan juga dapat diperluas dengan melibatkan lebih banyak lembaga dan wilayah provinsi, serta menggunakan data longitudinal agar perubahan kondisi anak dapat diamati dari waktu ke waktu. Dari sisi pemodelan, pengembangan dapat dilakukan dengan menyempurnakan pengaturan parameter, mencoba algoritma machine learning yang lebih beragam, serta menerapkan teknik untuk mengatasi data yang tidak seimbang. Selain itu, penggunaan metode yang dapat menjelaskan cara kerja model penting agar hasil prediksi lebih mudah dipahami. Validasi menggunakan data dari wilayah atau lembaga lain juga perlu dilakukan untuk memastikan hasil yang konsisten. Pengembangan dashboard ke depan dapat difokuskan pada penambahan fitur analisis yang lebih mendalam, integrasi dengan database dan API, serta peningkatan kinerja sistem agar lebih siap digunakan dalam kegiatan operasional. Fitur pengelolaan pengguna berdasarkan peran dan dukungan aplikasi mobile juga dapat dipertimbangkan untuk memudahkan penggunaan di lapangan.

Daftar Pustaka

- [1] K. Miyeki, S. I. Bonjol, and S. I. Bonjol, "Implementasi Penanganan Antar (Anak Terlantar) Di Dinas Sosial Kabupaten Lima Puluh Kota," vol. 1, no. 2, 2023.

- [2] S. J. Yanca and L. C. Johnson, *Generalist social work practice with families*. Princeton, N.J.: Recording for the Blind & Dyslexic, 2008.
- [3] G. I. P. Nastia, S. Sulastri, and E. Nuriyah, "UPAYA PENINGKATAN KAPASITAS KELUARGA DALAM PENGASUHAN ANAK (Studi Kasus Pada Proses Perlindungan Anak Terlantar oleh Rumah Perlindungan Sosial Asuhan Anak (RPSAA) Ciumbuleuit Kota Bandung)," *sswj*, vol. 11, no. 2, p. 81, Feb. 2022, doi: 10.24198/share.v11i2.37040.
- [4] M. Salidyn, H. A. Kadir, and W. Wahba, "Analisis Sistem Pengelolaan Penanganan Anak Terlantar Pada Dinas Sosial Kota Palu," *J Sin Man*, vol. 7, no. 2, pp. 95–104, Aug. 2020, doi: 10.56338/jsm.v7i2.1229.
- [5] D. S. Nurfadhilah, A. Setiawan, and R. Zulkifli, "Implementasi Geospatial Clustering Menggunakan Algoritma K-Means Untuk Penentuan Lokasi Strategis Promosi Kampus Universitas Langlangbuana," vol. 5, no. 2, 2025.
- [6] E. Susilowati, "PRAKTIK PERLINDUNGAN ANAK TERLANTAR DI LEMBAGA KESEJAHTERAAN SOSIAL ANAK".
- [7] "Child Protection," Child Protection Keeping children safe from violence, neglect and exploitation. [Online]. Available: <https://www.unicef.org/indonesia/child-protection>
- [8] F. Hilmi, K. Taqiyassar, N. R. P. Pratama, S. C. Kusuma, H. R. Nurwachid, and T. N. Fatyanosa, "ANALISIS PERBANDINGAN MODEL MACHINE LEARNING TREE-BASED DAN NON-TREE-BASED UNTUK TUGAS KLASIFIKASI".
- [9] H. Helmy, A. Nursyahid, and T. A. Setyawan, "Penerapan *Machine Learning* untuk Mengendalikan Parameter Budidaya Tanaman Hidroponik Berbasis *Edge* dan *Cloud Computing*," *JTIK*, vol. 11, no. 4, pp. 827–836, Aug. 2024, doi: 10.25126/jtiik.1148631.
- [10] S. F. Damanik, A. Wanto, and I. Gunawan, "Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok," *JKDN*, vol. 1, no. 2, pp. 21–32, Jan. 2022, doi: 10.58982/krisnadana.v1i2.108.
- [11] Eka Wulansari Fidayanthie, Asep Sayfullloh, Mardiana Rafa Alzena, and Nilam Kurnia Sari, "Analisis Perbandingan Algoritma Random Forest dan Algoritma Naive Bayes untuk Memprediksi Penyakit Paru-Paru di Indonesia," *Saturnus*, vol. 3, no. 3, pp. 123–135, Jul. 2025, doi: 10.61132/saturnus.v3i3.956.
- [12] A. H. Mubarak, P. Pujiono, D. Setiawan, D. F. Wicaksono, and E. Rimawati, "Parameter Testing on Random Forest Algorithm for Stunting Prediction," *Sinkron*, vol. 9, no. 1, pp. 107–116, Jan. 2025, doi: 10.33395/sinkron.v9i1.14264.
- [13] E. R. Susanto, M. R. Inzaghi, A. Amarudin, and N. Neneng, "Evaluasi Kinerja Model Random Forest Dalam Memprediksi Diabetes Berdasarkan Dataset Kesehatan di Indonesia," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 7, pp. 1857–1866, Jul. 2025, doi: 10.52436/1.jpti.871.
- [14] F. Rizki, S. Zahra, N. C. N. B. Tarigan, K. Ramadani, A. Alawiah, and T. S. Tambunan, "Analisis Faktor Penyebab Anak Terlantar di Kota Medan," *jishs*, vol. 3, no. 3, pp. 915–919, Dec. 2024, doi: 10.62379/jishs.v3i3.2478.
- [15] Ekin Adhi Guna, M. Davin Diza Ghifary, Esra Fransiska Sihombing, and Age Pius Datubara, "Implementasi Algoritma Decision Tree untuk Klasifikasi Data Evaluation Car Menggunakan Python," *jusiik-widyakarya*, vol. 1, no. 4, pp. 167–177, Nov. 2023, doi: 10.59581/jusiik-widyakarya.v1i4.1830.