

# KLASIFIKASI CALON PENDONOR DARAH POTENSIAL MENGGUNAKAN ALGORITMA DECISION TREE DI UTD PMI KOTA SURABAYA

Syifa Saskia Elfaretta<sup>1\*</sup>, Amalia Anjani Arifiyanti<sup>2</sup>, Anindo Saka Fitri<sup>3</sup>

<sup>1,2,3</sup>Universitas Pembangunan Nasional “Veteran” Jawa Timur; Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294; (031) 8706369

Received: 14 Juli 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

## Keywords:

Potential Blood Donor

Classification;

Decision Tree;

CART;

C4.5;

Random Forest.

**Abstrak.** Darah merupakan komponen yang vital dalam tubuh manusia. Kurangnya jumlah darah pada tubuh akan memengaruhi kerja dari organ lain. Oleh karena itu, PMI berperan aktif dalam menyediakan kebutuhan stok darah nasional. Untuk memastikan bahwa darah yang diterima oleh resipien aman dan berkualitas baik, maka perlu dilakukan klasifikasi calon pendonor darah potensial. Penelitian ini menggunakan beberapa algoritma *Decision Tree* dalam proses klasifikasi data. Algoritma yang digunakan adalah CART, C4.5, dan *Random Forest*. Hasil perbandingan dari tiga algoritma menunjukkan bahwa *Random Forest* memiliki nilai terbaik dibandingkan algoritma lainnya. Algoritma *Random Forest* mendapatkan akurasi dengan nilai 97% dan AUC ROC dengan nilai 99%. Oleh karena itu, algoritma *Random Forest* diimplementasikan dalam sistem klasifikasi calon pendonor darah potensial berbasis *web*. Hasil uji validasi sistem menunjukkan akurasi dengan angka 97%.

## Correspondent Email:

saskiaelfaretta@gmail.com

**Abstract.** Blood is one of the most important component in every human body. Deficiency of blood in a body could affects other organ's functionality. Therefore, the Indonesian Red Cross (PMI) plays an active role in providing the national blood supply. To ensure recipients received a safe and good quality blood supply, it is necessary to classify potential blood donors. This research uses several *Decision Tree* algorithms in the data classification process. The algorithms used are CART, C4.5, and *Random Forest*. The comparison results of the three algorithms show that *Random Forest* has the best performance compared to the others. The *Random Forest* algorithm achieved an accuracy of 97% and an AUC ROC of 99%. Therefore, the *Random Forest* algorithm is implemented in a web-based classification system for potential blood donors. System validation tests show an accuracy rate of 97%.

## 1. PENDAHULUAN

Darah merupakan komponen yang vital dalam tubuh manusia. Ketika darah dalam tubuh mengalami gangguan, maka kerja dari organ lain akan ikut berpengaruh. Salah satu contoh dari gangguan pada darah adalah

kurangnya jumlah darah pada tubuh. Sampai dengan saat ini, masih belum ditemukan penelitian di mana manusia dapat menciptakan darah. Oleh karena itu, tubuh manusia sangat berharga sebagai satu-satunya produsen yang dapat menghasilkan darah.

Tidak semua orang diberikan karunia memiliki kualitas darah yang baik dan cukup. Orang-orang dengan kondisi tertentu sangat beruntung ketika bisa mendapatkan donor darah dari sukarelawan. Donor darah adalah aktivitas pengambilan darah dari seseorang secara sukarela untuk disimpan ke dalam bank darah yang akan digunakan untuk keperluan transfusi darah [1]. Untuk mencapai kebutuhan akan jumlah darah tersebut, maka PMI (Palang Merah Indonesia) berperan aktif dalam menyediakan stok darah nasional.

Darah yang telah didonorkan oleh pendonor akan disalurkan oleh PMI kepada siapa saja yang berhak menerimanya. Akan tetapi, selain mengutamakan kebutuhan akan permintaan darah, PMI juga harus memastikan bahwa darah yang diterima oleh resipien aman dan berkualitas baik. Oleh karena itu, penelitian ini akan memprediksi apakah seorang calon pendonor darah berpotensi untuk mendonorkan darahnya atau tidak. Tujuan dari prediksi ini adalah memberikan manfaat kepada pihak luar, termasuk PMI dalam menentukan siapa saja yang berpotensi mendonorkan darahnya. Untuk menentukan hasil prediksi tersebut, penelitian ini akan menggunakan suatu algoritma klasifikasi. Klasifikasi adalah teknik yang digunakan untuk menentukan kelas dari suatu objek berdasarkan data-data yang sudah ada [2].

Proses klasifikasi calon pendonor darah untuk menentukan potensi tidaknya telah dilakukan oleh beberapa peneliti, yaitu Yuda Irawan dengan menggunakan algoritma *Decision Tree* C4.5 untuk prediksi kelayakan calon pendonor darah [3], Hermanto Wahono dkk. dengan menggunakan algoritma *Naïve Bayes*, *K-Nearest Neighbors*, dan *Decision Tree* C4.5 untuk prediksi calon pendonor darah darah potensial [4]. Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* menunjukkan akurasi yang lebih tinggi sebesar 93.83%, dibandingkan *Naïve Bayes* dengan akurasi sebesar 85.15% dan *K-Nearest Neighbors* dengan akurasi 84.10%. Selanjutnya penelitian dari Ni Ketut Melly Astuti dkk. dengan menggunakan algoritma *Decision Tree* C4.5 dan *K-Nearest Neighbors* untuk klasifikasi data donor darah [5]. Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* C4.5 menunjukkan akurasi yang lebih tinggi sebesar 92.9%, dibandingkan *K-Nearest Neighbors* sebesar 91%.

Dengan demikian, algoritma pada pengklasifikasian data yang digunakan dalam penelitian ini adalah *Decision Tree*. *Decision Tree* akan mengubah kumpulan data besar menjadi pohon keputusan yang merepresentasikan aturan-aturan keputusan [6]. Data masukan yang diberikan pengguna akan dianalisis menggunakan rumus perhitungan dari beberapa model algoritma *Decision Tree*. Penelitian ini akan membandingkan model CART, C4.5, dan *Random Forest*. Perbandingan dari tiga skenario yang terbentuk akan dievaluasi menggunakan *confusion matrix* dan AUC ROC. Skenario yang menunjukkan hasil evaluasi metrik terbaik akan diimplementasikan dalam sistem berbasis *web*. Hasil luaran sistem berupa prediksi calon pendonor darah apakah berpotensi untuk dapat mendonorkan darahnya atau tidak.

## 2. TINJAUAN PUSTAKA

### 2.1. Data Mining

Data dalam jumlah besar (*big data*) seringkali tidak diolah dengan benar, dan hanya dibiarkan tersimpan dalam memori, kemudian terbuang begitu saja sebagai sampah. Hal ini memang sesuai dengan definisi, bahwa data adalah fakta yang terekam, tanpa memiliki arti. Padahal *big data* tersebut dapat menghasilkan suatu informasi berguna yang tidak pernah diketahui sebelumnya. Untuk mengatasi sampah dari data tersebut agar dapat menjadi informasi, diperlukan proses pengolahan yang disebut dengan *data mining*.

*Data mining* atau penambangan data adalah analisis dari *big data* untuk menemukan pola tersembunyi, dan meringkas data agar dapat dipahami dan menjadi pengetahuan yang berguna bagi pemilik data [7]. Pengetahuan yang ditemukan dari penambangan data ini akan membantu perusahaan dalam pengambilan keputusan untuk strategi bisnis yang lebih baik.

### 2.2. Decision Tree

*Decision Tree* (pohon keputusan) adalah salah satu algoritma klasifikasi pada *data mining*. *Decision Tree* adalah diagram alir yang menyerupai struktur pohon, di mana setiap internal *node* menunjukkan pengujian terhadap suatu atribut, dan hasilnya ditunjukkan oleh setiap cabang, kemudian diakhiri dengan pembagian kelas yang ditunjukkan oleh *leaf*

*node* [8]. Struktur pohon keputusan terdiri dari simpul (*node*) dan cabang (*branch*).

Menurut Junita dan Bachtiar (2019), kelebihan dari algoritma *Decision Tree* adalah algoritma ini dapat menyelesaikan permasalahan yang terjadi pada KNN, SVM, dan *Naïve Bayes*. Algoritma *Decision Tree* mudah untuk diimplementasikan, membutuhkan waktu komputasi dan penyimpanan yang rendah, serta dapat mereduksi penggunaan fitur yang digunakan [9].

### 2.3. CART

CART (*Classification and Regression Trees*) adalah salah satu algoritma dari *Decision Tree*. Algoritma CART terdiri dari dua metode, yaitu pohon klasifikasi dan pohon regresi. CART akan menghasilkan pohon klasifikasi ketika variabel target bertipe kategorikal, sedangkan jika variabel target bertipe numerikal, maka CART akan menghasilkan pohon regresi. CART dalam pohon klasifikasi, penentuan *node* ditentukan berdasarkan nilai *gini index* terendah. *Gini index* adalah kriteria yang didasarkan pada ketidakmurnian yang mengukur perbedaan antara distribusi probabilitas dari nilai variabel target [10].

### 2.4. C4.5

Algoritma C4.5 adalah algoritma ID3 yang dikembangkan oleh Quinlan Ross pada tahun 1993. Jika algoritma ID3 menggunakan perhitungan *entropy* dan *information gain*, maka C4.5 menambahkan perhitungan *split info* dan *gain ratio* dalam menentukan atribut yang akan menjadi *node*. Menurut Qadrini dkk. (2021), beberapa karakteristik C4.5 yang juga merupakan perbaikan dari ID3:

- Dapat menangani data yang hilang (*missing value*).
- Dapat menangani atribut kontinu.
- Melakukan *pruning* untuk memperoleh performa yang lebih baik.
- Menggunakan *gain ratio* untuk mengatasi bias terhadap atribut dengan banyak nilai.

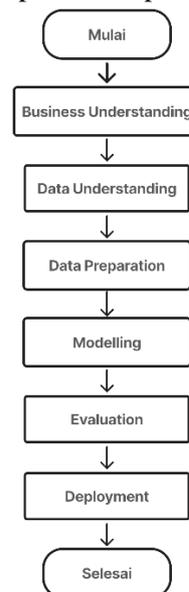
### 2.5. Random Forest

*Random Forest* adalah gabungan dari beberapa *Decision Tree*, di mana pada masing-masing pohon telah dilakukan pelatihan

menggunakan sampel individu dan setiap atribut dipecah pada pohon yang dipilih antara atribut subset yang bersifat acak [11]. Dalam pembuatan pohon keputusan pada algoritma *Random Forest* sama dengan proses yang dilakukan pada CART, hanya saja pada *Random Forest* tidak dilakukan *pruning* (pemangkasan). Indeks gini digunakan untuk memilih fitur di setiap simpul internal dari pohon keputusan.

## 3. METODE PENELITIAN

Tahapan dalam penelitian ini dilakukan dengan mengikuti kerangka kerja CRISP-DM. CRISP-DM (*Cross Industry Standard Process for Data Mining*) adalah suatu standarisasi dalam pemrosesan *data mining*, di mana data yang didapat akan melalui tahapan terstruktur dan terdefinisi dengan jelas dan efisien [12]. Kerangka kerja CRISP-DM memiliki enam tahapan yang dapat dilihat pada gambar 1.



Gambar 1. Metode Penelitian

Penelitian ini akan dimulai dengan *business understanding*, memahami permasalahan yang ingin diselesaikan, kemudian pada *data understanding*, data yang didapat diidentifikasi dan ditinjau apakah ada masalah dalam pengelolaan dan kualitasnya. Setelah itu, temuan masalah dan kualitas data akan diselesaikan pada tahap *data preparation*. Selanjutnya dalam *modelling*, data yang telah diproses dan dipastikan siap akan diterapkan algoritma *Decision Tree* untuk mengidentifikasi pola dalam data yang dimiliki. Setelah itu, pola yang ditemukan dari masing-masing model

akan ditinjau dan dinilai potensi yang dihasilkan pada tahap *evaluation*. Terakhir pada tahap *deployment*, model yang menghasilkan temuan pola dengan potensi tinggi akan diimplementasikan dalam sistem berbasis *web*.

#### 4. HASIL DAN PEMBAHASAN

##### 4.1. Business Understanding

Tujuan dari pemodelan data ini adalah untuk memprediksi calon pendonor darah potensial. Manfaatnya dapat digunakan oleh masyarakat umum untuk mengetahui apakah dirinya memiliki potensi untuk mendonorkan darahnya atau tidak. Data yang diperoleh dari *database* UTD PMI Kota Surabaya perlu digali untuk menemukan pola tersembunyi di dalamnya. Pola tersembunyi tersebut berupa klasifikasi data calon pendonor darah yang diproses menggunakan algoritma dengan fungsi klasifikasi. Temuan pola yang dihasilkan dari pemodelan dengan algoritma klasifikasi adalah potensi tidaknya calon pendonor.

Setelah melalui proses pemodelan, maka sistem *web* akan dibangun. Pengguna aplikasi ini adalah masyarakat umum yang ingin mendonorkan darahnya. Calon pendonor darah dapat memprediksi potensi mereka.

##### 4.2. Data Understanding

Data yang didapat dari UTD PMI Kota Surabaya akan dianalisis terkait kesesuaiannya untuk proses pemodelan. Data tersebut merupakan data calon pendonor darah pada Januari 2021 sampai dengan Februari 2024. *Dataset* yang diolah berjumlah 6 kolom dengan 544.504 baris.

Tabel 1. Atribut dalam *Dataset*

Atribut	Keterangan
usia	Usia calon pendonor ketika akan mendonorkan darahnya
tensi1	Angka atas dalam pembacaan tekanan darah
tensi2	Angka bawah dalam pembacaan tekanan darah
bb	Berat badan calon pendonor
hb	Jumlah hemoglobin yang ada dalam darah
tolak	Apakah ditolak atau tidak menjadi pendonor darah

##### 4.2.1 Missing Values Check

Proses pengecekan nilai yang hilang akan dilakukan pada tahap ini. Hasil dari pengecekan menunjukkan bahwa terdapat 1 dari 6 kolom yang memiliki *missing values*. Kolom tersebut adalah kolom 'tolak' yang dapat dilihat pada gambar 2.

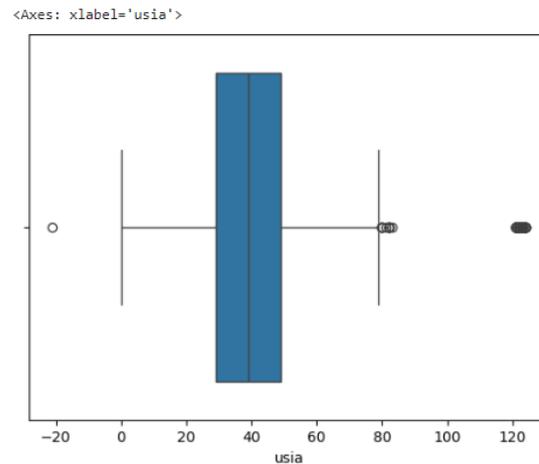
```

Your selected dataframe has 6 columns.
There are 1 columns that have missing values.
Missing Values % of Total Values
tolak 40494 7.4
    
```

Gambar 2. Kolom dengan *Missing Values*

##### 4.2.2 Outlier Check

Setelah melakukan pengecekan kolom dengan nilai yang hilang, pada tahap ini kolom akan diperiksa apakah semua nilai berada dalam batas yang sewajarnya, atautkah ada nilai yang tidak pada umumnya. Kolom yang berisi nilai di luar batas wajar akan terdeteksi sebagai *outlier*. Dalam penelitian ini, pengecekan *outlier* dilakukan dengan cara melihat visualisasi *boxplot* setiap kolom. Gambar 3 merupakan *boxplot* kolom 'usia' yang terdeteksi *outlier*.



Gambar 3. *Boxplot* Kolom 'usia'

##### 4.3. Data Preparation

Kolom yang terdeteksi *missing values* dan *outlier* akan diatasi pada tahap ini.

##### 4.3.1 Handle Missing Values

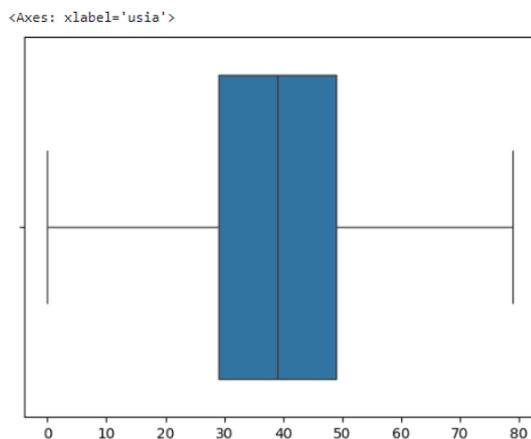
Pada sub 4.2.1 telah ditemukan satu kolom *missing values*, yaitu kolom 'tolak'. Kolom tersebut akan diatasi dengan mengisi nilai yang hilang dengan nilai modus. Hal ini dilakukan karena kolom 'tolak' bertipe kategorik.

```
Your selected dataframe has 6 columns.
There are 0 columns that have missing values.
Empty DataFrame
Columns: [Missing Values, % of Total Values]
Index: []
```

Gambar 4. *Missing Values* Berhasil Teratasi

### 4.3.2 Handle Outliers

Pada tahap ini akan diatasi kolom dengan nilai tidak wajar. Untuk menangani kondisi tersebut, maka digunakan metode IQR (*Interquartile Range*) sebagai penentu batas atas dan bawah untuk mendeteksi *outlier*. Nilai yang terdeteksi *outlier* akan diganti dengan nilai median dari kolom tersebut. Gambar 5 merupakan *boxplot* kolom 'usia' setelah *handle outlier*.



Gambar 5. Kolom 'usia' Berhasil Teratasi

### 4.3.3 Categorical Data Encoding

Kolom dengan tipe kategorik tidak akan bisa terbaca oleh sistem. Kolom tersebut harus diganti terlebih dahulu menjadi numerik dengan bantuan *encoding*. Kolom kategorik yang pada penelitian adalah kolom 'tolak'.

	usia	tensi1	tensi2	bb	tolak	hb
0	56.0	130.0	70.0	82.0	0	13.2
1	66.0	160.0	90.0	67.0	0	15.8
2	47.0	160.0	100.0	105.0	0	16.2
3	69.0	150.0	100.0	66.0	0	14.9
4	47.0	124.0	86.0	74.0	0	15.7

Gambar 6. Sampel *Dataset* Bersih

## 4.4. Modelling

Pelatihan dan pengujian pada 3 skenario dilakukan pada tahap ini.

### 4.4.1 Train Test Split

Pembagian data dalam penelitian ini akan menggunakan metode *hold-out*. *Hold-out* adalah metode pembagian data sederhana ke dalam data latih dan data uji secara acak dan saling bebas (tidak tumpang tindih). Metode *hold-out* juga mampu menjamin proporsional kelas dalam masing-masing data [13]. Penelitian ini akan membagi proporsi data menjadi 80:10:10. 80 sebagai data latih, 10 sebagai data uji, dan 10 disimpan untuk data validasi [14].

### 4.4.2 Skenario 1 (CART)

```
Accuracy: 0.9645552882408036
Confusion Matrix:
[[46945  812]
 [ 1118 5576]]
Classification Report:
              precision    recall  f1-score   support
0               0.98         0.98         0.98     47757
1               0.87         0.83         0.85     6694
accuracy                0.96         0.96     54451
macro avg              0.92         0.91         0.92     54451
weighted avg           0.96         0.96         0.96     54451
```

Gambar 7. *Confusion Matrix* dan Akurasi Algoritma CART

Gambar 7 menunjukkan perhitungan banyaknya data yang diprediksi dengan benar, dan banyaknya data yang diprediksi salah melalui *confusion matrix*. Data yang prediksinya benar dihasilkan dari perhitungan antara TP (*True Positive*) dan TN (*True Negative*), sedangkan data yang prediksinya salah dihasilkan dari perhitungan antara FP (*False Positive*) dan FN (*False Negative*), sehingga melalui algoritma CART data yang diprediksi dengan benar berjumlah 52.521 dan data yang diprediksi salah berjumlah 1.930 data. Dari *confusion matrix* tersebut juga dapat diketahui akurasi dari model CART melalui perhitungan jumlah data yang diprediksi dengan benar dibagi dengan jumlah seluruh data, sehingga akurasi yang didapat dari model CART sebesar 96%.

### 4.4.3 Skenario 2 (C4.5)

```

Accuracy: 0.9643165414776588
Confusion Matrix:
[[46947  810]
 [ 1133 5561]]
Classification Report:

```

	precision	recall	f1-score	support
0	0.98	0.98	0.98	47757
1	0.87	0.83	0.85	6694
accuracy			0.96	54451
macro avg	0.92	0.91	0.92	54451
weighted avg	0.96	0.96	0.96	54451

Gambar 8. Confusion Matrix dan Akurasi Algoritma C4.5

Berdasarkan *confusion matrix* pada gambar 8, ditemukan data yang diprediksi dengan benar oleh algoritma C4.5 berjumlah 52.508 data, sedangkan data yang diprediksi salah berjumlah 1.943 data. Kemudian dapat diketahui akurasi yang didapat oleh C4.5 sebesar 96%.

#### 4.4.4 Skenario 3 (Random Forest)

```

Accuracy: 0.9684119667223743
Confusion Matrix:
[[47048  709]
 [ 1011 5683]]
Classification Report:

```

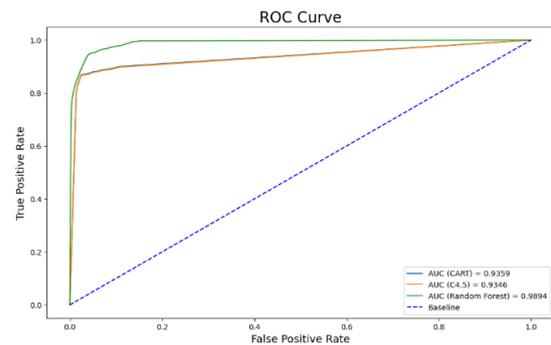
	precision	recall	f1-score	support
0	0.98	0.99	0.98	47757
1	0.89	0.85	0.87	6694
accuracy			0.97	54451
macro avg	0.93	0.92	0.93	54451
weighted avg	0.97	0.97	0.97	54451

Gambar 9. Confusion Matrix dan Akurasi Algoritma Random Forest

Berdasarkan *confusion matrix* pada gambar 9, ditemukan data yang diprediksi dengan benar oleh algoritma *Random Forest* berjumlah 52.731 data, sedangkan data yang diprediksi salah berjumlah 1.720 data. Kemudian dapat diketahui akurasi yang didapat oleh *Random Forest* sebesar 97%.

#### 4.5. Evaluation

Tiga skenario pemodelan data telah selesai dilakukan, selanjutnya pada tahap ini akan dibandingkan akurasi, nilai *error*, serta AUC ROC dari setiap skenario untuk menentukan algoritma terbaik yang akan digunakan dalam mengklasifikasikan calon pendonor darah potensial.



Gambar 10. Perbandingan AUC ROC Ketiga Skenario

AUC ROC dari ketiga skenario akan digabungkan seperti yang terlihat pada gambar 10. Tujuan dari penggabungan ini adalah agar perbandingan setiap skenario dapat terlihat dengan jelas. Berdasarkan perbandingan AUC ROC yang ditampilkan, ditemukan bahwa nilai AUC pada algoritma *Random Forest* tertinggi dibandingkan algoritma lainnya.

	Model	Akurasi	MAE	MSE	RMSE	AUC Score
0	CART	0.964555	0.035445	0.035445	0.188268	0.935931
1	C4.5	0.964317	0.035683	0.035683	0.188901	0.934559
2	Random Forest	0.968412	0.031588	0.031588	0.177730	0.989360

Gambar 11. Perbandingan AUC ROC Ketiga Skenario

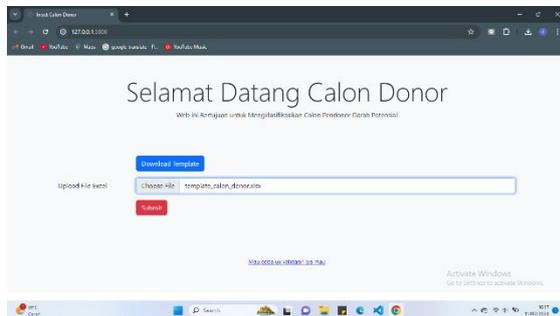
Dalam gambar 11 terlihat nilai akurasi, MAE (*Mean Absolute Error*), MSE (*Mean Squared Error*), RMSE (*Root Mean Squared Error*), serta AUC score dari masing-masing skenario. Nilai akurasi dan AUC score yang semakin tinggi menunjukkan bahwa model tersebut lebih unggul dari yang lainnya, sedangkan nilai MAE, MSE, dan RMSE berbanding terbalik dari akurasi, di mana jika nilai tersebut lebih rendah dari model lainnya, maka model tersebut lebih unggul dibanding model lain. Hal ini dikarenakan MAE, MSE, dan RMSE menghitung nilai *error* yang mungkin muncul dari prediksi yang dilakukan.

Berdasarkan gambar 11 ditemukan bahwa algoritma *Random Forest* memiliki akurasi dan AUC ROC yang tinggi dengan nilai *error* yang rendah, sehingga algoritma *Random Forest* akan dipilih untuk diimplementasikan dalam sistem klasifikasi calon pendonor darah berbasis *web*.

#### 4.6. Deployment

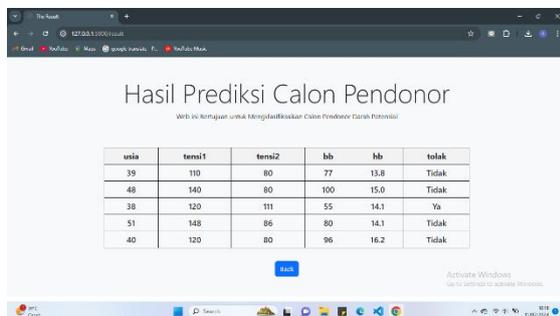
Pembuatan *website* ditujukan sebagai implementasi dari algoritma *Random Forest* dalam mengklasifikasikan data calon pendonor,

apakah berpotensi dalam mendonorkan darahnya atau tidak.



Gambar 12. Halaman Awal Web

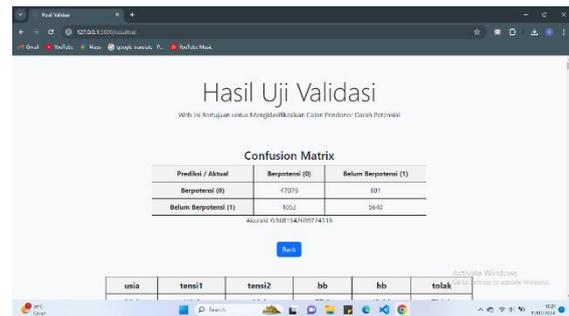
Gambar 12 merupakan tampilan awal dari halaman web. Tampilan tersebut ditujukan untuk siapa saja atau masyarakat umum yang ingin mencoba memprediksi apakah berpotensi sebagai pendonor atau tidak. Pada halaman tersebut, pengguna diminta untuk mengunggah file excel sesuai dengan template yang telah disediakan dalam button biru. Jika pengguna klik button tersebut, maka file template otomatis terunduh pada device. Setelah isi dari file disesuaikan, maka pengguna dapat klik submit dan masuk ke halaman hasil prediksi.



Gambar 13. Halaman Hasil Prediksi

Selanjutnya pada gambar 13 ditampilkan hasil prediksi dari calon pendonor yang telah melakukan submit file sesuai dengan template yang diberikan. Prediksi potensi akan terlihat pada kolom 'tolak', di mana 'Tidak' berarti tidak ditolak atau berpotensi dan 'Ya' berarti ditolak atau tidak berpotensi. Kemudian pada bagian bawah halaman terdapat tombol back yang mengarah ke halaman awal.

#### 4.6.1 Uji Validasi Sistem



Gambar 14. Pengujian Validasi Sistem

Hasil analisis dalam website selanjutnya perlu untuk dilakukan uji validasi untuk mengetahui seberapa akurat model dalam memprediksi data. Uji validasi dilakukan dengan cara menghitung nilai akurasi yang didapat dari confusion matrix. Pada penelitian ini, uji validasi akan dilakukan pada data validasi yang telah dipisahkan ketika train test split berjumlah 54.450 data. Berdasarkan perhitungan yang dilakukan, dapat disimpulkan bahwa akurasi dari sistem website dengan model Random Forest menghasilkan nilai sebesar 97%.

### 5. KESIMPULAN

- a. Pemodelan yang telah dilakukan dengan algoritma Decision Tree pada tiga skenario menghasilkan temuan bahwa algoritma Random Forest merupakan algoritma terbaik dibandingkan algoritma CART dan C4.5. Random Forest menghasilkan akurasi sebesar 97% dan AUC ROC sebesar 99%. Dengan demikian, maka algoritma Random Forest diimplementasikan dalam sistem klasifikasi calon pendonor darah potensial berbasis web.
- b. Hasil uji validasi sistem yang telah dibangun menggunakan algoritma Random Forest menghasilkan akurasi sebesar 97%.

### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

### DAFTAR PUSTAKA

[1] Fadilah dkk. (2024). Pentingnya Pengetahuan Tentang Donor Darah Terhadap Kesadaran Perilaku Masyarakat. *Jurnal Anestesi: Jurnal Ilmu Kesehatan dan Kedokteran*, 2(1), 77-87.

- [2] Handayani dkk. (2021). Komparasi Algoritma C4.5 dan Naïve Bayes dalam Penentuan Status Kelayakan Donor Darah. *SISTEMASI: Jurnal Sistem Informasi*, 10(3), 676-687.
- [3] Irawan, Y. (2021). Penerapan Algoritma Decision Tree C4.5 untuk Prediksi Kelayakan Calon Pendor Donor Darah dengan Klasifikasi Data Mining. *JTIM: Jurnal Teknologi Informasi dan Multimedia*, 2(4), 181-189.
- [4] Wahono, H., & Riana, D. (2020). Prediksi Calon Pendor Donor Darah Potensial dengan Algoritma Naïve Bayes, K-Nearest Neighbors dan Decision Tree C4.5. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 7-14.
- [5] Astuti, N. K., Utami, N. W., & Juliharta, I. G. (2022). Classification of Blood Donor Data Using C4.5 and K-Nearest Neighbor Method (Case Study: UTD PMI Bali Province). *PILAR Nusa Mandiri: Journal of Computing and Information System*, 18(1), 9-16.
- [6] Atmaja, K. J., Anandita, I. B., & Dewi, N. K. (2017). Penerapan Data Mining untuk Memprediksi Potensi Pendor Donor Darah Menjadi Pendor Tetap Menggunakan Metode Decision Tree C.45. *Jurnal Ilmu Komputer dan Sains Terapan*, 7(2), 101-108.
- [7] Yunus, M., Dahlan, H. S., & Santoso, P. B. (2014). SPK Pemilihan Calon Pendor Donor Darah Potensial dengan Algoritma C4.5 dan Fuzzy Tahani. *Jurnal EECCIS*, 8(1), 47-54.
- [8] Qadrini, L., Seppewali, A., & Aina, A. (2021). Decision Tree dan Adaboost pada Klasifikasi Penerima Program Bantuan Sosial. *JIP (Jurnal Inovasi Penelitian)*, 2(7), 1959-1966.
- [9] Junita, V., & Bachtar, F. A. (2019). Klasifikasi Aktivitas Manusia menggunakan Algoritma Decision Tree C4.5 dan Information Gain untuk Seleksi Fitur. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(10), 9426-9433.
- [10] Prabawati, N. I., Widodo, & Duskarnaen, M. F. (2019). Kinerja Algoritma Classification and Regression Tree (Cart) dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta. *Jurnal Pinter*, 3(2), 139-145.
- [11] Supriyadi dkk. (2020). Penerapan Algoritma Random Forest untuk Menentukan Kualitas Anggur Merah. *Jurnal Ilmiah Ekonomi dan Bisnis*, 13(2), 67-75.
- [12] Brown, M. S. (2014). *Data Mining for Dummies*. Penerbit John Wiley & Sons.
- [13] Abdulhakim, R., Carudin, & Dermawan, B. A. (2021). Analisis dan Penerapan Algoritma Convolutional Neural Network untuk Klasifikasi Kendaraan Prioritas. *Jurnal Sains dan Informatika*, 7(2), 135-144.
- [14] Kurniawan, W. A. & Salam, A. (2024). Penggunaan Feature Space SMOTE untuk Mengurangi Overfitting Akibat Imbalance Dataset. *Techno.COM*, 23(2), 328-337.