

**KLASIFIKASI CALON PENDONOR DARAH POTENSIAL
MENGUNAKAN ALGORITMA DECISION TREE
DI UTD PMI KOTA SURABAYA**

SKRIPSI

**Diajukan untuk memenuhi sebagian persyaratan
Dalam memperoleh gelar Sarjana Komputer
Program Studi Sistem Informasi**



Disusun Oleh:

SYIFA SASKIA ELFARETTA

20082010152

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
S U R A B A Y A
2024**

SKRIPSI

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

Disusun Oleh:
SYIFA SASKIA ELFARETTA
20082010152

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi
Program Studi Sistem Informasi Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur
pada Tanggal 12 Juli 2024

Pembimbing :

1.



Amalia Anjani Arifiyanti, S.Kom., M.Kom.
NIP. 19920812 2018032 001

2.



Anindo Saka Fitra, S.Kom., M.Kom.
NIP. 212199 30 325268

Tim Penguji :

1.


Agung Brastama Putra, S.Kom., M.Kom.
NIP. 19851124 2021211 003

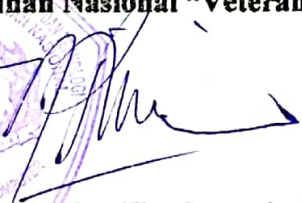
2.


Siti Mukaromah, S.Kom., M.Kom.
NIP. 19810704 2021212 011

3.


Reisa Permatasari, S.T., M.Kom.
NIP. 19920514 2022032 007

Mengetahui,
Dekan Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur


Prof. Dr. Ir. Novirina Hendrasarie, M.T.
NIP. 19681126 199403 2 001

LEMBAR PENGESAHAN

**KLASIFIKASI CALON PENDONOR DARAH POTENSIAL
MENGUNAKAN ALGORITMA DECISION TREE
DI UTD PMI KOTA SURABAYA**

Disusun Oleh:
SYIFA SASKIA ELFARETTA
20082010152


**Telah disetujui mengikuti Ujian Negara Lisan Gelombang Juli
Periode 2024 pada Tanggal 12 Juli 2024**

Menyetujui,


Dosen Pembimbing 1

Dosen Pembimbing 2


Amalia Anjani Arifiyanti, S.Kom., M.Kom.
NIP. 19920812 2018032 001


Anindo Saka Fitri, S.Kom., M.Kom.
NPT. 212199 30 325268

Mengetahui,
**Ketua Program Studi Sistem Informasi
Fakultas Ilmu Komputer**


Agung Brastama Putra, S.Kom., M.Kom.
NIP. 19851124 2021211 003



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
FAKULTAS ILMU KOMPUTER

KETERANGAN REVISI

Kami yang bertanda tangan di bawah ini menyatakan bahwa mahasiswa berikut:

Nama : Syifa Saskia Elfaretta

NPM : 20082010152

Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 12 Juli 2024 dengan judul:

**"KLASIFIKASI CALON PENDONOR DARAH POTENSIAL
MENGUNAKAN ALGORITMA DECISION TREE
DI UTD PMI KOTA SURABAYA"**

Oleh karenanya mahasiswa tersebut di atas dinyatakan bebas revisi Ujian Negara Lisan Skripsi dan diijinkan untuk membukukan laporan SKRIPSI dengan judul tersebut.

Surabaya, 22 Juli 2024

Dosen penguji yang memeriksa skripsi:

1. Agung Brastama Putra, S.Kom., M.Kom.
NIP. 19851124 2021211 003

2. Siti Mukaromah, S.Kom., M.Kom.
NIP. 19810704 2021212 011

3. Reisa Permatasari, S.T., M.Kom.
NIP. 19920514 2022032 007

Mengetahui,

Dosen Pembimbing 1

Dosen Pembimbing 2

Amalia Anjani Arifiyanti, S.Kom., M.Kom.
NIP. 19920612 2018032 001

Anindo Saka Fitri, S.Kom., M.Kom.
NPT. 212199 30 325268



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
FAKULTAS ILMU KOMPUTER

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Syifa Saskia Elfaretta

NPM : 20082010152

Program Studi : Sistem Informasi

Menyatakan bahwa Judul Skripsi / Tugas Akhir sebagai berikut:

**"KLASIFIKASI CALON PENDONOR DARAH POTENSIAL
MENGUNAKAN ALGORITMA DECISION TREE
DI UTD PMI KOTA SURABAYA"**

Bukan merupakan plagiat dari Skripsi / Tugas Akhir / Penelitian orang lain dan juga bukan merupakan Produk / Hasil Karya yang saya beli dari orang lain.

Saya juga menyatakan bahwa Skripsi / Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di Institusi Pendidikan lain. Jika ternyata dikemudian hari pernyataan terbukti benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari.

Surabaya, 22 Juli 2024

Hormat Saya,



Syifa Saskia Elfaretta

NPM. 20082010152

Judul : KLASIFIKASI CALON PENDONOR DARAH POTENSIAL
MENGUNAKAN ALGORITMA DECISION TREE DI UTD
PMI KOTA SURABAYA

Pembimbing 1 : Amalia Anjani Arifiyanti, S.Kom., M.Kom.

Pembimbing 2 : Anindo Saka Fitri, S.Kom., M.Kom.

ABSTRAK

Darah merupakan komponen yang vital dalam tubuh manusia. Kurangnya jumlah darah pada tubuh akan mempengaruhi 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. Skripsi ini menggunakan beberapa algoritma *Decision Tree* dalam proses klasifikasi data. Algoritma yang digunakan adalah CART, C4.5, dan *Random Forest*. Masing-masing algoritma tersebut akan ditambahkan dengan metode *oversampling* SMOTE dan metode *Random Search CV*, sehingga terbentuk sembilan skenario.

Hasil perbandingan dari sembilan skenario menunjukkan bahwa algoritma *Random Forest* tanpa bantuan SMOTE dan *Random Search CV* memiliki nilai terbaik dibandingkan skenario lainnya. Algoritma *Random Forest* mendapatkan akurasi dengan nilai 97%, AUC ROC dengan nilai 99%, dan tidak terdeteksi *overfitting*. 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%.

Kata Kunci: Klasifikasi, Calon Pendonor Potensial, *Decision Tree*, CART, C4.5, *Random Forest*, UTD PMI Kota Surabaya

KATA PENGANTAR

Segala puji bagi Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul, “Klasifikasi Calon Pendonor Darah Potensial Menggunakan Algoritma *Decision Tree* di UTD PMI Kota Surabaya” sebagai salah satu persyaratan menyelesaikan Program Studi Sarjana Sistem Informasi di Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam penyusunan skripsi ini, penulis telah banyak menerima bimbingan, dukungan, dan bantuan dari berbagai pihak. Oleh karena itu, dengan tulus penulis ingin mengucapkan terima kasih kepada:

1. Umma Heffi Puspasari Ningrum yang tanpa henti memberikan doa, perhatian, dan keceriaan dalam keseharian penulis.
2. Ayah Didit, Adik Haydar, dan Adik Keinarra yang senantiasa memberikan doa dan dukungan dalam setiap langkah.
3. Partner penulis dalam bercerita dan berkeluh kesah, *special thanks to* Daffa Harits yang telah kebersamai penulis sedari SMP, terlebih dalam pengerjaan Skripsi ini, waktu, tenaga, pikiran, semua *full* untukku, *wes pokok'e nde best*.
4. Ibu Amalia Anjani Arifiyanti, S.Kom., M.Kom. selaku dosen wali sekaligus dosen pembimbing 1 yang telah memberikan bimbingan, motivasi, dan juga bantuan dalam masa perkuliahan penulis dari awal hingga Skripsi ini selesai.

5. Ibu Anindo Saka Fitri, S.Kom., M.Kom. selaku dosen pembimbing 2 yang telah memberikan bimbingan dan arahan dalam pengerjaan Skripsi ini hingga selesai.
6. Bapak Agung Brastama Putra, S.Kom., M.Kom. selaku Koordinator Program Studi Sistem Informasi UPN “Veteran” Jawa Timur yang telah memberikan bantuan dalam perkuliahan, khususnya perihal administrasi.
7. Seluruh dosen Program Studi Sistem Informasi UPN “Veteran” Jawa Timur yang dengan sabar memberikan ilmu-ilmu kepada penulis.
8. Kakak-kakak mentor Program Studi Independen *Data Analytics* Batch 4 di Zenius, terutama mentor penulis, Kak Dede Brahma Arianto yang telah memberikan banyak ilmu dan juga inspirasi kepada penulis untuk menyusun Skripsi ini dengan tema *Data Mining*.
9. Pihak UTD PMI Kota Surabaya yang telah mengizinkan penulis mengumpulkan data calon pendonor.
10. Teman-teman ATENSI angkatan 2020, khususnya teman-teman kelas D yang telah kebersamai dan memberi warna dari awal menjadi Maba Angkatan Covid hingga Skripsi ini dapat selesai.

Penulis menyadari masih terdapat kekurangan dalam penyusunan Skripsi ini. Oleh karena itu, penulis memohon maaf atas segala kekurangan yang terdapat dalam Skripsi ini. Besar harapan penulis, semoga Skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan ke depan, khususnya pada Sistem Informasi.

Surabaya, Juli 2024

Penulis

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iv
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan.....	5
1.5 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 DASAR TEORI.....	7
2.1.1 Calon Pendorong Darah Potensial	7
2.1.2 Data Mining	8
2.1.4 Klasifikasi pada Data Mining	10
2.1.5 Decision Tree.....	11
2.1.6 CART.....	11
2.1.7 C4.5.....	12
2.1.8 Random Forest.....	12
2.1.9 Area Under Curve (AUC) ROC.....	13
2.1.10 Confusion Matrix	14
2.1.11 Flask.....	15

2.2	Penelitian Terdahulu.....	16
2.2.1	Hasil Penelitian Irawan (2021)	16
2.2.2	Hasil Penelitian Wahono dan Riana (2020).....	16
2.2.3	Hasil Penelitian Astuti dkk. (2022).....	17
2.2.4	Hasil Penelitian Hasanah dkk. (2021).....	18
2.2.5	Hasil Penelitian Abubakar dkk. (2023).....	18
2.2.6	Hasil Penelitian Sunarya dan Haryanti (2022)	19
BAB III METODOLOGI PENELITIAN.....		21
3.1	Studi Literatur.....	21
3.2	Business Understanding	22
3.2.1	Analisis Kebutuhan.....	22
3.2.1.1	Kebutuhan Data.....	23
3.2.1.2	Kebutuhan Sistem	23
3.2.2	Pengumpulan Data.....	23
3.3	Data Understanding	24
3.3.1	Atribut.....	24
3.4	Data Preparation	26
3.5	Modelling	26
3.6	Evaluation.....	27
3.7	Deployment	28
3.7.1	Visualisasi Data	28
3.7.2	Pembangunan Sistem.....	28
3.8	Penyusunan Laporan	28
BAB IV HASIL DAN PEMBAHASAN		30

4.1	Data Understanding	30
4.1.1	Drop Column	31
4.1.2	Pengelompokkan Data	32
4.1.3	Missing Values Check	33
4.1.4	Outlier Check	33
4.1.5	Exploratory Data Analysis	35
4.2	Data Preparation	38
4.2.1	Handle Column	38
4.2.1.1	Kolom ‘pekerjaan’	38
4.2.1.2	Kolom ‘jenis’	39
4.2.1.3	Kolom ‘kelamin’, ‘goldar’, dan ‘rhesus’	40
4.2.1.4	Kolom ‘kecamatan’	41
4.2.2	Handle Missing Values	42
4.2.3	Handle Outliers	44
4.2.4	Categorical Data Encoding	46
4.3	Modelling	47
4.3.1	Train Test Split	48
4.3.2	Skenario 1 (CART)	48
4.3.3	Skenario 2 (C4.5)	50
4.3.4	Skenario 3 (Random Forest)	51
4.3.5	Skenario 4 (CART+SMOTE)	52
4.3.6	Skenario 5 (C4.5+SMOTE)	53
4.3.7	Skenario 6 (Random Forest+SMOTE)	54
4.3.8	Skenario 7 (CART+SMOTE+Random Search CV)	55

4.3.9 Skenario 8 (C4.5+SMOTE+Random Search CV).....	56
4.3.10 Skenario 9 (Random Forest+SMOTE+Random Search CV).....	57
4.4 Evaluation.....	58
4.5 Deployment	60
4.5.1 Dashboard	60
4.5.2 Pembuatan Website.....	65
4.5.3 Uji Validasi Sistem	68
BAB V PENUTUP.....	71
5.1 Kesimpulan.....	71
5.2 Saran	72
LAMPIRAN.....	73
Lampiran 1. Jadwal Penelitian	73
Lampiran 2. Surat Penelitian Tugas Akhir.....	74
DAFTAR PUSTAKA	75

DAFTAR GAMBAR

Gambar 2.1 Model CRISP-DM (Hasanah et al., 2021).....	9
Gambar 2.2 Confusion Matrix (Afifah, 2023)	14
Gambar 3.1 Metodologi Penelitian.....	21
Gambar 4.1 Informasi dari Dataset Calon Pendorong.....	30
Gambar 4.2 Kolom dengan Nilai Tidak Valid..	31
Gambar 4.3 Sampel Dataset Sementara.....	32
Gambar 4.4 Source Code Pengelompokkan Data.....	32
Gambar 4.5 Kolom dengan Missing Values.....	33
Gambar 4.6 Outlier pada Kolom ‘usia’	34
Gambar 4.7 Outlier pada Kolom ‘ht’	35
Gambar 4.8 Pie Chart Kolom ‘goldar’ berdasarkan Rhesus.....	35
Gambar 4.9 Bar Chart Kolom ‘tolak’ Berdasarkan Jenis Kelamin.....	36
Gambar 4.10 Histogram Kolom ‘usia’	37
Gambar 4.11 Count Plot Kolom ‘tolak’	37
Gambar 4.12 Daftar Nilai Kolom ‘pekerjaan’ Setelah Penyesuaian.....	39
Gambar 4.13 Daftar Nilai Kolom ‘jenis’ Sebelum Penyesuaian.....	39
Gambar 4.14 Daftar Nilai Kolom ‘jenis’ Setelah Penyesuaian.....	40
Gambar 4.15 Source Code Penyesuaian Kolom ‘goldar’	40
Gambar 4.16 Kolom ‘goldar’, ‘kelamin’, ‘rhesus’ Setelah Penyesuaian.....	41
Gambar 4.17 Daftar Nilai Kolom ‘kecamatan’ Sebelum Penyesuaian.....	41
Gambar 4.18 Daftar Nilai Kolom ‘kecamatan’ Setelah Penyesuaian.....	42
Gambar 4.19 Source Code Handle Missing Values Tahap 1.....	42
Gambar 4.20 Kolom Missing Values yang Tersisa.....	43
Gambar 4.21 Source Code Handle Missing Values Tahap 2.....	43
Gambar 4.22 Missing Values Berhasil Teratasi.....	44
Gambar 4.23 Source Code Handle Outliers.....	44
Gambar 4.24 Kolom ‘usia’ Setelah Handle Outliers.....	45
Gambar 4.25 Kolom ‘ht’ Setelah Handle Outliers.....	45
Gambar 4.26 Persentase Value Kolom ‘ht’	46
Gambar 4.27 Categorical Data Encoding.....	47
Gambar 4.28 Sampel Dataset Bersih.....	47

Gambar 4.29 Train Test Split.....	48
Gambar 4.30 Confusion Matrix dan Classification Report Skenario 1.....	49
Gambar 4.31 AUC ROC Skenario 1.....	49
Gambar 4.32 Confusion Matrix dan Classification Report Skenario 2.....	50
Gambar 4.33 AUC ROC Skenario 2.....	50
Gambar 4.34 Confusion Matrix dan Classification Report Skenario 3.....	51
Gambar 4.35 AUC ROC Skenario 3.....	51
Gambar 4.36 Confusion Matrix dan Classification Report Skenario 4.....	52
Gambar 4.37 AUC ROC Skenario 4.....	52
Gambar 4.38 Confusion Matrix dan Classification Report Skenario 5.....	53
Gambar 4.39 AUC ROC Skenario 5.....	53
Gambar 4.40 Confusion Matrix dan Classification Report Skenario 6.....	54
Gambar 4.41 AUC ROC Skenario 6.....	54
Gambar 4.42 Confusion Matrix dan Classification Report Skenario 7.....	55
Gambar 4.43 AUC ROC Skenario 7.....	55
Gambar 4.44 Confusion Matrix dan Classification Report Skenario 8.....	56
Gambar 4.45 AUC ROC Skenario 8.....	56
Gambar 4.46 Confusion Matrix dan Classification Report Skenario 9.....	57
Gambar 4.47 AUC ROC Skenario 9.....	57
Gambar 4.48 Perbandingan AUC ROC 9 Skenario.....	58
Gambar 4.49 Perbandingan Akurasi dan Tingkat Error 9 Skenario.....	59
Gambar 4.50 Dashboard Distribusi Calon Pendorong Darah.....	61
Gambar 4.51 Dashboard Distribusi Calon Pendorong Darah (2)	62
Gambar 4.52 Dashboard Perbandingan Algoritma.....	64
Gambar 4.53 Halaman Awal Web.....	65
Gambar 4.54 Halaman Input Form.....	66
Gambar 4.55 Halaman Hasil Prediksi dari Form.....	67
Gambar 4.56 Halaman Upload File.....	67
Gambar 4.57 Halaman Hasil Prediksi dari Upload File.....	68
Gambar 4.58 Pengujian Validasi Sistem.....	69

DAFTAR TABEL

Tabel 2.1 Kategori Nilai AUC (Qadrini et al., 2021).....	13
Tabel 3.1 Atribut Prediksi.....	25
Tabel 3.2 Atribut Visualisasi.....	26
Tabel 3.3 Skenario Pemodelan.....	27
Tabel 4.1 Handle Column ‘pekerjaan’	38
Tabel 4.2 Tabel Pengecekan Overfitting.....	60
Tabel 4.3 Sampel Perbandingan Hasil Prediksi.....	69