



SKRIPSI

KLASIFIKASI OBAT ANTI TUBERKULOSIS MENGUNAKAN ALGORITMA *CATEGORICAL BOOSTING* DENGAN OPTIMASI OPTUNA

YOSUA SATRIA BARA HARMONI
NPM 21083010029

DOSEN PEMBIMBING

Kartika Maulida Hindrayani, S.Kom., M.Kom.

Dr.Eng. Ir. Dwi Arman Prasetya.,S.T.,M.T.,IPU., Asean. Eng.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

KLASIFIKASI OBAT ANTI TUBERKULOSIS MENGUNAKAN ALGORITMA *CATEGORICAL BOOSTING* DENGAN OPTIMASI OPTUNA

YOSUA SATRIA BARA HARMONI
NPM 21083010029

DOSEN PEMBIMBING

Kartika Maulida Hindrayani, S.Kom., M.Kom.
Dr.Eng. Ir. Dwi Arman Prasetya.,S.T.,M.T.,IPU., Asean. Eng.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

KLASIFIKASI OBAT ANTI TUBERKULOSIS MENGUNAKAN ALGORITMA *CATEGORICAL BOOSTING* DENGAN OPTIMASI OPTUNA

YOSUA SATRIA BARA HARMONI
NPM 21083010029

DOSEN PEMBIMBING

Kartika Maulida Hindrayani, S.Kom., M.Kom.

Dr.Eng. Ir. Dwi Arman Prasetya.,S.T.,M.T.,IPU., Asean. Eng.

**KEMENTERIAN PENDIDIKAN, TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**

LEMBAR PENGESAHAN

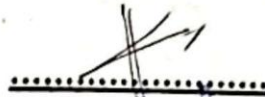
KLASIFIKASI OBAT ANTI TUBERKULOSIS MENGGUNAKAN ALGORITMA CATEGORICAL BOOSTING DENGAN OPTIMASI OPTUNA

Oleh:
YOSUA SATRIA BARA HARMONI
NPM. 21083010029

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 4 Juni 2025:

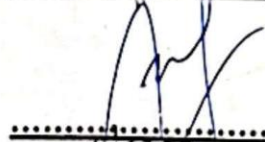
Menyetujui,

Kartika Maulida Hindrayani, S.Kom., M.Kom.
NIP. 199209092022032009



(Pembimbing I)

Dr.Eng. Ir. Dwi Arman Prasetya, S.T., M.T.,
IPU., Asean. Eng.
NIP. 198012052005011002



(Pembimbing II)

Amri Muhaimin, S.Stat., M.Stat., M.S.
NIP. 199507232024061002



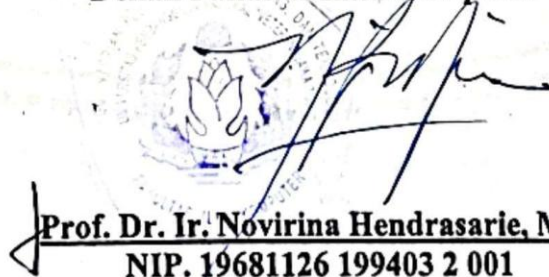
(Ketua Penguji)

Andri Fauzan Adziima S.Si, M.Si.
NIP. 199502122024061001



(Penguji I)

Mengetahui,
Dekan Fakultas Ilmu Komputer



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

LEMBAR PERSETUJUAN

KLASIFIKASI OBAT ANTI TUBERKULOSIS MENGGUNAKAN ALGORITMA CATEGORICAL BOOSTING DENGAN OPTIMASI OPTUNA

Oleh:

YOSUA SATRIA BARA HARMONI
NPM. 21083010029

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng.
NIP. 198012052005011002

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertandatangan di bawah ini:

Nama Mahasiswa : Yosua Satria Bara Harmoni
NPM : 21083010029
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 26 Mei 2025
Yang Membuat Pernyataan,



YOSUA SATRIA BARA HARMONI
NPM. 21083010029

ABSTRAK

Nama Mahasiswa / NPM : Yosua Satria Bara Harmoni / 21083010029
Judul Skripsi : Klasifikasi Obat Anti Tuberkulosis Menggunakan
Algoritma *Categorical Boosting* dengan Optimasi
Optuna
Dosen Pembimbing : 1. Kartika Maulida Hindrayani, S.Kom., M.Kom.
2. Dr. Eng. Ir. Dwi Arman Prasetya., S.T., M.T., IPU.,
Asean. Eng.

Penyakit tuberkulosis merupakan salah satu penyebab utama kematian global, dengan angka kematian mencapai 1,30 juta jiwa pada tahun 2022, meningkat sebesar 3,2% dibandingkan tahun sebelumnya. Indonesia menjadi salah satu negara dengan jumlah kasus tuberkulosis tertinggi di dunia. Pendekatan *Directly Observed Treatment Short-course* (DOTS) berperan dalam meningkatkan efektivitas terapi tuberkulosis dengan memastikan ketersediaan Obat Anti Tuberkulosis yang tepat. Namun, kesalahan dalam pemilihan obat dapat memicu kegagalan terapi, kambuhan, dan kasus *Multi-Drug Resistant* (MDR). Untuk mengatasi hal ini, model klasifikasi berbasis data rekam medis pasien dapat digunakan guna meningkatkan ketepatan pemilihan obat. Penelitian ini berfokus pada pengembangan model klasifikasi untuk menentukan jenis obat menggunakan algoritma *Categorical Boosting* (CatBoost) yang dioptimalkan dengan Optuna menggunakan pendekatan *Tree-structured Parzen Estimator* (TPE). Data penelitian terdiri dari variabel numerik, seperti usia dan durasi pengobatan, serta variabel kategorik, seperti riwayat diabetes melitus, status HIV, dan kombinasi obat. Algoritma CatBoost dipilih karena kemampuannya dalam menangani data kategorik dengan baik. Optimasi hyperparameter dilakukan menggunakan Optuna guna memperoleh parameter terbaik untuk CatBoost. Tahapan preprocessing melibatkan penghapusan outlier, normalisasi fitur, serta reduksi dimensi pada 620 sampel data, yang kemudian dibagi menjadi 90% data latih dan 10% data uji. Hasil eksperimen menunjukkan bahwa model CatBoost menghasilkan akurasi awal sebesar 90%. Setelah menerapkan teknik oversampling dengan SMOTE dan optimasi parameter menggunakan Optuna, akurasi meningkat menjadi 96%, menunjukkan peningkatan sebesar 6%. Model ini mampu mengklasifikasikan jenis obat seperti isoniazid, rifampisin, dan pirazinamid secara akurat, sehingga dapat mendukung pemilihan terapi yang lebih efektif bagi pasien tuberkulosis.

Kata kunci : *Categorical Boosting*, Tuberkulosis, Obat Anti Tuberkulosis, Optuna.

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Yosua Satria Bara Harmoni / 21083010029
Thesis Title : *Classification of Anti-Tuberculosis Drugs Using Categorical Boosting algorithm with Optuna Optimization*
Advisor : 1. Kartika Maulida Hindrayani, S.Kom., M.Kom.
2. Dr. Eng. Ir. Dwi Arman Prasetya.,S.T.,M.T.,IPU., Asean. Eng.

ABSTRACT

Tuberculosis is one of the leading causes of death globally, with death rate reaching 1.30 million by 2022, an increase of 3.2% compared to the previous year. Indonesia is one of the countries with the highest number of tuberculosis cases in the world. The Directly Observed Treatment Short-course (DOTS) plays a role in improving the effectiveness of tuberculosis therapy by ensuring the availability of appropriate antituberculosis drugs. However, errors in drug selection can lead to therapy failure, relapses, and Multi-Drug Resistant (MDR) cases. To overcome this, classification models based on patient medical record data can be used to improve the accuracy of drug selection. This research focuses on developing classification model to determine the type of drug using Categorical Boosting algorithm optimized with Optuna using Tree-structured Parzen Estimator. The data consisted of numerical variables, such as age, treatment duration, and categorical variables, such as history of diabetes, HIV status, drug combination. The CatBoost algorithm was chosen due to its ability to handle categorical data. Hyperparameter optimization was performed to obtain the best parameters. The preprocessing stage involved memory reduction, feature normalization, and encoding on 620 data samples, which were then divided into 90% training and 10% test data. Experimental results show CatBoost model produces an initial accuracy of 90%. After applying oversampling with SMOTE and parameter optimization techniques using Optuna, the accuracy increased to 96%, showing 6% improvement. The model can accurately classify drugs combination, which can support the selection of more effective therapies for tuberculosis patients.

Keywords: *Anti-Tuberculosis Drugs, Categorical Boosting, Tuberculosis, Optuna.*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Segala syukur dan puji hanya bagi Tuhan Yesus Kristus, oleh karena anugerah-Nya yang melimpah, kemurahan dan kasih setia yang besar kepada penulis sehingga skripsi dengan judul **“Klasifikasi Obat Anti Tuberkulosis menggunakan Algoritma *Categorical Boosting* dengan Optimasi Optuna”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Ibu Kartika Maulida Hindrayani, S.Kom., M.Kom selaku Dosen Pembimbing utama yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Dan penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Kedua orang tua, keluarga, dan teman-teman yang senantiasa memberikan dorongan dan doa.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur
4. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur dan Pembimbing 2.
5. Bapak Trimono S.Si., M.Si., selaku koordinator skripsi di program studi sains data.
6. Ibu Kartika Maulida Hindrayani, S.Kom., M.Kom. selaku Dosen Wali dan Dosen Pembimbing 1.
7. Bapak dan Ibu dosen Program Studi Sains Data UPN “Veteran” Jawa Timur yang sudah berkenan untuk memberikan waktu untuk berkontribusi dalam penyelesaian skripsi.
8. Pihak jajaran Puskesmas Gending dan Rumah Sakit Umum Daerah Dr. Iskak Tulungagung yang telah membantu dalam pengurusan data skripsi yang dibutuhkan.

9. *Last but not least, I wanna thank me. I wanna thank me for believing in me. I wanna thank me for all doing this hard work. I wanna thank me for having no days off. I wanna thank me for never quitting. I wanna thank me for just being me at all times.*

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 16 Juni 2025

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN SKRIPSI	ii
LEMBAR PERSETUJUAN SKRIPSI.....	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
DAFTAR NOTASI.....	xxi
BAB I. PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	6
1.4. Tujuan Penelitian.....	6
1.5. Manfaat Penelitian.....	6
BAB II. TINJAUAN PUSTAKA.....	9
2.1. Penelitian Terdahulu	9
2.2. Dasar Teori.....	15
2.2.1. Farmakoterapi	15
2.2.2. <i>Clinical Decision Support System</i>	18
2.2.3. Tuberkulosis.....	20
2.2.4. Pengobatan Tuberkulosis	22
2.2.5. <i>Machine Learning</i>	26
2.2.6. Klasifikasi	28
2.2.7. <i>Categorical Boosting</i> (CatBoost).....	30
2.2.8. Optuna	37
2.2.9. <i>Synthetic Minority Oversampling Technique</i> (SMOTE)	40

2.2.10. <i>Confusion Matrix</i>	42
2.2.11. Analisis ROC (<i>Receiver Operating Characteristic</i>)	44
2.2.12. Streamlit	47
BAB III. DESAIN DAN IMPLEMENTASI SISTEM	49
3.1. Variabel Penelitian dan Sumber Data	49
3.2. Langkah Analisis.....	51
3.3. Desain Sistem.....	63
BAB IV. HASIL DAN PEMBAHASAN	65
4.1. Pengumpulan Data Pasien Tuberkulosis.....	65
4.2. Persiapan Dataset	66
4.3. Pembagian Data	78
4.4. Penyeimbangan Data.....	79
4.5. Klasifikasi Obat Anti Tuberkulosis dengan Model CatBoost.....	81
4.6. Klasifikasi Model CatBoost dengan Optimalisasi Optuna	88
4.7. Perbandingan Model CatBoost dan CatBoost – Optuna.....	97
4.8. Pemilihan Model Terbaik.....	99
4.9. Implikasi Klinis.....	103
4.10. Deployment Aplikasi	106
BAB V. PENUTUP	113
5.1. Kesimpulan.....	113
5.2. Saran Pengembangan	114
DAFTAR PUSTAKA	115
LAMPIRAN	125

DAFTAR GAMBAR

Gambar 2.1 Farmakoterapi	16
Gambar 2.2 Interaksi Tubuh – Obat	17
Gambar 2.3 CDSS	18
Gambar 2.4 Penyakit Tuberkulosis	20
Gambar 2.5 Gejala Penyakit Tuberkulosis	21
Gambar 2.6 Obat Anti Tuberkulosis.....	23
Gambar 2.7 Cara kerja Obat Anti Tuberkulosis	24
Gambar 2.8 <i>Machine learning</i>	26
Gambar 2.9 Klasifikasi	28
Gambar 2.10 CatBoost	31
Gambar 2.11 CatBoost <i>architecture</i>	31
Gambar 2.12 Optuna	37
Gambar 2.13 Optuna <i>Architecture</i>	38
Gambar 2.14 SMOTE.....	40
Gambar 2.15 <i>confusion matrix</i>	42
Gambar 2.16 Metrik – metrik klasifikasi.....	44
Gambar 2.17 Grafik ROC.....	45
Gambar 2.18 Kurva ROC	46
Gambar 2.19 AUC	47
Gambar 2.20 <i>Streamlit</i>	47
Gambar 2.21 <i>Streamlit use case</i>	48
Gambar 3.1 Diagram alir penelitian	52
Gambar 3.2 Pengumpulan Data dan Studi Literatur.....	53
Gambar 3.3 <i>Preprocessing</i>	54
Gambar 3.4 Penyeimbangan kelas.....	56
Gambar 3.5 Pemodelan dan Optimasi Parameter	58
Gambar 3.6 Interpretasi model	63
Gambar 3.7 Desain sistem	64
Gambar 4.1 Distribusi Tipe Obat	67
Gambar 4.2 Presentase Paduan Pengobatan pada Pasien	68
Gambar 4.3 Distribusi Klasifikasi Anatomi	69

Gambar 4.4 Distribusi Jenis Kelamin Pasien	70
Gambar 4.5 Jenis Kelamin dengan Tipe Obat.....	70
Gambar 4.6 Distribusi Usia	71
Gambar 4.7 Distribusi Lama Pengobatan Berdasarkan Kelompok Umur.....	72
Gambar 4.8 Distribusi Pasien Berdasarkan Usia dan Regimen Obat.....	73
Gambar 4.9 Distribusi Usia Berdasarkan beberapa faktor	74
Gambar 4.10 <i>Missing Value</i>	75
Gambar 4.11 Perbandingan Distribusi Kelas	81
Gambar 4.12 Split Data 70:30 non SMOTE	83
Gambar 4.13 Split Data 70:30 SMOTE	83
Gambar 4.14 Split Data 80:20 Non SMOTE.....	84
Gambar 4.15 Split Data 80:20 SMOTE	85
Gambar 4.16 Split Data 90:10 Non SMOTE.....	85
Gambar 4.17 Split Data 90:10 SMOTE	86
Gambar 4.18 <i>Feature Importance</i> CatBoost SMOTE.....	88
Gambar 4.19 Arsitektur Optuna	90
Gambar 4.20 Split Data 70:30 dengan optimasi.....	91
Gambar 4.21 Split Data 80:20 dengan optimasi.....	92
Gambar 4.22 Split Data 90:10 dengan optimasi.....	93
Gambar 4.23 <i>Feature Importance</i> setelah dioptimasi	96
Gambar 4.24 Evaluasi Performa Split Data 70:30	97
Gambar 4.25 Evaluasi Performa Split Data 80:20	98
Gambar 4.26 Evaluasi Performa Split Data 90:10	98
Gambar 4.27 <i>Classification Report</i> Model Terbaik	99
Gambar 4.28 Evaluasi Model (K = 5) Model Terbaik	100
Gambar 4.29 ROC Model Terbaik	101
Gambar 4.30 <i>Feature Importance</i> Model Terbaik	103
Gambar 4.31 Skema Pemberian Terapi.....	104
Gambar 4.32 CDSS	105
Gambar 4.33 <i>User Interface</i>	107
Gambar 4.34 <i>Dashboard</i> Prediksi	107
Gambar 4.35 Input Data Demografi & Kontak	108

Gambar 4.36 Input Diagnosis & Klasifikasi TBC.....	108
Gambar 4.37 Input Riwayat Medis & Pengobatan.....	109
Gambar 4.38 <i>Input</i> Komorbid	110
Gambar 4.39 Akurasi pada <i>Deployment</i>	110
Gambar 4.40 Fitur Prediksi <i>File</i>	111
Gambar 4.41 Hasil Prediksi dari Excel	111
Gambar 4.42 <i>Download</i> Hasil Prediksi	112

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu	9
Tabel 2.2 Algoritma 1	34
Tabel 2.3 Algoritma 2	35
Tabel 2.4 Tabel AUC	46
Tabel 3.1 Tabel variabel data	49
Tabel 4.1 Dataset Pasien Tuberkulosis.....	65
Tabel 4.2 Transformasi Data	76
Tabel 4.3 Data setelah <i>Label Encoding</i>	78
Tabel 4.4 Pembagian Data.....	78
Tabel 4.5 Jumlah Sampel <i>Imbalanced Ratio</i>	80
Tabel 4.6 Perbandingan Performa Model.....	87
Tabel 4.7 Perbandingan Performa Model Setelah Optimasi	93
Tabel 4.8 <i>Setting Parameter</i>	94
Tabel 4.9 Prediksi Data Aktual dan Prediksi.....	95
Tabel 4.10 Parameter Terbaik	101

Halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

Lampiran 1. LOA Publikasi Jurnal	125
Lampiran 2. Surat Pengajuan Data.....	126
Lampiran 3. Surat Izin Penelitian.....	127
Lampiran 4. Surat Keterangan Etik Lolos Penelitian	128
Lampiran 5. Lampiran Hasil Penelitian di RSUD Dr. Iskak.....	129
Lampiran 6. Kode Program Pemodelan	130
Lampiran 7. Kode Program Aplikasi	131

Halaman ini sengaja dikosongkan

DAFTAR NOTASI

1. $l(\theta)$: *Likelihood* fungsi untuk parameter model θ , dengan kondisi bahwa data y lebih kecil dari nilai a , yang diberikan oleh hipotesis H .
2. $z(\theta)$: Probabilitas bahwa y lebih kecil dari a , dengan kondisi a dan hipotesis H .
3. A : Nilai minimum dari sekumpulan data $\{y_1, y_2, \dots, y_n\}$.
4. H : Himpunan data yang terdiri dari pasangan $(\theta_1, y_1), (\theta_2, y_2), \dots, (\theta_n, y_n)$.
5. $ctri_i$: Rasio kelas mayoritas yang dikalibrasi dengan *prior* dan *total count* untuk menghindari bias terhadap kelas minoritas.
6. $Fm(X)$: *Output* dari model m pada data X , yang diperoleh melalui penambahan hasil sebelumnya dan pembaruan berdasarkan pembelajaran (*learning rate* dan *residual*).
7. $\%$: Persentase perbandingan antara jumlah data kelas mayoritas dengan kelas minoritas.
8. T : Vektor data masukan, yang terdiri dari nn elemen $\{x_1, x_2, \dots, x_n\}$.
9. Y_T : Vektor label data T , yang terdiri dari nn elemen $\{y_1, y_2, \dots, y_n\}$.
10. $d(x, y)$: Jarak *Euclidean* antara dua titik x dan y .
11. x_{syn} : Data sintetis yang dihasilkan oleh SMOTE, yang dihitung berdasarkan data yang ada dan data tetangga terdekat.
12. x_i : Data yang akan direplikasi menggunakan SMOTE.
13. x_{knn} : Data yang memiliki jarak terdekat dengan data x_i .
14. δ : Bilangan acak antara 0 dan 1 yang digunakan untuk menentukan seberapa jauh data sintetis $x_{\{syn\}}$ diposisikan antara x_i dan $x_{\{knn\}}$.

Halaman ini sengaja dikosongkan

