

**PERBANDINGAN METODE KNN DAN RANDOM FOREST UNTUK
KLASIFIKASI PENYAKIT TUBERKULOSIS BERDASARKAN GEJALA
KLINIS PASIEN**

TUGAS AKHIR

No. 1266/WM.FT.H6/T.ILKOM/TA/2025

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Komputer**



Disusun oleh:

CORY ALVANIA SELAN

23122037

PROGRAM STUDI ILMU KOMPUTER

FAKULTAS TEKNIK

UNIVERSITAS KATOLIK WIDYA MANDIRA

KUPANG

2026

HALAMAN PERSETUJUAN

TUGAS AKHIR

No. 1266/WM.FT.H6/T.ILKOM/TA/2025

PERBANDINGAN METODE KNN DAN RANDOM FOREST UNTUK KLASIFIKASI PENYAKIT TUBERKULOSIS BERDASARKAN GEJALA KLINIS PASIEN

Oleh :

CORY ALVANIA SELAN

NIM : 23122037

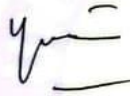
TELAH DIPERIKSA DAN DIPERTAHANKAN DIDEPAN PENGUJI :

DI : KUPANG

PADA TANGGAL : JANUARI 2026

DOSEN PENGUJI I

DOSEN PENGUJI II



Yulianti Paula Bria, S.T., M.T., Ph.D

NIDN : 0823078702



Yovina Carmeneja Hoar Siki, S.T., M.T

NIDN : 0805058803

DOSEN PENGUJI III



Alfry Aristo Jansen Sinlae, S.Kom., M.Cs

NIDN : 0807078704

KETUA PELAKSANA

SEKRETARIS PELAKSANA



Alfry Aristo Jansen Sinlae, S.Kom., M.Cs

NIDN : 0807078704



Paul Filson M. Tengangatu, S.Kom., M.T.I

NUPTK : 4955769670130332

HALAMAN PENGESAHAN

TUGAS AKHIR

No. 1266/WM.FT.II6/T.ILKOM/TA/2025

PERBANDINGAN METODE KNN DAN RANDOM FOREST UNTUK KLASIFIKASI PENYAKIT TUBERKULOSIS BERDASARKAN GEJALA KLINIS PASIEN

Oleh :

CORY ALVANIA SELAN

NIM : 23122037

TELAH DIPERIKSA DIDEPAN PEMBIMBING:

DOSEN PEMBIMBING I



Alfry Aristo Jansen Sinlae, S.Kom., M.Cs

NIDN : 0807078704

DOSEN PEMBIMBING II



Paul Filson M. Tengangatu, S.Kom., M.T.I

NUPTK : 4955769670130332

MENGETAHUI,
KETUA PROGRAM STUDI ILMU
KOMPUTER UNIKA WIDYA
MANDIRA



Yovina Carmencia Hoar Siki, S.T., M.T

NIDN : 0805058803

MENGETAHUI,
DEKAN FAKULTAS TEKNIK
UNIKA WIDYA MANDIRA



Dr. Don Gema N. Da Costa, S.T., M.T

NIDN : 0820036801

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur, haru, tawa, tangis, dan segala warna emosi yang telah menemani perjalanan ini, Tugas Akhir ini penulis persembahkan kepada:

1. Tuhan Yang Maha Esa, Sang Pemilik Hidup. Terima kasih telah memegang tangan penulis saat hampir menyerah, memberikan cahaya di tengah gelapnya keputusasaan, dan membuktikan bahwa bersama-Mu, tidak ada yang mustahil.
2. Kedua Orang Tua Tercinta, alasan utama saya tidak pernah menyerah. Terima kasih telah menjadi sandaran paling kokoh dan pendoa paling setia. Keberhasilan ini tidak ada artinya tanpa doa kalian. ini adalah kado kecil untuk membalas peluh dan pengorbanan yang tak mungkin mampu saya bayar seumur hidup saya.
3. Kaka Lens dan Adik Gebi, penguat hati yang selalu memberikan alasan bagi penulis untuk terus melangkah. Terima kasih telah menjadi rumah yang hangat untuk pulang.
4. Dosen Pembimbing I, Bapak Alfry Aristo Jansen SinlaE, S.Kom., M.Cs, terima kasih atas kasih sayang layaknya orang tua yang tidak hanya membimbing akademik saya, tetapi juga mempercayai kemampuan saya saat saya sendiri meragukannya. Untuk Dosen Pembimbing II, Bapak Paul Filson M. Tengangatu, S.Kom., M.T.I, terima kasih telah menjadi cahaya penunjuk dengan kesabaran yang tak terhingga, membimbing saya hingga saya mampu berdiri tegak menyelesaikan tugas akhir ini.
5. Para Dosen dan Tenaga Pendidik, terima kasih atas warisan ilmu dan didikan yang akan penulis jaga selamanya sebagai bekal kehidupan.
6. Keluarga Selan dan Nomleni, terima kasih atas doa tulus yang terus mengalir dari kejauhan, menjadi energi tambahan bagi saya.

7. Persekutuan Doa Sehati, terima kasih telah menjadi wadah yang menguatkan iman dan menjadi tempat bertumbuh bersama dalam doa selama masa-masa sulit perkuliahan.
8. *Sahabat 3,5 Tahun* (Inda, Eka, Elga, dan Prily), kita bukan sekadar teman yang bertemu di ruang kelas, melainkan keluarga yang lahir dari perjuangan. Terima kasih telah menjadi rumah saat saya merasa asing, menjadi tangan yang merangkul saat saya jatuh, dan menjadi suara yang meyakinkan saat saya hampir menyerah pada keadaan. Kita telah melewati badai bersama, dan berdiri di garis *finish* ini bersama kalian adalah bagian terindah sekaligus paling mengharukan dalam hidup saya.
9. *Keluarga Kejar*, yang telah mewarnai hari-hari saya. Terima kasih telah menjadi tempat berbagi tawa konyol yang meringankan beban, membuktikan bahwa persahabatan sejati adalah *support system* terbaik.
10. *Teman-teman Ilkom A 22 dan Angkatan 2022*, rekan seperjuangan yang telah berproses bersama dari nol. Terima kasih atas setiap bantuan dan kebersamaan yang takkan terlupakan.
11. Semua pihak yang tidak dapat saya sebutkan satu per satu.

MOTTO

"Takut akan Tuhan adalah awal dari segala pengetahuan, kejujuran adalah jalannya, dan keyakinan bahwa setiap persoalan memiliki solusi adalah kekuatannya."

PERNYATAAN DAN KEASLIAN KARYA

Yang bertanda tangan di bawah ini:

Nama : Cory Alvania Selan

NIM : 23122037

Fakultas/Prodi : Teknik/Ilmu Komputer

Dengan ini, saya menyatakan bahwa Tugas Akhir ini yang berjudul “Perbandingan Metode KNN dan *Random Forest* untuk Klasifikasi Penyakit *Tuberkulosis* berdasarkan Gejala Klinis Pasien” adalah hasil karya saya sendiri. Karya ini disusun tanpa melakukan plagiarisme dan sepenuhnya berdasarkan penelitian, analisis, serta pemikiran yang telah saya lakukan. Apabila di kemudian hari, terbukti bahwa karya ini merupakan hasil plagiarisme atau terdapat pelanggaran akademik lainnya, saya bersedia menerima sanksi sesuai dengan peraturan yang berlaku di Universitas Katolik Widya Mandira Kupang. Demikian pernyataan ini saya buat dengan sebenar-benarnya dan penuh tanggung jawab.

Mengetahui,
Pembimbing



Alfry Aristo Jansen Sinlae, S.Kom., M.Cs

NIDN : 0807078704

Kupang, Januari 2026
Mahasiswa



Cory Alvania Selan

NIM : 23122037

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa karena atas berkat, rahmat, dan karunia-Nya yang besar, Tugas Akhir penelitian berjudul Perbandingan Metode KNN dan *Random Forest* Untuk Klasifikasi Penyakit *Tuberculosis* Berdasarkan Gejala Klinis Pasien ini dapat diselesaikan. Tugas Akhir ini disusun untuk diajukan sebagai salah satu syarat kelulusan pada Program Strata-1 di Program Studi Ilmu Komputer, Fakultas Teknik, Universitas Katolik Widya Mandira Kupang.

Penulis menyadari bahwa penyelesaian penyusunan Tugas Akhir ini tidak lepas dari bantuan beberapa pihak. Oleh karena itu, pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada:

1. Pater Dr. Stefanus Lio, SVD, S.Fil., MA, selaku Rektor Universitas Katolik Widya Mandira Kupang.
2. Bapak Dr. Don Gaspar N. Da Costa, S.T., M.T, selaku Dekan Fakultas Teknik Universitas Katolik Widya Mandira.
3. Ibu Yovinia Carmeneja Hoar Siki, S.T., M.T, selaku Ketua Program Studi Ilmu Komputer Universitas Katolik Widya Mandira.
4. Kedua dosen pembimbing, Bapak Alfry Aristo Jansen SinlaE, S.Kom., M.Cs, selaku Pembimbing I dan Bapak Paul Filson M. Tengangatu, S.Kom., M.T.I, sebagai Pembimbing II, atas bimbingan, arahan, kritik, dan saran yang diberikan selama proses penyusunan Tugas Akhir ini. Tanpa bimbingan yang tulus dan penuh kesabaran, Tugas Akhir ini tidak mungkin dapat terselesaikan.
5. Kedua dosen penguji Ibu Yulianti Paula Bria, S.T., M.T., Ph.D, selaku Penguji I dan Ibu Yovinia Carmeneja Hoar Siki, S.T., M.T., selaku penguji II.
6. Bapak Donatus Joseph Manehat, S.Si, M.Kom sebagai dosen Pembimbing Akademik yang telah memberikan arahan dan dukungan dari awal semester hingga penyelesaian Tugas Akhir ini.
7. Bapak Juventinho Jose Maria De Araujo, S.Kom., selaku Laboran di Program Studi Ilmu Komputer yang selalu mendukung.

8. Seluruh Dosen dan Pegawai Program Studi Ilmu Komputer di Universitas Katolik Widya Mandira Kupang.
9. Kedua Orang Tua tercinta, yang telah menjadi sumber kekuatan utama bagi penulis. Terima kasih atas doa yang tidak pernah putus, kasih sayang yang luar biasa, serta pengorbanan baik secara moril maupun materil sehingga penulis dapat menyelesaikan studi hingga tahap ini. Tugas Akhir ini penulis persembahkan spesial untuk kalian.
10. Teman-teman Kelas Ilmu Komputer A Angkatan 2022, terima kasih atas kebersamaan, bantuan, dan semangat yang kita bagi selama masa perkuliahan. Suka dan duka yang kita lalui bersama menjadi kenangan yang sangat berharga bagi penulis.
11. Seluruh Teman-teman Angkatan 2022 Program Studi Ilmu Komputer, terima kasih atas dukungan dan solidaritasnya. Semoga kesuksesan menyertai kita semua di masa depan.

Penulis menyadari bahwa penyusunan Tugas Akhir ini tidak luput dari kesalahan. Untuk itu, penulis mengharapkan saran dan kritik dari berbagai pihak sehingga Tugas Akhir ini dapat memberikan manfaat bagi semuanya.

Kupang, Januari 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN.....	i
HALAMAN PENGESAHAN.....	i
HALAMAN PERSEMBAHAN.....	iii
MOTTO	v
PERNYATAAN DAN KEASLIAN KARYA.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
ABSTRAK	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan.....	5
1.7 Daftar Istilah.....	7
BAB II LANDASAN TEORI	12
2.1 Penelitian Terdahulu	12
2.2 Landasan Teori.....	24
2.2.1 <i>Tuberkulosis (TBC)</i>	24

2.2.2	<i>Machine Learning</i>	25
2.2.3	Model <i>Machine Learning</i>	26
2.2.4	<i>Jupyter Notebook</i>	27
BAB III METODOLOGI PENELITIAN		29
3.1	Pengumpulan Data	30
3.2	<i>Pre-processing</i> Data.....	31
3.2.1	<i>Cleaning</i> Data	31
3.2.2	Transformasi Data.....	33
3.2.3	Normalisasi Data.....	35
3.2.4	Seleksi Fitur	39
3.2.5	<i>KFold Cross Validation</i>	47
3.2.6	Penerapan Model Klasifikasi.....	48
3.3	Evaluasi Kinerja Model.....	53
3.4	Pengujian Model Klasifikasi.....	55
BAB IV HASIL DAN PEMBAHASAN.....		56
4.1	Pengumpulan Data.....	56
4.2	<i>Preprocessing</i> Data	60
4.2.1	<i>Cleaning</i>	64
4.2.2	Transformasi (<i>Encoding</i>).....	67
4.2.3	Normalisasi (<i>Scaling</i>)	83
4.2.4	Penanganan <i>Imbalance</i> Data (SMOTE)	85
4.2.5	Seleksi Fitur	87
4.2.6	Penerapan Model dan Evaluasi Model	93
BAB V PENGUJIAN DAN ANALISIS HASIL		110
5.1	Pengujian.....	110

5.2 Analisis Hasil	112
BAB VI PENUTUP	115
6.1 Kesimpulan.....	115
6.2 Saran.....	116
DAFTAR PUSTAKA.....	117

DAFTAR TABEL

Tabel 1. 1 Daftar Istilah dan Singkatan.....	7
Tabel 2. 1 Daftar Penelitian Terdahulu	17
Tabel 3. 1 Atribut dengan <i>Missing Value</i>	31
Tabel 3. 2 Ilustrasi <i>dataset tuberkulosis</i>	35
Tabel 3. 3 Perhitungan Normalisasi <i>Z-Score</i>	37
Tabel 3. 4 Ilustrasi <i>dataset tuberkulosis</i> (hasil transformasi numerik)	42
Tabel 3. 5 Ilustrasi <i>dataset tuberkulosis</i> ternormalisasi	43
Tabel 3. 6 Ilustrasi <i>dataset tuberkulosis</i> sederhana.....	48
Tabel 3. 7 <i>Dataset</i> hasil normalisasi	49
Tabel 3. 8 <i>Confusion matrix</i>	54
Tabel 4. 1 Deskripsi Atribut Data Awal.....	57
Tabel 4. 2 Keterangan Atribut yang dihapus	60
Tabel 4. 3 Karakteristik Data	62
Tabel 4. 4 Deskripsi Kategori Atribut Klinis	63
Tabel 4. 5 Ringkasan Distribusi Gejala.....	75
Tabel 4. 6 Hasil Evaluasi Performa Model Skenario	99

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi <i>Random Forest</i>	27
Gambar 3. 1 Tahapan Penelitian	29
Gambar 3. 2 Dokumentasi Pengumpulan Data	30
Gambar 3. 3 Pohon keputusan	52
Gambar 4. 1 Implementasi <i>Source code</i> muat data CSV dengan <i>pandas</i>	60
Gambar 4. 2 Implementasi <i>source code</i> menghapus 15 atribut tidak relevan	65
Gambar 4. 3 Implementasi <i>source code</i> dan hasil identifikasi nilai hilang	66
Gambar 4. 4 Implementasi <i>source code</i> pengisian data kosong dengan modus	66
Gambar 4. 5 Hasil penanganan nilai hilang	67
Gambar 4. 6 Implementasi <i>source code</i> transformasi <i>ordinal encoding</i>	68
Gambar 4. 7 Implementasi <i>source code</i> transformasi atribut biner dan target	69
Gambar 4. 8 Tampilan <i>dataset</i> setelah transformasi keseluruhan.....	70
Gambar 4. 9 Ringkasan statistik atribut numerik (bagian 1)	70
Gambar 4. 10 Ringkasan statistik atribut numerik (bagian 2)	71
Gambar 4. 11 Implementasi <i>source code</i> pengecekan nilai unik atribut	72
Gambar 4. 12 Hasil pengecekan nilai unik atribut	73
Gambar 4. 13 Implementasi <i>source code</i> visualisasi distribusi gejala.....	74
Gambar 4. 14 Implementasi <i>source code</i> normalisasi dengan <i>z-score</i>	85
Gambar 4. 15 Implementasi normalisasi <i>z-score</i>	85
Gambar 4. 16 Visualisasi Sebelum SMOTE.....	86
Gambar 4. 17 Implementasi <i>source code</i> penggunaan SMOTE	87
Gambar 4. 18 Hasil penyeimbangan data menggunakan SMOTE	87
Gambar 4. 19 Implementasi <i>source code</i> 10 fitur <i>Stratified K-Fold 5-fold</i>	89
Gambar 4. 20 Implementasi <i>source code</i> 15 fitur <i>Stratified K-Fold 5-fold</i>	89
Gambar 4. 21 Implementasi <i>source code</i> 10 fitur <i>Stratified K-Fold 10-fold</i>	90
Gambar 4. 22 Implementasi <i>source code</i> 15 fitur <i>Stratified K-Fold 10-fold</i>	90
Gambar 4. 23 Hasil Peringkat fitur untuk 10 fitur dan 5 <i>fold</i>	91
Gambar 4. 24 Hasil Peringkat fitur untuk 15 fitur dan 5 <i>fold</i>	91
Gambar 4. 25 Hasil Peringkat fitur untuk 10 fitur dan 10 <i>fold</i>	91
Gambar 4. 26 Hasil Peringkat fitur untuk 15 fitur dan 10 <i>fold</i>	92

Gambar 4. 27 Implementasi <i>source code</i> 10 Fitur dan 5 <i>fold</i>	95
Gambar 4. 28 Implementasi <i>source code</i> 15 Fitur dan 5 <i>fold</i>	96
Gambar 4. 29 Implementasi <i>source code</i> 10 Fitur dan 10 <i>fold</i>	97
Gambar 4. 30 Implementasi <i>source code</i> 15 Fitur dan 10 <i>fold</i>	98
Gambar 4. 31 <i>source code</i> Visualisasi <i>Confusion Matrix</i>	102
Gambar 4. 32 Hasil <i>Confusion Matrix</i> 10 dan 15 fitur dengan 5 <i>fold</i>	102
Gambar 4. 33 Hasil <i>Confusion Matrix</i> 10 fitur dengan 10 <i>fold</i>	106
Gambar 4. 34 Hasil <i>Confusion Matrix</i> 15 fitur dengan 10 <i>fold</i>	107
Gambar 5. 1 Implementasi <i>source code</i> muat model	113
Gambar 5. 2 Implementasi <i>source code</i> prediksi data uji	113
Gambar 5. 3 Hasil prediksi data uji pada <i>fold</i> ke-1	113
Gambar 5. 4 Hasil prediksi data uji pada <i>fold</i> ke-2.....	114
Gambar 5. 5 Hasil prediksi data uji pada <i>fold</i> ke-3.....	114
Gambar 5. 6 Hasil prediksi data uji pada <i>fold</i> ke-4.....	114
Gambar 5. 7 Hasil prediksi data uji pada <i>fold</i> ke-5.....	114
Gambar 5. 8 Implementasi <i>source code</i> voting mayoritas dan <i>accuracy</i>	115
Gambar 5. 9 Hasil <i>voting</i> mayoritas prediksi <i>Random Forest</i>	115

ABSTRAK

Tuberkulosis (TBC) merupakan penyakit menular yang disebabkan oleh *Mycobacterium tuberculosis* dan masih menjadi salah satu penyebab kematian tertinggi di dunia, termasuk di Indonesia. Penyakit ini menyerang paru-paru dan memerlukan diagnosis yang cepat dan akurat. Namun, keterbatasan tenaga medis dan fasilitas laboratorium, khususnya di puskesmas daerah pedalaman, sering menyebabkan keterlambatan diagnosis. Oleh karena itu, diperlukan sistem pendukung keputusan berbasis data gejala klinis untuk membantu diagnosis dini. Penelitian ini bertujuan membangun model klasifikasi TBC menggunakan data gejala klinis pasien. Seleksi fitur dilakukan menggunakan algoritma *Random Forest* berbasis *Gini Importance*, kemudian kinerjanya dibandingkan dengan *K-Nearest Neighbors* (KNN) menggunakan variasi 10 dan 15 fitur serta skema *5-Fold* dan *10-Fold Cross Validation*. Evaluasi model dilakukan menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Hasil pelatihan menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan *Accuracy* 99,0093%, *Precision* 99,0584%, *Recall* 99,0183%, dan *F1-Score* 99,0088% pada konfigurasi 15 fitur dan *5-Fold Cross Validation*. Pengujian data riil menggunakan metode *Majority Voting* menghasilkan *Accuracy* sebesar 98,00%, dengan 48 dari 50 data uji berhasil diklasifikasikan dengan benar. Berdasarkan hasil tersebut, algoritma *Random Forest* dalam penelitian ini ditetapkan sebagai metode terbaik dan dinilai andal untuk klasifikasi TBC pada fasilitas kesehatan tingkat pertama.

Kata Kunci: *Tuberkulosis*, Gejala Klinis, *Random Forest*, KNN, Klasifikasi, *Cross Validation*

ABSTRACT

Tuberculosis (TB) is an infectious disease caused by Mycobacterium tuberculosis and remains one of the leading causes of death worldwide, including in Indonesia. This disease primarily affects the lungs and requires rapid and accurate diagnosis. However, limitations in medical personnel and laboratory facilities, particularly in rural primary healthcare centers, often result in delayed diagnosis. Therefore, a decision support system based on clinical symptom data is needed to assist early diagnosis. This study aims to develop a TB classification model using patients' clinical symptom data. Feature selection was performed using the Random Forest algorithm based on Gini Importance, and its performance was compared with the K-Nearest Neighbors (KNN) algorithm using variations of 10 and 15 features and 5-Fold and 10-Fold Cross Validation schemes. Model performance was evaluated using Accuracy, Precision, Recall, and F1-Score metrics. The training results indicate that Random Forest achieved the best performance with an Accuracy of 99.0093%, Precision of 99.0584%, Recall of 99.0183%, and F1-Score of 99.0088% using 15 features and 5-Fold Cross Validation. Testing on real data using the Majority Voting method produced an Accuracy of 98.00%, with 48 out of 50 test samples correctly classified. Based on these results, the Random Forest algorithm is identified as the most effective method and is considered reliable for TB classification in primary healthcare facilities.

Keywords: *Tuberculosis, Clinical Symptoms, Random Forest, KNN, Classification, Cross Validation*