

**PENGEMBANGAN ALGORITMA *RANDOM FOREST*
MENGUNAKAN METODE *REGULARIZED*
*IMPURITY REDUCTION (GREEDY RANDOM FOREST)***



**Disusun Untuk Memenuhi Syarat Menyelesaikan Tugas Akhir
Pendidikan Sarjana Terapan Pada Jurusan Teknik Elektro
Program Studi Teknik Telekomunikasi
Politeknik Negeri Sriwijaya**

OLEH:

Khairunnisa'

062140352374

POLITEKNIK NEGERI SRIWIJAYA

PALEMBANG

2025

TUGAS AKHIR

PENGEMBANGAN ALGORITMA *RANDOM FOREST* MENGUNAKAN METODE *REGULARIZED IMPURITY* *REDUCTION (GREEDY RANDOM FOREST)*



**Disusun Untuk Memenuhi Syarat Menyelesaikan Proposal Tugas Akhir
Pendidikan Sarjana Terapan Pada Jurusan Teknik Elektro
Program Studi Teknik Telekomunikasi**

Oleh:

**Nama : Khairunnisa'
(062140352374)**
Dosen Pembimbing I : Sopian Soim, S.T., M.T.
Dosen Pembimbing II : Lindawati, S.T., M.T.I.

**POLITEKNIK NEGERI SRIWIJAYA
PALEMBANG
2025**

**PENGEMBANGAN ALGORITMA *RANDOM FOREST*
MENGUNAKAN METODE *REGULARIZED*
*IMPURITY REDUCTION (GREEDY RANDOM FOREST)***



TUGAS AKHIR

**Disusun untuk Memenuhi Syarat Menyelesaikan Tugas Akhir
Pendidikan Sarjana Terapan pada Jurusan Teknik Elektro
Program Studi Teknik Telekomunikasi
Politeknik Negeri Sriwijaya**

OLEH:

**Khairunnisa'
062140352374**

Menyetujui,

**Palembang,
Dosen Pembimbing II**

2025

Dosen Pembimbing I

**Sopian Seim, S.T., M.T.
NIP. 197103142001121001**

**Lindawati, S.T., M.T.I.
NIP. 197105182006042001**

Mengetahui,

Ketua Jurusan Teknik Elektro

**Koordinator Program Studi
Sarjana Terapan Teknik
Telekomunikasi**



**Dr. Ir. Selamat Muslimin, S.T., M.Kom. IPM.
NIP. 197907222008011007**

**Mohammad Fadhli, S.Pd., M.T.
NIP. 199004032018031001**

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini menyatakan :

Nama : Khairunnisa'
Jenis Kelamin : Perempuan
Tempat, Tanggal Lahir : Palembang, 24 April 2004
Alamat : Jl. May Mahidin I No.1228
NIM : 062140352374
Program Studi : Sarjana Terapan Teknik Telekomunikasi
Jurusan : Teknik Elektro
Judul Skripsi/Laporan Akhir : Pengembangan Algoritma *Random Forest*
Menggunakan Metode *Regularized Impurity*
Reduction (Greedy Random Forest)

Menyatakan dengan sesungguhnya bahwa:

1. Skripsi/Laporan Akhir ini adalah hasil karya saya sendiri serta bebas dari tindakan plagiasi dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.
2. Dapat menyelesaikan segala urusan terkait pengumpulan revisi Skripsi/Laporan Akhir yang sudah disetujui oleh dewan penguji paling lama 1 bulan setelah ujian Skripsi/Laporan Akhir.
3. Dapat menyelesaikan segala urusan peminjaman/penggantian alat/buku dan lainnya paling lama 1 bulan setelah ujian Skripsi/Laporan Akhir.

Apabila dikemudian hari diketahui ada pernyataan yang terbukti tidak benar dan tidak dapat dipenuhi, maka saya siap bertanggung jawab dan menerima sanksi tidak diikutsertakan dalam prosesi wisuda serta dimasukkan dalam daftar hitam oleh jurusan Teknik Elektro sehingga berdampak tertundanya pengambilan Ijazah & Transkrip (ASLI & COPY). Demikian surat pernyataan ini dibuat dengan sebenar-benarnya dan dalam keadaan sadar tanpa paksaan.



Palembang, Agustus 2025
Yang Menyatakan



(Khairunnisa')

MOTTO DAN PERSEMBAHAN

“Jihad Terbesar Adalah Melawan Jiwamu Sendiri untuk Melawan Kejahatan di dalam Dirimu” – Nabi Muhammad SAW

"Sesungguhnya bersama kesulitan ada kemudahan." -QS. Al-Insyirah: 6

"Barang siapa yang menempuh jalan untuk mencari ilmu, maka Allah akan memudahkan baginya jalan menuju surga." -HR. Muslim, No. 2699

Tugas Akhir ini kupersembahkan kepada:

- 1. Ayah dan Ibu, terimakasih kepada kedua orang tua yang telah berjasa dalam mendukung, memberikan do'a, cinta, dan semangat yang tidak dapat dihitung.**
- 2. Saudari saya tersayang, Mbak Puput, yang tak henti-hentinya memberikan saya support.**
- 3. Bapak Sopian Soim, S.T., M.T. dan Ibu Lindawati, S.T., M.T.I, selaku Dosen Pembimbing I dan Dosen Pembimbing II, terimakasih atas bimbingan, arahan, dan dukungannya selama pengerjaan Tugas Akhir ini.**
- 4. Seluruh teman-teman Kebel yang telah berjuang dari awal semester 1 hingga saat ini, semoga kedepannya diberikan rezeki, kemudahan, dan kelancaran dalam menjalani kehidupan.**

ABSTRAK

PENGEMBANGAN ALGORITMA *RANDOM FOREST* MENGGUNAKAN METODE *REGULARIZED IMPURITY REDUCTION (GREEDY RANDOM FOREST)*

(2025: XIII+ 78 halaman + 41 gambar + 11 tabel)

KHAIRUNNISA

062140352374

JURUSAN TEKNIK ELEKTRO

**PROGRAM STUDI SARJANA TERAPAN TEKNIK TELEKOMUNIKASI
POLITEKNIK NEGERI SRIWIJAYA**

Meningkatnya kompleksitas dan besarnya data di bidang kesehatan menuntut pengembangan model *machine learning* yang lebih adaptif dan andal, khususnya untuk tugas klasifikasi berskala besar. *Random Forest* merupakan salah satu metode *ensemble* yang banyak digunakan karena ketangguhannya dalam menghadapi permasalahan seperti ketidakseimbangan kelas dan dimensi fitur yang tinggi. Namun, model *Random Forest* standar masih memiliki kelemahan terhadap *overfitting* dan dominasi fitur, terutama akibat mekanisme pemilihan *split* yang bersifat lokal. Oleh karena itu, penelitian ini mengusulkan *Greedy Random Forest*, sebuah algoritma yang disempurnakan melalui mekanisme pemilihan *split* menggunakan *Regularized Impurity Reduction* yang mempertimbangkan *balance*, *efficiency*, dan *discrimination*. Evaluasi dilakukan pada dua dataset kesehatan, yaitu Alzheimer dan Hypothyroid. Hasil pengujian menunjukkan bahwa *Greedy Random Forest* mampu meningkatkan performa dibandingkan *Random Forest* standar. Pada dataset Alzheimer, *Greedy Random Forest* menghasilkan akurasi sebesar 0,9372 sedikit lebih baik dari *Random Forest* yang memiliki akurasi 0,9349, signifikan terlihat pada dataset Hypothyroid, di mana *Greedy Random Forest* mencapai akurasi 0,9677, mengungguli *Random Forest* yang hanya memperoleh akurasi 0,8991. Dengan hasil tersebut, *Greedy Random Forest* terbukti efektif dalam mengatasi *overfitting*, meningkatkan generalisasi model, serta menghasilkan performa yang lebih konsisten dan interpretatif.

Kata Kunci: *Greedy Algorithm, Health Data, Machine Learning, Random Forest.*

ABSTRACT

DEVELOPMENT OF THE RANDOM FOREST ALGORITHM USING THE REGULARIZED IMPURITY REDUCTION METHOD (GREEDY RANDOM FOREST)

(2025: XIII + 78 halaman + 41 gambar + 11 tabel)

KHAIRUNNISA

062140352374

JURUSAN TEKNIK ELEKTRO

PROGRAM STUDI SARJANA TERAPAN TEKNIK TELEKOMUNIKASI

POLITEKNIK NEGERI SRIWIJAYA

The increasing complexity and scale of healthcare data demand the development of more adaptive and reliable machine learning models, particularly for large-scale classification tasks. Random Forest is one of the most widely used ensemble methods due to its robustness in handling issues such as class imbalance and high-dimensional feature spaces. However, the standard Random Forest model still suffers from overfitting and feature dominance, mainly due to its locally based split selection mechanism. Therefore, this study proposes the Greedy Random Forest, an enhanced algorithm that applies a split selection mechanism based on Regularized Impurity Reduction, which considers balance, efficiency, and discrimination. Evaluation was conducted on two healthcare datasets, namely Alzheimer and Hypothyroid. The experimental results show that Greedy Random Forest outperforms the standard Random Forest. On the Alzheimer dataset, Greedy Random Forest achieved an accuracy of 0.9372, slightly higher than Random Forest with 0.9349. A more significant improvement was observed on the Hypothyroid dataset, where Greedy Random Forest achieved an accuracy of 0.9677, surpassing Random Forest which only reached 0.8991. These results demonstrate that Greedy Random Forest is effective in reducing overfitting, improving model generalization, and providing more consistent and interpretable performance.

Keywords: *Greedy Algorithm, Health Data, Machine Learning, Random Forest.*

DAFTAR ISI

MOTTO DAN PERSEMBAHAN.....	v
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
KATA PENGANTAR.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Metode Penulisan	4
1.7 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Machine Learning	7
2.2 Decision Tree	8
2.3 Random Forest	9
2.4 Split criterion.....	10
2.5 <i>Regularized Impurity Reduction (Greedy Random Forest)</i>	12
2.6 <i>Overfitting</i>	12
2.7 Akurasi, <i>Precision</i> , <i>F1-Score</i> , dan <i>Recall</i>	13
2.8 <i>Area Under Curve (AUC) Metric</i>	14
2.9 Python	14
2.9.1 Keunggulan <i>Python</i> dalam Pengembangan <i>Machine Learning</i>	14
2.9.2 Scikit-learn.....	15
2.9.3 <i>Numpy</i> dan <i>Pandas</i>	15
2.9.4 <i>Seaborn</i>	15
2.9.5 Matplotlib	16
2.10 Perbandingan Penelitian.....	16
BAB III METODOLOGI PENELITIAN	18
3.1 Kerangka Penelitian	18
3.2 Pengumpulan Data	21
3.2.1 <i>Alzheimers</i>	22
3.2.2 <i>Hyphotyroid</i>	24
3.3 Pre-processing Data.....	27
3.3.1 <i>Encoding Data</i>	27
3.3.2 <i>Remove Missing Values</i>	28
3.3.3 Penghapusan <i>Outlier</i>	28
3.3.4 Pembagian Data	28
3.3.5 <i>Select Feature</i>	29
3.3.6 <i>Scaling Feature</i>	29
3.4 Model <i>Standard Random Forest</i>	29
3.5 Model <i>Greedy Random Forest</i>	30

3.6 Evaluasi Model.....	33
BAB IV HASIL DAN PEMBAHASAN	36
4.1 <i>Pre-Processing</i> Data.....	36
4.1.1 Encoding Data	36
4.1.2 Remove Missing Values	37
4.1.3 Penghapusan <i>Outlier</i>	41
4.1.4 Pembagian Data	45
4.1.5 Select Fitur.....	46
4.1.6 <i>Scaling Fitur</i>	49
4.2 Hasil Model	50
4.2.1 Alzheimer	55
4.2.2 Hypothyroid.....	57
4.3 Analisa Hasil	58
BAB V KESIMPULAN DAN SARAN	62
5.1 Kesimpulan	62
5.2 Saran.....	63
REFERENSI.....	64

DAFTAR GAMBAR

Gambar 2. 1 Decision Tree[13]	8
Gambar 2. 2 Random Forest[12]	9
Gambar 3. 1 Bagan Penelitian.....	19
Gambar 3. 2 Tahapan Kerja	20
Gambar 3. 3 Dataset Alzheimer	22
Gambar 3. 4 Heatmap Correlation Alzheimer	24
Gambar 3. 5 <i>Heatmap Correlation</i> Hypothyroid.....	26
Gambar 3. 6 Tahapan Preprocessing.....	27
Gambar 3. 7 Model Standard Random Forest.....	29
Gambar 3. 8 <i>Greedy Random Forest</i>	31
Gambar 4. 1 <i>Encoding Data</i>	36
Gambar 4. 2 Dataset Alzheimer sebelum di Encoding Data.....	37
Gambar 4. 3 Hasil Dataset Alzheimer yang Telah di Encoding Data.....	37
Gambar 4. 4 Codingan Untuk Memeriksa NaN dan Duplikat	38
Gambar 4. 5 Hasil Pemeriksaan Pada Dataset Alzheimer	38
Gambar 4. 6 Hasil Pemeriksaan NaN dan Duplikat Pada Dataset Hypothyroid ..	39
Gambar 4. 7 Codingan untuk Menghapus Duplikat	39
Gambar 4. 8 Hasil Penghapusan Baris Duplikat pada Dataset Hypothyroid.....	39
Gambar 4. 9 Heatmap Correlation Pada Dataset Hypothyroid	40
Gambar 4. 10 <i>Remove Missing Values</i>	40
Gambar 4. 11 <i>Heatmap Correlation Hypothyroid</i>	41
Gambar 4. 12 Penampilan <i>Outlier</i>	42
Gambar 4. 13 Jumlah Outlier Pada Dataset Alzheimer	42
Gambar 4. 14 Hapus <i>Outlier</i>	43
Gambar 4. 15 Jumlah Outlier yang Telah Dihapus Pada Dataset Alzheimer	43
Gambar 4. 16 Menampilkan <i>Outlier</i> Pada <i>Hypothyroid</i>	44
Gambar 4. 17 Jumlah Outlier pada Dataset Hypothyroid.....	44
Gambar 4. 18 Menghapus <i>Outlier</i> Pada Dataset Hypothyroid	45
Gambar 4. 19 Jumlah Outlier yang Telah Dihapus Pada dataset Hypothyroid	45
Gambar 4. 20 <i>Train dan Test Split</i>	45
Gambar 4. 21 <i>F-Score Feature</i>	46
Gambar 4. 22 Daftar F-Score dan P-Value Pada Dataset Alzheimer	47
Gambar 4. 23 <i>SelectKBest</i>	47
Gambar 4. 24 Menghitung Nilai F-Score dan P-Value Hypothyroid.....	48
Gambar 4. 25 Daftar F-Score dan P-Value Pada Dataset Hypothyroid.....	48
Gambar 4. 26 <i>SelectKBest</i>	49
Gambar 4. 27 <i>Scaling Feature</i>	50
Gambar 4. 28 Perbandingan Akurasi Pada Dataset Alzheimer dan Hypothyroid	52
Gambar 4. 29 ROC Comparison	55
Gambar 4. 30 <i>Confusion Matrix</i> Alzheimer.....	55
Gambar 4. 31 <i>Confusion Matrix Hypothyroid</i>	57
Gambar 4. 32 Perbandingan ROC Keseluruhan	59

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i> [18].....	13
Tabel 2. 2 Perbandingan Penelitian.....	16
Tabel 3. 1 Karakteristik Dataset.....	22
Tabel 3. 2 Tipe Data Alzheimer	23
Tabel 3. 3 Tipe Data <i>Hypothyroid</i>	25
Tabel 4. 1 Hyperparameter di Tiap Model.....	50
Tabel 4. 2 Evaluasi Accuracy Model	51
Tabel 4. 3 Evaluasi Area Under Curve	53
Tabel 4. 4 <i>Accuracy Alzheimer</i>	56
Tabel 4. 5 Evaluasi Hypothyroid	57
Tabel 4. 6 Perbandingan Akurasi Keseluruhan.....	58