

TUGAS AKHIR

**IMPLEMENTASI *FEATURE ENGINEERING* DAN
HYPERPARAMETER TUNING UNTUK MENINGKATKAN
AKURASI MODEL *RANDOM FOREST* PADA
KLASIFIKASI RISIKO KREDIT**



NADEA PUTRI NUR FAUZI

20110031

**PROGRAM STUDI SAINS DATA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO
2024**

TUGAS AKHIR

**IMPLEMENTASI *FEATURE ENGINEERING* DAN
HYPERPARAMETER TUNING UNTUK MENINGKATKAN
AKURASI MODEL *RANDOM FOREST* PADA
KLASIFIKASI RISIKO KREDIT**

***IMPLEMENTATION OF FEATURE ENGINEERING AND
HYPERPARAMETER TUNING TO IMPROVE THE
ACCURACY OF RANDOM FOREST MODELS ON
CREDIT RISK CLASSIFICATION***

Disusun Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Sains Data



NADEA PUTRI NUR FAUZI

20110031

**PROGRAM STUDI SAINS DATA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO**

2024

HALAMAN PERSETUJUAN PEMBIMBING

**IMPLEMENTASI *FEATURE ENGINEERING* DAN
HYPERPARAMETER TUNING UNTUK MENINGKATKAN
AKURASI MODEL *RANDOM FOREST* PADA
KLASIFIKASI RISIKO KREDIT**

***IMPLEMENTATION OF FEATURE ENGINEERING AND
HYPERPARAMETER TUNING TO IMPROVE THE
ACCURACY OF RANDOM FOREST MODELS ON
CREDIT RISK CLASSIFICATION***

Dipersiapkan dan Disusun Oleh
NADEA PUTRI NUR FAUZI
20110031

**Fakultas Informatika
Institut Teknologi Telkom Purwokerto**

Pada Tanggal: 10 Januari 2024

Pembimbing Utama,



(Siti Khomsah S.Kom., M.Cs.)

NIDN: 0517108101

Pembimbing Pendamping,



(Aditya Dwi Putro W. S.Kom., M.Kom.)

NIDN: 0624119303

**IMPLEMENTASI *FEATURE ENGINEERING* DAN
HYPERPARAMETER TUNING UNTUK MENINGKATKAN
AKURASI MODEL *RANDOM FOREST* PADA
KLASIFIKASI RISIKO KREDIT**

***IMPLEMENTATION OF FEATURE ENGINEERING AND
HYPERPARAMETER TUNING TO IMPROVE THE
ACCURACY OF RANDOM FOREST MODELS ON
CREDIT RISK CLASSIFICATION***

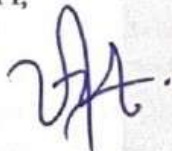
Disusun Oleh

NADEA PUTRI NUR FAUZI

20110031

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tugas
Akhir pada Hari Kamis, Tanggal 18 Januari 2024

Penguji I,



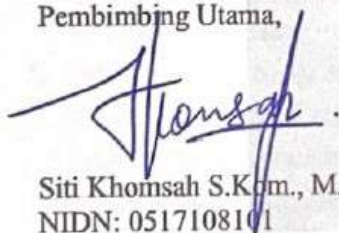
Atika Ratna Dewi S.Si., M.Sc.
NIDN: 0625088202

Penguji II,



Yesy Dian Rosita S.Kom., M.Kom.
NIDN: 0705108503

Pembimbing Utama,



Siti Khomsah S.Kom., M.Cs.
NIDN: 0517108101

Pembimbing Pendamping,



Aditya Dwi Putro W., S.Kom., M.Kom.
NIDN: 0624119303


Auliya Burhanuddin S.Si., M.Kom.
NIK. 19820008

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nadea Putri Nur Fauzi
NIM : 20110031
Program Studi : S1 Sains Data

Menyatakan bahwa Tugas Akhir dengan judul berikut:
**IMPLEMENTASI *FEATURE ENGINEERING* DAN
HYPERPARAMETER TUNING UNTUK MENINGKATKAN AKURASI
MODEL *RANDOM FOREST* PADA KLASIFIKASI RISIKO KREDIT**

Dosen Pembimbing Utama : Siti Khomsah S.Kom., M.Cs
Dosen Pembimbing Pendamping : Aditya Dwi Putro W. S.Kom., M.Kom

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Institut Teknologi Telkom Purwokerto maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan, dan penelitian Saya Sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing.
3. Dalam Karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggungjawab Saya, bukan tanggungjawab Institut Teknologi Telkom Purwokerto.
5. Pernyataan ini Saya buat dengan sesungguhnya, apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima Sanksi Akademik dengan pencabutan gelar yang sudah diperoleh serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Purwokerto, 10 Januari 2024,

Yang Menyatakan,



(Nadea Putri Nur Fauzi)

KATA PENGANTAR

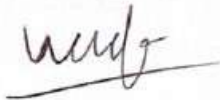
Puji syukur dipanjatkan kepada Allah Subhanahu Wa Ta'ala, karena dengan nikmat dan karuniaNya penulis dapat menyelesaikan penelitian ini yang berjudul **“Implementasi *Feature Engineering* dan *Hyperparameter Tuning* Untuk Meningkatkan Akurasi Model *Random Forest* Pada Klasifikasi Risiko Kredit”** sebagai bagian dari perjalanan akademis. Sholawat serta salam semoga tetap tercurah limpahkan kepada Baginda Rasulullah Shallallahu ‘Alaihi Wasallam. Pada kesempatan ini penulis mendapatkan banyak ilmu, pengalaman, petunjuk, dan bimbingan yang sangat baik. Untuk itu penulis ingin mengucapkan terima kasih kepada semua pihak yang telah memberikan arahan, dukungan serta bantuan dalam penulisan Tugas Akhir ini, terlebih kepada:

1. Keluarga yang selalu mendoakan, memberikan semangat, dan mendukung baik secara moral maupun material.
2. Ibu Dr. Tenia Wahyuningrum S.Kom., M.T, selaku Rektor Institut Teknologi Telkom Purwokerto.
3. Bapak Auliya Burhanuddin S.Si., M.Kom, selaku Dekan Fakultas Informatika Institut Teknologi Telkom Purwokerto.
4. Ibu Siti Khomsah S.Kom., M.Cs., selaku Kepala Program Studi S1 Sains Data sekaligus Dosen Pembimbing Utama yang selalu mendukung, memberikan banyak saran dan masukan dalam pengerjaan Tugas Akhir.
5. Bapak Aditya Dwi Putro Wicaksono S.Kom., M.Kom selaku Dosen Pembimbing Kedua yang selalu mendukung dan memberikan arahan dalam pengerjaan Tugas Akhir.
6. Ibu Atika Ratna Dewi S.Si., M.Sc., Ibu Yesy Diah Rosita S.Kom., M.Kom. dan Ibu Rima Dias Ramadhani S.Kom., M.Kom. selaku Dosen Penguji yang telah memberikan banyak saran, arahan, dan bimbingan dalam penyelesaian Tugas Akhir ini.
7. Seluruh Staff Dosen Fakultas Informatika, Institut Teknologi Telkom Purwokerto yang telah memberikan bimbingan, dukungan, dan pengajaran pada segenap ilmu yang diberikan.

8. Teman – teman terdekat saya, Elisabeth Angeline, Niesya Ayunda, Fanisa Nimastiti, Ike Fitria, Afifatunni'mah, Rio Ghaniy, Abdul Latief, Teddy Syamkalla, dan Miko Ardian yang selalu mendukung, memberikan banyak saran, dan meluangkan waktunya untuk diskusi serta berbagi ide bersama.
9. Seluruh pihak yang tidak bisa disebutkan satu – persatu atas dukungan, berbagi ide, dan menjadi pilar semangat selama penelitian ini berlangsung.

Semoga laporan tugas akhir ini dapat memberikan manfaat dan menjadi sumbangan kecil bagi perkembangan ilmu pengetahuan. Akhir kata, penulis memohon maaf jika terdapat kekurangan dalam penelitian ini dan berharap agar laporan tugas akhir ini dapat diterima dengan baik oleh semua pihak.

Puwokerto, 10 Januari 2024



Nadea Putri Nur Fauzi

DAFTAR ISI

HALAMAN PERSETUJUAN PEMBIMBING	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
ABSTRAK.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah.....	3
1.3. Pertanyaan Penelitian	3
1.4. Tujuan Penelitian.....	4
1.5. Batasan Masalah.....	4
1.6. Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1. Tinjauan Pustaka	5
2.2. Landasan Teori	14
2.2.1. <i>Credit Risk</i> / Risiko Kredit.....	14
2.2.2. <i>Machine Learning</i>	14
2.2.3. <i>Exploratory Data Analysis</i>	16
2.2.4. <i>Feature Engineering</i>	18
2.2.5. <i>Random Forest</i>	21
2.2.6. <i>Hyperparameter Tuning</i>	22
2.2.7. Matriks Konfusi	25
2.2.8. <i>Classification Report</i>	27
2.2.9. Kurva AUC (<i>Area Under the Curve</i>).....	27
BAB III METODOLOGI PENELITIAN.....	28

3.1. Subjek dan Objek Penelitian	28
3.2. Alat dan Bahan Penelitian	28
3.2.1. Perangkat Keras / <i>Hardware</i>	28
3.2.2. Perangkat Lunak / <i>Software</i>	28
3.2.3. Bahan / Data.....	28
3.3. Diagram Alir Penelitian.....	28
3.3.1. Kajian Pustaka	29
3.3.2. Metodologi SEMMA	29
3.3.3. Analisis Hasil Eksperimen dan Kesimpulan.....	43
BAB IV HASIL DAN PEMBAHASAN	44
4.1. <i>Modify</i>	44
4.1.1. <i>Data Preprocessing</i>	44
4.1.2. <i>Feature Engineering</i>	47
4.2. <i>Model</i>	49
4.2.1. Model 1 (Model Baseline)	49
4.2.2. Model 2 (<i>Feature Engineering</i> & RF Default).....	50
4.2.3. Model 3 (<i>Feature Engineering</i> & HT Random Search).....	50
4.2.4. Model 4 (<i>Feature Engineering</i> & HT Grid Search)	52
4.3. <i>Assess</i>	55
4.3.1. Nilai Akurasi.....	55
4.3.2. <i>Confusion Matrix</i>	55
4.3.3. <i>Classification Report</i>	58
4.3.4. Grafik AUC.....	59
BAB V KESIMPULAN & SARAN	61
5.1. Kesimpulan.....	61
5.2. Saran.....	61
DAFTAR PUSTAKA	62
LAMPIRAN.....	67

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	9
Tabel 2.2 Contoh Hasil Standardisasi Menggunakan Robust Scaler.....	21
Tabel 3.1 Fitur Data Risiko Kredit.....	30
Tabel 3.2 Informasi Fitur Data.....	30
Tabel 3.3 Hasil Statistik Deskriptif.....	31
Tabel 3.4 Jumlah Data Null	31
Tabel 3.5 Jumlah Data Outliers.....	32
Tabel 3.6 Contoh Perhitungan KNNImputer	36
Tabel 3.7 Contoh Penerapan Label Encoder.....	37
Tabel 3.8 Perlakuan Terhadap Pemodelan.....	43
Tabel 4.1 Hasil Penginputan Data Null	44
Tabel 4.2 Alasan Pengubahan Nilai (Replace Value).....	45
Tabel 4.3 Hasil Replace Value.....	46
Tabel 4.4 Hasil Encoding Data	47
Tabel 4.5 Parameter Model 1 (Default Random Forest).....	49
Tabel 4.6 Parameter Model 2 (Default Random Forest).....	50
Tabel 4.7 Parameter Random Forest Randomized Search CV	51
Tabel 4.8 Parameter Terbaik Randomized Search CV	51
Tabel 4.9 Parameter Random Forest Grid Search CV	53
Tabel 4.10 Parameter Terbaik Grid Search CV	53
Tabel 4.11 Hasil Akurasi Model	55
Tabel 4.12 Skor AUC 4 Model	60
Tabel 4.13 Durasi Testing Model.....	60

DAFTAR GAMBAR

Gambar 2.1 Sub Pembagian Machine learning.....	15
Gambar 2.2 Alur Klasifikasi Menggunakan Machine learning	16
Gambar 2.3 Contoh Data Berdistribusi Normal.....	17
Gambar 2.4 Data Outliers	17
Gambar 2.5 Feature Selection	19
Gambar 2.6 Oversampling	19
Gambar 2.7 Contoh Random Ovesampling	20
Gambar 2.8 Cara Kerja Algoritma Random Forest	21
Gambar 2.9 Percobaan Jumlah $n_{estimators}$	22
Gambar 2.10 Kedalaman pohon.....	23
Gambar 2. 11 Contoh $Min_Samples_Split$	24
Gambar 2.12 Contoh $Min_Samples_Leaf$	25
Gambar 2.13 Confusion Matrix	26
Gambar 3.1 Diagram Alir Penelitian	29
Gambar 3.2 Nilai Korelasi Antar Variabel	33
Gambar 3.3 Nilai Feature Importance.....	34
Gambar 3.4 Alur Kerja KNNImputer	35
Gambar 3.5 Alur Kerja Random Oversampling	38
Gambar 3.6 Contoh Perbandingan Kelas Pada Data.....	39
Gambar 3.7 Contoh Data Hasil Oversampling	39
Gambar 3.8 Data Sebelum Standardisasi	40
Gambar 3.9 Data Setelah Standardisasi	40
Gambar 3.10 Algoritma Random Forest.....	41
Gambar 3.11 Contoh Penggunaan Random Forest Pada Data Baru.....	42
Gambar 3.12 Penerapan Hyperparameter Tuning Pada Random Forest	42
Gambar 4.1 Jumlah Data Setelah Oversampling	48
Gambar 4.2 Data Hasil Standardisasi	48
Gambar 4.3 Confusion Matrix Model 1	56
Gambar 4.4 Confusion Matrix Model 2.....	56

Gambar 4.5 Confusion Matrix Model 3.....	57
Gambar 4.6 Confusion Matrix Model 4.....	57
Gambar 4.7 Classification Report Model 1.....	58
Gambar 4.8 Classification Report Model 2, 3, dan 4.....	58
Gambar 4.9 Skor AUC Model 1	59
Gambar 4.10 Skor AUC Model 2, 3, dan 4.....	59

DAFTAR LAMPIRAN

Lampiran 1. Data.....	65
Lampiran 2. Code Exploratory Data Analysis.....	65
Lampiran 3. Code Preprocessing.....	66
Lampiran 4. Code Modelling.....	68
Lampiran 5. Code Evaluasi Model.....	69