



**IMPLEMENTASI ALGORITMA RANDOM FOREST
TERHADAP PREDIKSI GOOD LOAN/BAD LOAN KREDIT
NASABAH BANK DI JAKARTA**

SKRIPSI

SULTAN FAREL SYAH REZA

1910511105

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN
JAKARTA**

**FAKULTAS ILMU KOMPUTER
PROGRAM STUDI S-1 INFORMATIKA**

2023



UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

**IMPLEMENTASI ALGORITMA RANDOM FOREST TERHADAP
PREDIKSI GOOD LOAN/BAD LOAN KREDIT NASABAH BANK DI
JAKARTA**

SKRIPSI

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

SULTAN FAREL SYAH REZA

1910511105

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S-1 INFORMATIKA

2023

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Sultan Farel Syah Reza

NIM : 1910511105

Tanggal : 3 Juli 2023

Judul Skripsi : **Implementasi Algoritma *Random Forest* Terhadap Prediksi *Good Loan/Bad Loan* Kredit Nasabah Bank di Jakarta**

Bilamana pada kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 3 Juli 2023

Yang Menyatakan,



Sultan Farel Syah Reza

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademika Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Sultan Farel Syah Reza
NIM : 1910511105
Fakultas : Ilmu Komputer
Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan karya ilmiah saya kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royalti Non-Eksklusif (*Non-Exchange Royalty Free Right*) untuk dipublikasikan dengan judul:

Implementasi Algoritma *Random Forest* terhadap Prediksi *Good Loan / Bad Loan* Kredit Nasabah Bank di Jakarta

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih media atau memformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 3 Juli 2023

Yang Menyatakan,



Sultan Farel Syah Reza

LEMBAR PENGESAHAN

Tugas akhir ini diajukan oleh:

Nama : Sultan Farel Syah Reza
NIM : 1910511105
Program Studi : S1 Informatika
Judul Tugas : IMPLEMENTASI ALGORITMA RANDOM FOREST
Akhir TERHADAP PREDIKSI GOOD LOAN/BAD LOAN
KREDIT NASABAH BANK DI JAKARTA

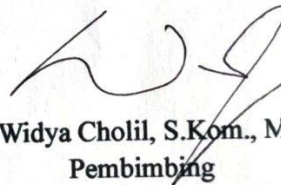
Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



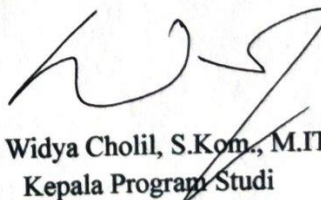
Yuni Widiastiwi, S.Kom, M.Si.
Penguji I



Anita Muliawati, S.Kom., MTI.
Penguji II



Dr. Widya Cholil, S.Kom., M.IT.
Pembimbing



Dr. Widya Cholil, S.Kom., M.IT.
Kepala Program Studi

Ditetapkan di : Jakarta
Tanggal Ujian : 3 Juli 2023



Abstrak

Kredit merupakan pemberian (penjaminan) barang atau jasa oleh satu pihak dengan uang untuk memenuhi segala kebutuhan, keinginan, dan aspirasi masyarakat, berdasarkan persaingan masyarakat yang semakin kompetitif. Risiko kredit adalah risiko kerugian yang terkait dengan ketidakmampuan dan/atau keengganan peminjam untuk memenuhi kewajibannya untuk membayar kembali dana pinjaman secara penuh pada atau setelah tanggal jatuh tempo. Dalam pemberian kredit, bank harus mengidentifikasi, mengelola, dan memastikan risiko kredit pada seluruh produk dan harus melalui proses pengendalian manajemen risiko yang layak. Oleh karena itu, dibutuhkan sebuah sistem dimana yang mampu memprediksi risiko kredit yang ditimbulkan oleh nasabah bank yang tidak mampu membayar pinjaman kredit agar bank tidak merugi. Menggunakan data yang didapatkan dari ID/X untuk membuat sebuah model *machine learning* menggunakan algoritma *Random Forest*. Keluaran yang dihasilkan dari model yang telah dibuat adalah pengklasifikasian nasabah bank terbilang *good loan / bad loan*. Model klasifikasi yang diperoleh akan dievaluasi menggunakan nilai *accuracy*, *precision*, *recall*, dan *F1-Score*. Hasil evaluasi terbaik didapatkan oleh Model rasio perbandingan 70% data latih dan 30% data uji dengan nilai akurasi sebesar 84,32%, nilai *precision* sebesar 96,79%, nilai *recall* sebesar 86,44% dan *F1-score* sebesar 91.3%.

Kata Kunci: Kredit, Manajemen Risiko, *Machine Learning*, Prediksi, *Random Forest*

Abstract

Credit is the granting (guarantee) of goods or services by one party with money to meet all the needs, wants and aspirations of society, based on increasingly competitive society. Credit risk is the risk of loss associated with the inability and/or unwillingness of the borrower to fulfill its obligation to repay the loan in full on or after the maturity date. In granting credit, banks must identify, manage and ensure credit risk in all products and must go through an appropriate risk management control process. Therefore, a system is needed which is able to predict the credit risk posed by bank customers who are unable to pay off credit loans so that the bank does not lose money. Using data obtained from ID/X to created a machine learning model using the Random Forest algorithm. The output generated from the model that has been made is the classification of bank customers into good loans / bad loans. The classification model obtained will be evaluated using accuracy, precision, recall, and F1-Score values. The best evaluation results are obtained with a model ratio of 70% training data and 30% testing data, achieving an accuracy of 84.32%, precision of 96.79%, recall of 86.44%, and F1-score of 91.3%.

Keywords: *Credit, Risk Management, Machine Learning, Prediction, Random Forest*

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah Yang Maha Esa atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi dengan baik sehingga dapat menyelesaikan jenjang pendidikan sarjana S1 tepat waktu. Penulisan skripsi ini bertujuan untuk memenuhi salah satu persyaratan untuk memperoleh gelar Sarjana Komputer, Jurusan Informatika. Penulisan skripsi dengan judul “Implementasi Algoritma *Random Forest* terhadap Prediksi *Good Loan / Bad Loan* Kredit Nasabah Bank di Jakarta” yang dilakukan sejak September 2022 ini tidak terlepas dari bantuan beberapa pihak-pihak lain. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih sebesar-besarnya kepada:

1. Allah Yang Maha Esa, yang telah memberikan kesehatan serta ilmu kepada penulis sehingga dapat menyelesaikan skripsi ini dengan baik.
2. Ibu Dr. Ermatita, M.Kom selaku Dekan Fakultas Ilmu Komputer UPN Veteran Jakarta beserta jajarannya.
3. Ibu Dr. Widya Cholil, M.I.T. selaku Kepala Program Studi Informatika dan Dosen Pembimbing Skripsi yang membantu penulis dalam penyusunan skripsi dan memberikan saran dengan baik.
4. Ibu Helena Nurramdhani Irmada, S.Pd, M.Kom selaku Dosen Pembimbing Akademik yang membantu penulis selama duduk di bangku perkuliahan.
5. Orang tua dan keluarga, yang selalu mendoakan serta mendukung peneliti sehingga dapat menyelesaikan skripsi dengan baik.
6. Seluruh teman dekat penulis yang telah memberikan dukungan, doa, dan semangat untuk menyelesaikan skripsi ini, yang namanya tidak dapat penulis sebutkan satu per satu.
7. Seluruh pihak yang terlibat dalam kelancaran pembuatan skripsi ini dan yang belum disebutkan di atas, penulis ucapkan terimakasih.

Penulis menyadari bahwa penyusunan skripsi ini jauh dari sempurna. Oleh karena itu, penulis harap kepada para peneliti selanjutnya untuk memberikan saran dan kritik yang bersifat membangun untuk kesempurnaan skripsi ini. Akhir kata penulis berharap semoga skripsi ini mampu bermanfaat untuk semua pihak.

DAFTAR ISI

PERNYATAAN ORISINALITAS.....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS.....	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
ABSTRACT.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian.....	2
1.4. Manfaat Penelitian.....	2
1.5. Ruang Lingkup	3
1.6. Luaran Yang Diharapkan	3
1.7. Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1. Bank.....	5
2.2. Kredit.....	5
2.2.1 Definisi Kredit.....	5
2.2.2 Definisi Risiko Kredit	6
2.3. Data Mining.....	7
2.4. Python.....	7

2.5.	<i>Knowledge Discovery Database (KDD)</i>	8
<u>2.5.1.</u>	<i>Definisi Knowledge Discovery Database (KDD)</i>	8
<u>2.5.2.</u>	<i>Data Selection</i>	9
<u>2.5.3.</u>	<i>Transformation</i>	9
<u>2.5.4.</u>	<i>Interpretation/Evaluation</i>	9
2.6.	<i>Machine Learning</i>	9
2.7.	<i>Random Forest</i>	10
2.8.	<i>Confusion Matrix</i>	11
2.8.1.	<i>Accuracy</i>	12
2.8.2.	<i>Precision</i>	12
2.8.3.	<i>Recall</i>	12
2.8.4.	<i>F-1 Score</i>	13
2.9.	<i>ROC</i>	13
2.10.	<i>AUC</i>	14
2.11.	<i>Imbalanced Class</i>	15
2.12.	<i>Oversampling</i>	16
2.13.	<i>Entropy</i>	16
2.14.	<i>Gini Index</i>	17
2.15.	<i>Review Penelitian Terdahulu</i>	17
BAB III METODOLOGI PENELITIAN		21
3.1.	<i>Alur Penelitian</i>	21
3.2.	<i>Tahapan Penelitian</i>	22
3.2.1.	<i>Identifikasi Masalah</i>	22
3.2.2.	<i>Studi Literatur</i>	22
3.2.3.	<i>Pengumpulan Data</i>	22
3.2.4.	<i>Preprocessing data</i>	23

3.2.5.	Exploratory Data Analysis	23
3.2.6.	Training dan Testing	24
3.2.7.	Pemodelan	24
3.2.8.	Evaluasi	24
3.3.	Waktu dan Tempat Pelaksanaan.....	25
3.4.	Alat dan Bahan yang digunakan.....	25
<u>3.4.1.</u>	Perangkat keras	25
<u>3.4.2.</u>	Perangkat lunak.....	25
3.5.	Jadwal Pelaksanaan	26
BAB IV	HASIL DAN PEMBAHASAN	27
4.1.	Data	27
4.2.	Import Library Package	31
4.3.	<i>Preprocessing</i>	32
<u>4.3.1</u>	<i>Data Cleaning</i>	32
<u>4.3.2</u>	<i>Exploratory Data Analysis</i>	34
<u>4.3.3</u>	Feature Selection	45
<u>4.3.4</u>	<i>Encoding</i>	45
4.4.	Pemodelan	51
4.5.	Evaluasi	55
BAB V	PENUTUP.....	78
5.1.	Kesimpulan.....	78
5.2.	Saran.....	79
DAFTAR PUSTAKA	80
DAFTAR RIWAYAT HIDUP	84
LAMPIRAN	86

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix (Prasojo dan Haryatmi, 2021)	11
Tabel 3. 1 Data Atribut.....	22
Tabel 3. 2 Jadwal Penelitian.....	26
Tabel 4. 1 Data pinjaman kredit nasabah bank di Jakarta tahun 2014 – 2021	29
Tabel 4. 2 Import Library Package	31
Tabel 4. 3 Jumlah Missing Value pada kolom data kredit nasabah bank di Jakarta	33
Tabel 4. 4 Nasabah kredit bank yang terverifikasi dan tidak terverifikasi.....	33
Tabel 4. 5 Jumlah Missing Value setelah dilakukan drop pada data null	34
Tabel 4.6 Persebaran data pada fitur target Loan Status	35
Tabel 4.7 Pembagian Loan Status antara good loan dan bad loan.....	37
Tabel 4.8 Persebaran data good loan dan bad loan	37
Tabel 4. 9 Persebaran data term	39
Tabel 4.10 Persebaran data Home Ownership	40
Tabel 4.11 Persebaran data Employee Length.....	41
Tabel 4.12 Persebaran data purpose.....	43
Tabel 4.13 Encoding Value loan status.....	46
Tabel 4.14 Encoding Value Employee length.....	47
Tabel 4.15 Encoding Value Home Ownership	48
Tabel 4.16 Encoding Value Verification Status.....	49
Tabel 4. 17 Encoding Value Term	49
Tabel 4.18 Encoding Value Purpose	50
Tabel 4.19 Variabel Data Input dan Data Output	51
Tabel 4.20 Contoh Jumlah Kasus terhadap data set.....	54
Tabel 4.21 Nilai Confusion Matrix data latih 80%	57
Tabel 4.22 Nilai Confusion Matrix data uji 20%	60
Tabel 4.23 Nilai Confusion Matrix data latih 70%	62
Tabel 4.24 Nilai Confusion Matrix data uji 30%	65
Tabel 4.25 Nilai Confusion Matrix data latih 60%	67
Tabel 4.26 Nilai Confusion Matrix data uji 40%	69

Tabel 4.27 Evaluasi model pada rasio data latih 80% dan data uji 20%	72
Tabel 4.28 Evaluasi model pada rasio data latih 70% dan data uji 30%	74
Tabel 4.29 Evaluasi model pada rasio data latih 60% dan data uji 40%	75
Tabel 4.30 Perbandingan Hasil Evaluasi Model keseluruhan.....	76

DAFTAR GAMBAR

Gambar 2. 1 Survei Stackoverflow tahun 2022 terkait teknologi populer.....	8
Gambar 2. 2 Tahap Proses KDD (Arifin, Enri, & Sulistiyowati, 2021).	8
Gambar 2. 3 Alur kerja Random Forest (Suliztia, 2020).	11
Gambar 2.4 Kurva ROC (Kristiawan, 2021)	14
Gambar 2.5 Kurva AUC (Kristiawan, 2021)	15
Gambar 3. 1 Tahapan Penelitian	21
Gambar 4.1 Kode program pengklasifikasian kolom loan status menjadi good loan dan bad loan	36
Gambar 4.2 Visualisasi pie chart loan status antara good loan dengan bad loan..	38
Gambar 4.3 Visualisasi bar chart term antara 36 months dan 60 months.....	39
Gambar 4.4 Visualisasi bar chart home ownership antara own, rent, dan mortgage	41
Gambar 4.5 Visualisasi bar chart employee length dari <1 year sampai 10+ years	42
Gambar 4.6 Visualisasi bar chart purpose.....	44
Gambar 4.7 Visualisasi heatmap korelasi Pearson	45
Gambar 4.8 Kode program untuk encoding kolom loan status.....	46
Gambar 4.9 Kode program untuk encoding kolom employee length	47
Gambar 4.10 Kode program untuk encoding kolom home ownership	48
Gambar 4.11 Kode program untuk encoding kolom verification status	49
Gambar 4.12 Kode program untuk encoding kolom term	49
Gambar 4.13 Kode program untuk encoding kolom purpose	50
Gambar 4.14 Gambaran trees dari data latih dan data uji	53
Gambar 4.15 Confusion Matrix rasio data latih 80%	56
Gambar 4.16 Confusion Matrix rasio data uji 20%	59
Gambar 4.17 Confusion Matrix rasio data latih 70%	61
Gambar 4.18 Confusion Matrix rasio data uji 30%	64
Gambar 4.19 Confusion Matrix rasio data latih 60%	66
Gambar 4.20 Confusion Matrix rasio data uji 40%	69
Gambar 4.21 Kurva ROC.....	76