

IMPLEMENTASI MODEL HYBRID RANDOM FOREST-XGBOOST UNTUK KLASIFIKASI SKOR KREDIT PADA DATA TIDAK SEIMBANG

SKRIPSI

Oleh:

ARTOLITO CHANDRA

NIM. 221110767

BRYAN YAPITER

NIM. 221110930

NICHOLAS TIO

NIM. 221111917



**PROGRAM STUDI S-1 TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2026**

IMPLEMENTATION OF A HYBRID RANDOM FOREST-XGBOOST MODEL FOR CREDIT SCORE CLASSIFICATION ON IMBALANCED DATA

UNDERGRADUATE THESIS

By:

**ARTOLITO CHANDRA
ID NUMBER. 221110767
BRYAN YAPITER
ID NUMBER. 221110930
NICHOLAS TIO
ID NUMBER. 221111917**



**UNDERGRADUATE PROGRAM OF COMPUTER SCIENCE
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2026**

LEMBARAN PENGESAHAN

**IMPLEMENTASI MODEL HYBRID RANDOM FOREST-XGBOOST
UNTUK KLASIFIKASI SKOR KREDIT PADA DATA TIDAK
SEIMBANG**

SKRIPSI

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Sarjana
Program Studi S-1 Teknik Informatika

Oleh:

ARTOLITO CHANDRA

NIM. 221110767

BRYAN YAPITER

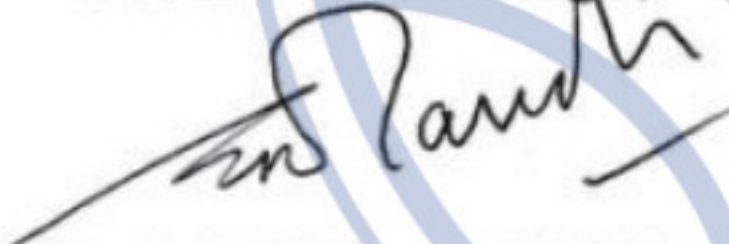
NIM. 221110930

NICHOLAS TIO

NIM. 221111917

Disetujui Oleh:

Dosen Pembimbing I,



Andri, S.Kom., M.T.I.

Dosen Pembimbing II,



Heru Kurniawan, S.Kom., M.Kom.

Medan, 23 Juli 2026

Diketahui dan Disahkan Oleh:



Ketua Program Studi
S-1 Teknik Informatika

Carles Juliandy, S.Kom., M.Kom.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Artolito Chandra

NIM : 221110767

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Implementasi Model Hybrid Random Forest-XGBoost untuk Klasifikasi Skor Kredit pada Data Tidak Seimbang

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 29 Juni 2026

Saya yang membuat pernyataan,



Artolito Chandra

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Bryan Yapiter

NIM : 221110930

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Implementasi Model Hybrid Random Forest-XGBoost untuk Klasifikasi Skor Kredit pada Data Tidak Seimbang

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 29 Juni 2026

Saya yang membuat pernyataan,



Bryan Yapiter

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Nicholas Tio

NIM : 221111917

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Implementasi Model Hybrid Random Forest-XGBoost untuk Klasifikasi Skor Kredit pada Data Tidak Seimbang

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 29 Juni 2026

Saya yang membuat pernyataan,



Nicholas Tio

IMPLEMENTASI MODEL HYBRID RANDOM FOREST-XGBOOST UNTUK KLASIFIKASI SKOR KREDIT PADA DATA TIDAK SEIMBANG

Abstrak

Penilaian skor kredit merupakan instrumen krusial bagi lembaga keuangan untuk mengelola risiko, namun dataset yang digunakan umumnya tidak seimbang (imbalanced) sehingga menyebabkan bias pada model klasifikasi konvensional. Penelitian ini mengimplementasikan model hybrid Random Forest-XGBoost dengan pendekatan stacking menggunakan Logistic Regression sebagai meta-learner untuk mengklasifikasikan skor kredit (Good, Standard, Poor) pada data tidak seimbang. Dataset Credit Score Classification dari Kaggle (100.000 sampel) diproses melalui tahapan preprocessing (penanganan missing value, encoding, scaling, dan analisis korelasi), pembagian data stratified 80:20, serta evaluasi menggunakan metrik accuracy, precision, recall, F1-score (macro), dan AUC-ROC. Hasil pengujian menunjukkan model hybrid stacking mencapai akurasi 72,78%, F1-score macro 0,7056, dan AUC macro 0,8737, lebih unggul dibandingkan Random Forest (71,56%) dan XGBoost (71,41%). Model hybrid ini berhasil memanfaatkan kekuatan bagging dari Random Forest dan boosting dari XGBoost secara komplementer, sehingga memberikan performa yang lebih baik dan stabil dalam menangani ketidakseimbangan kelas. Penelitian ini menawarkan solusi efektif untuk meningkatkan akurasi penilaian risiko kredit pada data keuangan yang kompleks.

Kata kunci: credit scoring, random forest, xgboost, hybrid stacking, data tidak seimbang

Abstract

Credit scoring is a crucial instrument for financial institutions in managing credit risk. However, credit datasets are typically imbalanced, causing bias in conventional classification models. This study implements a hybrid Random Forest-XGBoost model using a stacking approach with Logistic Regression as the meta-learner to classify credit scores (Good, Standard, Poor) on imbalanced data. The Credit Score Classification dataset from Kaggle (100,000 samples) underwent preprocessing (missing value handling, encoding, scaling, and correlation analysis), stratified 80:20 split, and evaluation using accuracy, precision, recall, macro F1-score, and AUC-ROC metrics. Experimental results show that the hybrid stacking model achieved an accuracy of 72.78%, macro F1-score of 0.7056, and macro AUC of 0.8737, outperforming standalone Random Forest (71.56%) and XGBoost (71.41%). The hybrid model effectively combines the strengths of bagging from Random Forest and boosting from XGBoost, delivering superior and more stable performance in handling class imbalance. This study provides an effective solution to improve credit risk assessment accuracy on complex financial data.

Keywords: credit scoring, random forest, xgboost, hybrid stacking, imbalanced data

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala Rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik, yang berjudul “Implementasi Model Hybrid Random Forest-XGBoost untuk Klasifikasi Skor Kredit pada Data Tidak Seimbang”

Tugas Akhir ini disusun sebagai bagian dari persyaratan kurikulum pada Program Studi S-I Teknik Informatika, Universitas Mikroskil Medan.

Penulis juga menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Andri, S.Kom., M.T.I., selaku Dosen Pembimbing I.
2. Bapak Heru Kurniawan, S.Kom., M.Kom., selaku Dosen Pembimbing II.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Sunaryo Winardi, S.Kom., M.T., selaku Dekan Fakultas Informatika Universitas Mikroskil Medan.
5. Bapak Carles Juliandy, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Teknik Informatika Fakultas Informatika Universitas Mikroskil Medan.
6. Bapak dan Ibu Dosen Universitas Mikroskil yang telah memberikan ilmunya selama menempuh pendidikan di Universitas Mikroskil Medan.
7. Keluarga penulis, khususnya orang tua, yang senantiasa memberikan dukungan moral dan materil sepanjang masa studi hingga Tugas Akhir ini terselesaikan.
8. Rekan-rekan serta semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan bantuan dan dukungan selama proses penyusunan Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih memiliki keterbatasan dan kekurangan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang bersifat membangun demi penyempurnaan di masa mendatang. Akhir kata, semoga Tugas Akhir ini dapat memberikan manfaat bagi pembaca dan perkembangan ilmu pengetahuan. Terima Kasih.

Medan, 19 Juni 2026

Penulis,

Artolito Chandra

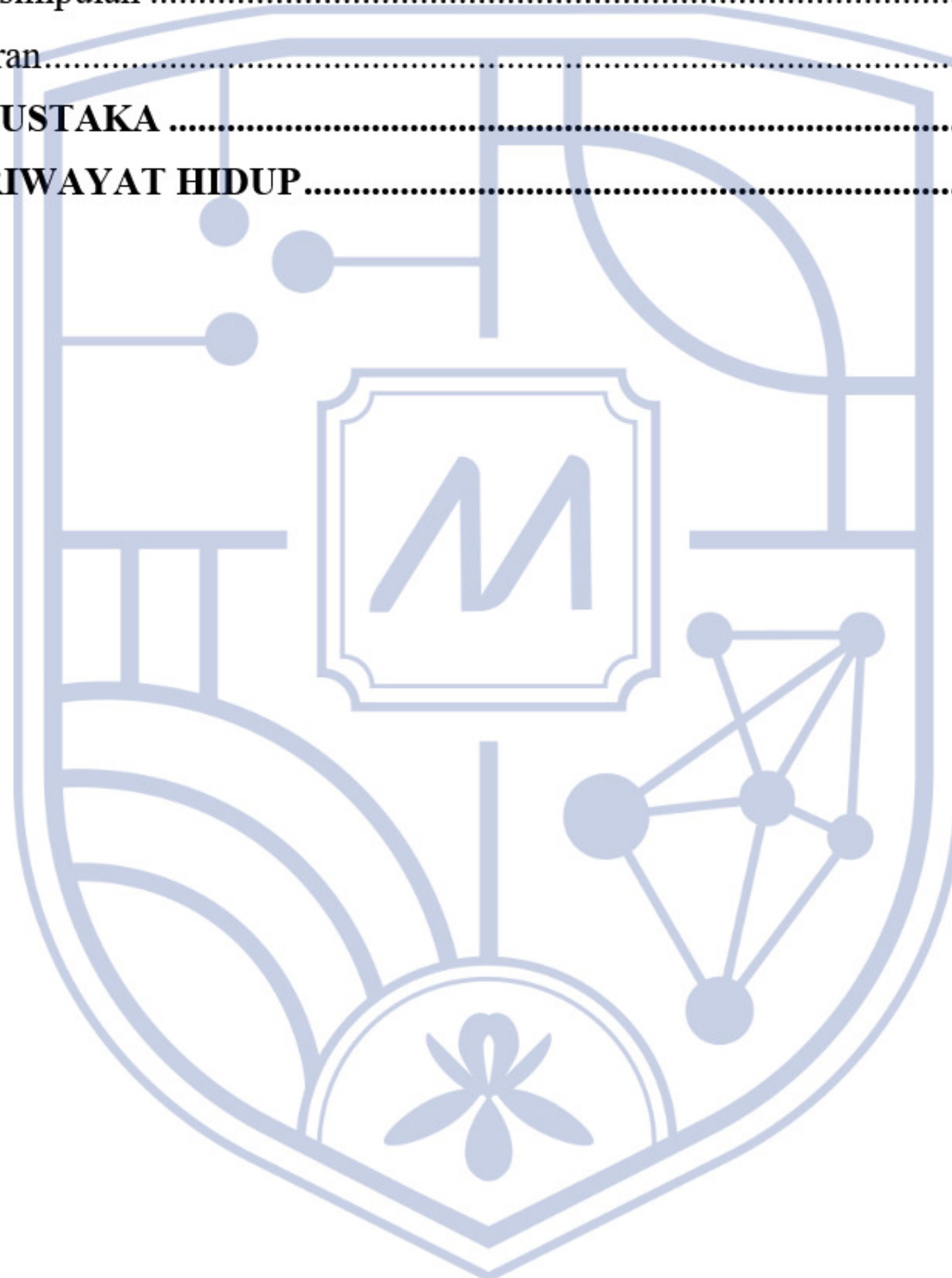
Bryan Yapiter

Nicholas Tio

DAFTAR ISI

ABSTRAK.....	i
KATA PENGANTAR	ii
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL.....	vii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup.....	4
BAB II KAJIAN LITERATUR.....	5
2.1 Skor Kredit.....	5
2.2 Ketidakseimbangan Data	6
2.3 <i>Preprocessing</i>	7
2.3.1 <i>Missing Value Imputation</i>	7
2.3.2 <i>Matrix Correlation Analysis</i>	8
2.3.3 <i>Data Cleansing</i>	9
2.3.4 <i>Data Encoding</i>	10
2.3.5 <i>Data Scaling</i>	10
2.4 <i>Machine Learning</i>	11
2.5 <i>Ensemble Learning</i>	12
2.6 <i>Random Forest</i>	13
2.7 <i>XGBoost</i>	15
2.8 <i>Hybrid</i>	17
2.9 <i>Stacking</i>	17
2.10 <i>Confusion Matrix</i>	19
BAB III TAHAPAN PELAKSANAAN	22
3.1 Kerangka Tahapan Pelaksanaan.....	22
3.2 Analisis.....	22
3.3 Training / Test Model	32
3.4 Perancangan Tampilan.....	45
3.4.1 Tab Analisis Model.....	46

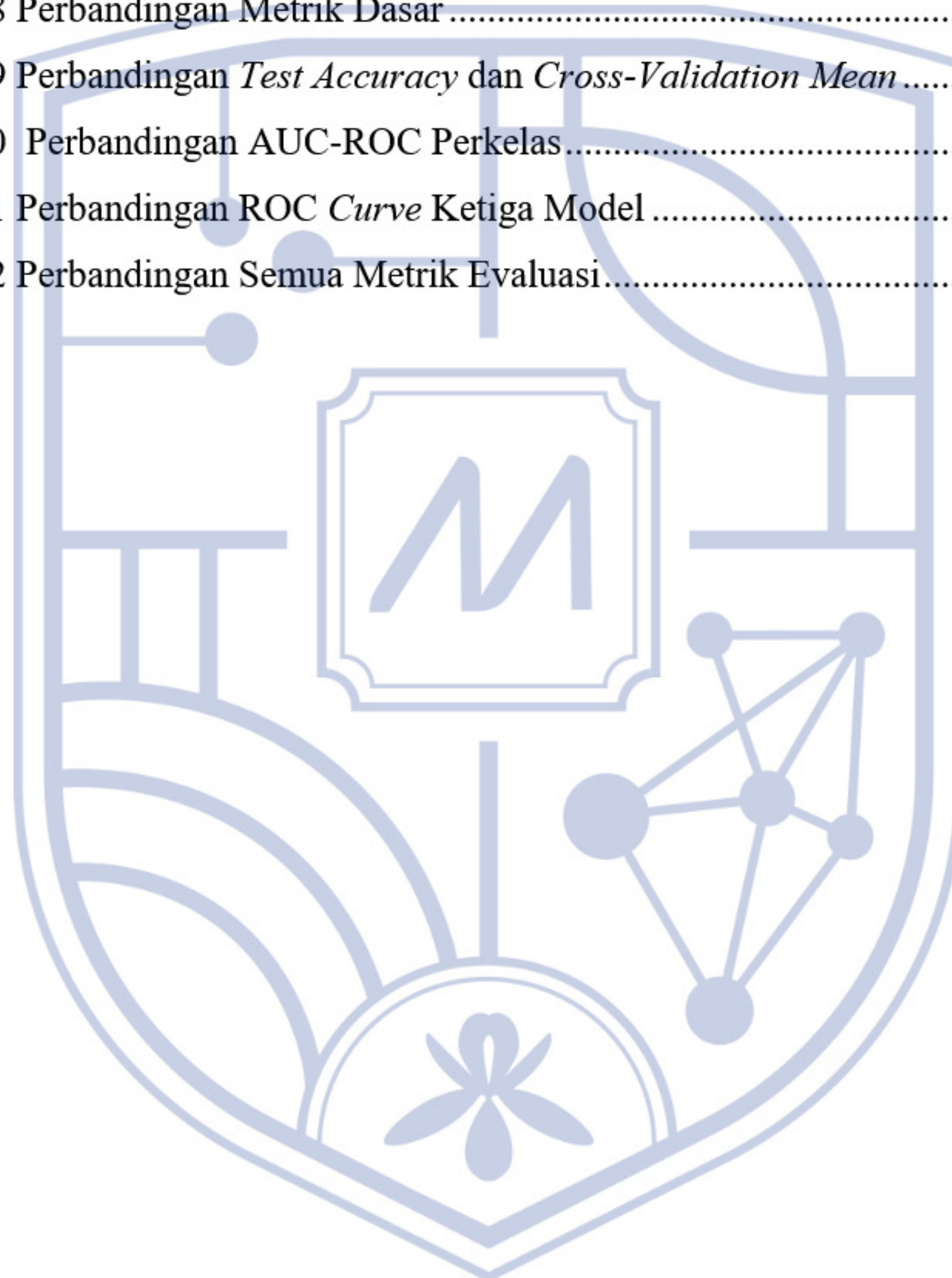
3.4.2 Tab Uji Model.....	48
BAB IV HASIL DAN PEMBAHASAN	54
4.1 Hasil	54
4.1.1 Tab Analisis Model.....	54
4.1.2 Tab Uji Model.....	59
4.2 Pembahasan.....	65
BAB V PENUTUP	72
5.1 Kesimpulan	72
5.2 Saran.....	72
DAFTAR PUSTAKA	74
DAFTAR RIWAYAT HIDUP.....	78



DAFTAR GAMBAR

Gambar 2.1 Konsep <i>Stacking</i> [40].....	18
Gambar 2.2 <i>Confusion Matrix</i> [42]	19
Gambar 2.3 ROC Curve [41].....	21
Gambar 2.4 AUC [41]	21
Gambar 3.1 <i>Flowchart</i> Analisis Proses	23
Gambar 3.2 <i>Sample Dataset</i>	24
Gambar 3.3 <i>Flowchart</i> Tahapan <i>Preprocessing</i>	25
Gambar 3.4 Matriks Korelasi Sebelum Standarisasi (<i>Heatmap</i>).....	27
Gambar 3.5 Matriks Korelasi Setelah Standarisasi (<i>Heatmap</i>).....	28
Gambar 3.6 <i>Flowchart</i> Training dan Testing Model.....	32
Gambar 3.7 Tampilan Analisis Model	46
Gambar 3.8 Tampilan Analisis Model	47
Gambar 3.9 Tampilan Uji Model	49
Gambar 3.10 <i>Header</i>	50
Gambar 3.11 <i>Form Input</i>	50
Gambar 3.12 <i>Input Form</i>	51
Gambar 3.13 Tombol <i>Randomize</i>	51
Gambar 3.14 Tombol Prediksi.....	52
Gambar 3.15 Hasil Prediksi.....	52
Gambar 3.16 Ringkasan Data.....	52
Gambar 3.17 Konsensus Model.....	53
Gambar 4.1 Tab Analisis Model 1.....	55
Gambar 4.2 Tab Analisis Model 2.....	55
Gambar 4.3 <i>Header</i> Tab Analisis Model	56
Gambar 4.4 Informasi Dataset.....	56
Gambar 4.5 Metrik <i>Performance Model</i>	57
Gambar 4.6 <i>Confusion Matrix</i>	58
Gambar 4.7 Sepuluh Sampel Prediksi (Data Test)	58
Gambar 4.8 Kesimpulan <i>Tab Analisis Model</i>	59
Gambar 4.9 Tab Uji Model.....	60
Gambar 4.10 <i>Header</i> Tab Uji Model	60

Gambar 4.11 <i>Form Input Data Nasabah</i>	61
Gambar 4.12 <i>Form Pengisian Data Informasi Pribadi</i>	61
Gambar 4.13 <i>Form Pengisian Data Informasi Keuangan</i>	62
Gambar 4.14 <i>Form Pengisian Data Informasi Kredit</i>	62
Gambar 4.15 Tombol <i>Randomize Data</i>	63
Gambar 4.16 Tombol <i>Prediksi Credit Score</i>	63
Gambar 4.17 Hasil <i>Prediksi</i>	64
Gambar 4.18 <i>Perbandingan Metrik Dasar</i>	66
Gambar 4.19 <i>Perbandingan Test Accuracy dan Cross-Validation Mean</i>	67
Gambar 4.20 <i>Perbandingan AUC-ROC Perkelas</i>	68
Gambar 4.21 <i>Perbandingan ROC Curve Ketiga Model</i>	68
Gambar 4.22 <i>Perbandingan Semua Metrik Evaluasi</i>	69



DAFTAR TABEL

Tabel 3.1 Distribusi Kelas Target.....	24
Tabel 3.2 Contoh Penanganan <i>Missing Values</i>	26
Tabel 3.3 Contoh Konversi Tipe Data.....	29
Tabel 3.4 Contoh Penanganan <i>Placeholder</i>	30
Tabel 3.5 Contoh Penanganan Nilai Negatif.....	30
Tabel 3.6 Hasil <i>Scaling (Z-Score)</i>	31
Tabel 3.7 Contoh Transformasi Fitur <i>Occupation</i> menggunakan <i>One-Hot Encoding</i>	31
Tabel 3.8 Hasil <i>Encoding (LabelEncoder)</i>	31
Tabel 3.9 Pembagian <i>Dataset</i> untuk <i>Training</i> dan <i>Testing</i>	33
Tabel 3.10 Delapan Data Sampel Acak.....	33
Tabel 3.11 Simulasi Perhitungan 100 Iterasi.....	39
Tabel 3.12 Parameter Model <i>Random Forest</i> yang Diinisialisasi	41
Tabel 3.13 Parameter Model XGBoost yang Diinisialisasi.....	42
Tabel 3.14 Parameter Model <i>Base Learner</i> RF yang Diinisialisasi	42
Tabel 3.15 Parameter Model <i>Base Learner</i> XGboost yang Diinisialisasi.....	43
Tabel 3.16 Parameter Model <i>Logistic Regression</i> yang Diinisialisasi	43
Tabel 3.17 Parameter Model <i>StackingClassifier</i> yang Diinisialisasi	43
Tabel 3.18 Hasil Metrik Evaluasi Model.....	45
Tabel 4.1 Contoh Data Input Nasabah.....	63
Tabel 4.2 Perbandingan Metrik Dasar	66
Tabel 4.3 Perbandingan <i>Test Accuracy</i> vs <i>Cross-Validation Mean</i>	67
Tabel 4.4 Tabel Perbandingan AUC-ROC	67
Tabel 4.5 Ringkasan Semua Metrik Evaluasi.....	68
Tabel 4.6 <i>Confusion Matrix Random Forest</i>	69
Tabel 4.7 <i>Confusion Matrix XGBoost</i>	70
Tabel 4.8 <i>Confusion Matrix Hybrid Stacking</i>	70