OPTIMALISASI MODEL *GRADIENT BOOSTING MACHINE* DENGAN *GRID SEARCH* UNTUK MENENTUKAN KELAYAKAN PEMBERIAN KREDIT BANK

WAHIDA RAMADHANIA
H1091211043

SKRIPSI

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

OPTIMALISASI MODEL *GRADIENT BOOSTING MACHINE* DENGAN *GRID SEARCH* UNTUK MENENTUKAN KELAYAKAN PEMBERIAN KREDIT BANK

WAHIDA RAMADHANIA
H1091211043

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika pada Program Studi Statistika

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

LEMBAR PENGESAHAN

Judul Tugas Akhir	: Optimalisasi Model *Gradient Boosting Machine* Dengan *Grid Search* untuk Menentukan Kelayakan Pemberian Kredit Bank
Nama Mahasiswa	: Wahida Ramadhania
NIM	: H1091211043
Jurusan/Program Studi	: Matematika/Statistika
Tanggal Lulus	: 19 Maret 2025
SK Pembimbing	: No. 2618/UN22.8/TD.06/2024/Tanggal 2 September 2024
SK Penguji	: No. 806/ UN22.8/TD.06/2025/Tanggal 10 Maret 2025

Dosen Pembimbing

Pembimbing I	Pembimbing II
Neva Satyahadewi, M.Sc., CRA., CRP., CRMP	Nurfitri Imro'ah, M.Si.
NIP. 198212042005012001	NIP. 198907182019032021

Dosen Penguji

Ketua Penguji	Anggota Penguji
Shantika Martha, M.Si.	Drs. Helmi, M.Si.
NIP. 198403082008122003	NIP. 196410171998021001

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Neva Satyahadewi, M.Sc., CRA., CRP., CRMP.	Nurfitri Imro'ah, M.Si.
NIP. 198212042005012001	NIP. 198907182019032021

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, M.Si.
NIP.197108022000031001

ii

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama : Wahida Ramadhania

NIM : H1091211043

Program Studi/ Jurusan : Statistika/Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyatan ini untuk digunakan sebagaimana mestinya.

Pontianak, 19 Maret 2025

Wahida Ramadhania

Optimalisasi Model *Gradient Boosting Machine* Dengan *Grid Search* untuk Menentukan Kelayakan Pemberian Kredit Bank

Abstrak

Pengambilan keputusan dalam menentukan kelayakan pemberian kredit melibatkan berbagai faktor kompleks seperti riwayat kredit, pendapatan pekerjaan, dan jumlah tanggungan. Tujuan dari penelitian ini ialah menentukan tingkat akurasi model *Gradient Boosting Machine* (GBM) menggunakan teknik optimasi parameter dengan *Grid Search* serta menganalisis hasil klasifikasi kelayakan pemberian kredit. Pendekatan yang digunakan termasuk dalam metode *supervised learning*. *Supervised Learning* merupakan salah satu pembelajaran mesin yang bertujuan untuk menyelesaikan masalah seperti klasifikasi dengan mengindentifikasi variabel target menggunakan model hasil pelatihan dari data berkategori. GBM merupakan salah satu algoritma klasifikasi *supervised learning* berbasis *decision tree*. Permasalahan pada GBM adalah kecenderungan mengalami *overfitting* yang disebabkan oleh pengaturan *hyperparameter* yang tidak optimal. Hal ini dapat diatasi dengan teknik optimasi parameter seperti *grid search*. Data yang digunakan yaitu data historis nasabah yang berasal dari *website* www.kaggle.com sebanyak 730 data debitur dengan 162 pinjaman yang disetujui dan 568 pinjaman tidak disetujui. Langkah penelitian yang pertama yaitu *preprocessing* data, kemudian membagi data menjadi data *training* dan data *testing* dengan proporsi 80:20, menangani *imbalance class* pada data training menggunakan *Random Oversampling (ROS)*, membuat model GBM tanpa *grid search* dan model GBM menggunakan *grid search*. Hasil penelitian didapatkan nilai untuk membentuk model GBM dengan *n trees* = 100, *interaction depth* = 3, dan *learning rate* = 0,1 dengan akurasi model yaitu 83,43%, sedangkan untuk membentuk model GBM optimal menggunakan optimasi parameter *grid search* yaitu dengan *n trees* = 50, *interaction depth* = 5, dan *learning rate* = 0,3 diperoleh akurasi model yaitu sebesar 95,15%. Hal ini menunjukkan bahwa metode GBM menggunakan optimasi *grid search* menghasilkan kinerja yang lebih baik dibandingkan tanpa optimasi.

Kata kunci: ROS, Optimasi Parameter, *Supervised Learning*

Optimization Of Gradient Boosting Machine Model With Grid Search to Determine The Feasibility Of Bank Credit Granting

Abstract

Decision-making in determining creditworthiness involves various complex factors such as credit history, employment income, and number of dependents. The purpose of this research is to determine the accuracy of the Gradient Boosting Machine (GBM) model using parameter optimization techniques with Grid Search and analyze the results of the classification of creditworthiness. The approach used falls under the supervised learning method. Supervised Learning is a type of machine learning aimed at solving problems such as classification by identifying the target variable using a model trained on categorized data. GBM is one of the supervised learning classification algorithms based on decision trees. The problem with GBM is its tendency to overfit due to suboptimal hyperparameter settings. This can be addressed with parameter optimization techniques such as grid search. The data used is historical customer data source from the website www.kaggle.com consisting of 730 debtor records with approved and rejected loans. The first step of the research is data preprocessing, then dividing the data into training data and testing data, creating a GBM model without grid search and a GBM model using grid search. The research results obtained values to form the GBM model with n trees = 100, interaction depth = 3, and learning rate = 0.1 with a model accuracy of 83.43%, while to form the optimal GBM model using grid search parameter optimization with n trees = 50, interaction depth = 5, and learning rate = 0.3, a model accuracy of 95.15% was achieved. This shows that the GBM method using grid search optimization yields better performance compared to without optimization.

***Keywords**: ROS, Parameter Optimization, Supervised Learning*

PRAKATA

Dengan mengucap puji dan syukur penulis panjatkan kepada Allah Subhanahu wata'ala atas segala rahmat dan hidayah-Nya, serta perlindungan, pertolongan, dan ridho-Nya, tidak terlupakan iringan salam dan sholawat bagi junjungan kita Nabi Muhammad Sallahu Alaihi Wassalam, sehingga penulis dapat menyelesaikan Tugas Akhir yang berujudul "**Optimalisasi Model *Gradient Boosting Machine* Dengan *Grid Search* Untuk Menentukan Kelayakan Pemberian Kredit Bank**" dengan tepat pada waktunya. Penulisan skripsi ini bertujuan untuk memenuhi syarat memperoleh gelar Sarjana Statistika pada Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Tanjungpura.

Dalam penulisan skripsi ini tidak luput dari bimbingan, dukungan, serta motivasi dan doa dari berbagai pihak, sehingga penulis dapat menyelesaikan skripsi dengan lancar. Penulis menyampaikan penghargaan dan terima kasih kepada:

1. Kedua orang tua penulis, Mulyadi dan Halwa yang selalu mendoakan agar penulis segera wisuda serta kedua adik, Waldi Ramadhan dan Windi Novisha yang memberikan dukungan dan semangat yang luar biasa.
2. Ibu Neva Satyahadewi, M.Sc., CRA., CRP., CRMP. selaku dosen pembimbing pertama yang telah mengarahkan dan membimbing selama penulisan skripsi ini.
3. Ibu Nurfitri Imro'ah, M.Si. selaku dosen pembimbing kedua yang telah mengarahkan dan membimbing selama penulisan skripsi ini.
4. Shantika Martha, M.Si. selaku dosen penguji pertama yang telah memberikan saran, masukan, serta koreksi dalam penulisan skripsi ini.
5. Drs. Helmi, M.Si. selaku dosen penguji kedua yang telah memberikan saran, masukan, serta koreksi dalam penulisan skripsi ini.
6. Bapak/Ibu Dosen Jurusan Matematika yang telah memberikan ilmu bermanfaat selama proses perkuliahan.
7. Sahabat-sahabat penulis, Ghina, Nanda, Hera, Della, Putri, Ulya, Luis, Muti, dan Siska yang telah membersamai dan saling menyemangati selama perkuliahan hingga saat proses penulisan skripsi ini selesai.

8. Teman-teman ”Expos” Statistika Angkatan 2021 yang membersamai penulis selama proses kuliah di Program Studi Statistika FMIPA Untan.
9. Sahabat sekaligus saudara penulis yaitu Ningum, Wanda, Pida, Dania, Wulan, Nova dan Amelia yang selalu mendukung dan menyemangati penulis agar segera mendapatkan gelar.
10. Semua pihak yang telah memberikan semangat kepada penulis, namun tidak dapat disebutkan satu persatu.

Semoga segala hal yang telah diberikan kepada penulis menjadi amal ibadah serta diberkahi Allah Subhanahu wata’ala. Penulis menyadari penyusunan skripsi ini jauh dari kata sempurna maka dari itu penulis menerima saran dan kritikan yang membangun.

Pontianak, 19 Maret 2025

Wahida Ramadhania

DAFTAR ISI

LEMBAR PENGESAHAN ..... **Error! Bookmark not defined.**
PERNYATAAN INTEGRITAS AKADEMIK iii
Abstrak iv
Abstract v
PRAKATA vi
DAFTAR ISI viii
DAFTAR TABEL x
DAFTAR GAMBAR xi
DAFTAR SIMBOL xii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan Penelitian 3
1.4 Batasan Masalah 3
1.5 Tinjauan Pustaka 4
1.6 Metodologi Penelitian 5
BAB II LANDASAN TEORI 8
2.1 Bank 8
2.2 *Machine Learning* 8
2.3 Klasifikasi 9
2.4 *Binary Cross Entropy Loss Function* 10
2.5 *Hyperparameter tuning* dengan *Grid Search* 11
BAB III *GRADIENT BOOSTING MACHINE – GRID SEARCH* 12
3.1 *Gradient Boosting Machine* (GBM) 12
3.2 *Grid Search* 14
3.3 Evaluasi Kinerja Model Klasifikasi 15
3.4 *Feature Importance* 16
BAB IV HASIL DAN PEMBAHASAN 18

4.1 Data Penelitian 18
4.2 *Preprocessing* Data 19
4.3 Pembagian Data *Training* dan *Testing* 22
4.4 Hasil *Random Oversampling* (ROS) 22
4.5 Pembentukan Model *Gradient Boosting Machine* tanpa *Grid Search* 23
4.6 Pembetukan Model *Gradient Boosting Machine* dengan *Grid Search* 28
4.7 Perbandingan Kinerja Model Klasifikasi GBM tanpa *Grid Search* dan GBM dengan *Grid Search* 32
BAB V KESIMPULAN 34
DAFTAR PUSTAKA 35
LAMPIRAN 39

DAFTAR TABEL

Tabel 2.1 *Hyperparameter Tuning* dengan *Grid Search* 11
Tabel 3.1 *Confusion Matrix* 16
Tabel 4.1 Statistik Deskriptif 19
Tabel 4.2 Keterangan Atribut 20
Tabel 4.3 Pelabelan Data Kategori 21
Tabel 4.4 Pembagian Data *Training* dan *Testing* 22
Tabel 4.5 Hasil ROS pada Data *Training* 23
Tabel 4.6 Hasil Akurasi model optimal GBM tanpa *Grid Search* 26
Tabel 4.7 *Confusion Matrix* model GBM tanpa *Grid Search* 26
Tabel 4.8 Nilai *features importance* GBM tanpa *Grid Search* 28
Tabel 4.9 Parameter optimasi model GBM 29
Tabel 4.10 Hasil Akurasi terbaik model GBM dengan *Grid Search* 29
Tabel 4.11 *Confusion Matrix* model GBM dengan *Grid Search* 30
Tabel 4.12 Nilai *features importance* GBM dengan *Grid Search* 31
Tabel 4.13 Perbandingan Kinerja model GBM tanpa *Grid Search* dan GBM dengan *Grid Search* 32

DAFTAR GAMBAR

Gambar 3.1 Skema Kerja GBM 12
Gambar 3.2 Ilustrasi *k-fold cross validation* 15
Gambar 4. 1 Diagram lingkaran status debitur 18
Gambar 4. 2 Pohon Iterasi Pertama, Kedua, dan Ketiga 25
Gambar 4. 3 *Features Importance* GBM tanpa *Grid Search* 27
Gambar 4. 4 *Features Importance* GBM dengan *Grid Search* 31

DAFTAR SIMBOL

y_i	:	Nilai aktual variabel target pada data ke-*i*
i		Urutan data, dengan nilai $1, 2, 3, \ldots, n$
x	:	Nilai proporsi kelas positif yaitu untuk pinjaman disetujui (1)
$1 - x$	:	Nilai proporsi kelas negatif yaitu untuk pinjaman tidak disetujui (0)
L	:	Fungsi kerugian
$F_0(x)$	:	Nilai prediksi konstan awal
m	:	Indeks iterasi pohon keputusan
M	:	Total banyak iterasi pada pohon keputusan
r_{im}	:	Residual pada iterasi m untuk data ke-*i*
p_i	:	Nilai probabilitas prediksi pada data ke-*i*
$F_{m-1}(x)$	:	Nilai prediksi keseluruhan model pada iterasi sebelumnya
$F_m(x)$	:	Nilai prediksi baru untuk masing-masing data *Training* pada pengulangan ke - *m*
R_{jm}	:	Himpunan sampel pada *j* indeks simpul daun untuk indeks iterasi pohon keputusan ke-*m*
j	:	Indeks simpul daun dalam pohon keputusan
J_m	:	Banyak simpul daun yang dihasilkan pohon keputusan hingga ke-*m* dalam pohon keputusan
Y_{jm}	:	Prediksi target pada iterasi ke-*j* indeks simpul daun untuk indeks iterasi pohon keputusan ke-*m*

BAB I
PENDAHULUAN

1.1 Latar Belakang

Bank sangat efektif dalam memfasilitasi hubungan antara pihak-pihak ekonomi yang memiliki sumber daya keuangan berlebihan dengan pihak-pihak yang memerlukan pendanaan. Bank juga sebagai tempat penyimpanan yang efisien dan bermanfaat bagi masyarakat, juga mendukung kelancaran transaksi keuangan di semua sektor ekonomi (Simatupang, Yanti, dan Mardila, 2021). Tujuan utama lembaga perbankan, baik konvensional maupun syariah, adalah untuk mengumpulkan dana dan mengalokasikan dana tersebut untuk berbagai keperluan seperti investasi, modal kerja, dan konsumsi. Bank sebagai institusi keuangan yang fokus pada penerimaan simpanan berupa giro, tabungan, dan deposito, serta sebagai sumber pinjaman (kredit) bagi masyarakat yang memerlukannya.

Secara mendasar, kredit merujuk pada transaksi dimana pihak pemberi kredit menyerahkan uang, barang, atau jasa kepada pihak penerima kredit, dengan perjanjian untuk mengembalikan jumlah tersebut dalam jangka waktu tertentu, biasanya dengan tambahan biaya atau pokok (Barus dan Erick, 2017). Berdasarkan ketentuan Undang-undang perbankan, yakni UU no. 7 tahun 1998, tentang kredit didefinisikan sebagai fasilitas penyaluran dana atau tagihan yang setara, berdasarkan kesepakatan serta persetujuan peminjaman di antara pihak bank dengan pihak lainnya, yang mengharuskan pihak peminjam untuk membayar kembali pinjamannya dengan tambahan bunga atau pokok dalam jangka waktu yang telah ditentukan atau disepakati.

Pengambilan keputusan dalam pemberian kredit merupakan proses yang kompleks karena melibatkan banyak faktor yang harus dipertimbangkan, seperti riwayat kredit, pendapatan, pekerjaan, dan lain sebagainya. Dalam konteks ini, penggunaan model klasifikasi menjadi solusi yang efektif untuk membantu bank dalam menentukan kelayakan pemberian kredit kepada nasabah (Mardiyyah, Rahaningsih, dan Ali, 2024). Klasifikasi merupakan proses analisis data yang mencari kelas yang sama dalam suatu objek data dan mengklasifikasikannya ke dalam salah satu kategori (kelas) yang telah ditentukan. *Machine learning* merupakan teknik model

klasifikasi yang mampu menangani berbagai jenis data dengan baik dan memberikan hasil klasifikasi yang akurat.

Salah satu metode dalam *Machine Learning* ialah metode *boosting* dikenal sebagai *gradient boosting* yang merupakan pendekatan *supervised learning* yang menggunakan *decision tree* sebagai dasarnya. Proses algoritma *gradient boosting* dimulai dengan pembuatan pohon klasifikasi pertama, yang kemudian terus disesuaikan dengan menyesuaikan pohon baru melalui upaya untuk meminimalkan fungsi kerugian (Natekin dan Knoll, 2013).

Gradient Boosting Machine (GBM) adalah algoritma *machine learning* yang dapat membangun sebuah model klasifikasi dengan menggabungkan serangkaian model lemah (*weak learners*) secara bertahap dan kemudian menggabungkannya untuk membentuk model yang lebih kuat dan akurat. Algoritma ini bekerja secara berulang-ulang dengan menambahkan model yang lemah menjadi model baru untuk memperbaiki kesalahan yang dibuat oleh model sebelumnya (Pratiwi, dkk, 2024). Dengan ini, dapat terbukti bahwa model GBM mampu menghasilkan model dengan tingkat akurasi yang lebih tinggi.

Metode klasifikasi pada *machine learning* sering ditemukan permasalahan dalam prosesnya, yaitu ketika salah satu kelasnya mempunyai jumlah data yang jauh berbeda dengan kelas lainnya pada data *training*. Permasalahan ini disebut sebagai *imbalance class* atau ketidakseimbangan kelas yang menyebabkan menurunnya kinerja klasifikasi pada kelas minoritas. Kinerja algoritma *machine learning* dievaluasi dengan akurasi hasil prediksi, namun hal ini tidak sesuai apabila terjadi ketidakseimbangan kelas. Model klasifikasi yang dibuat menggunakan data tidak seimbang berakibat menghasilkan akurasi klasifikasi pada kelas minoritas yang rendah (Fitriani, Yasin, dan Tarno, 2021).

Imbalance data merupakan suatu kondisi data yang tidak seimbang dengan jumlah data suatu kelas melebihi jumlah data pada kelas yang lain. Permasalahan kelas tidak seimbang (*imbalance*) dapat diatasi dengan melakukan teknik *resampling* yakni teknik distribusi data diseimbangkan kembali untuk mengurangi efek distribusi kelas tidak seimbang. *Random Oversampling* (ROS) adalah teknik penambahan data dari kelas minoritas ke dalam data *training* secara acak (Fitriani, Yasin, dan Tarno, 2021).

Proses ini diulang hingga jumlah data dari kelas minoritas setara dengan jumlah data dari kelas mayoritas. ROS bekerja untuk menyeimbangkan data sehingga proporsi data yang dihasilkan akan sama besar (*balance*).

Kekurangan dari algoritma GBM adalah kecenderungan untuk mengalami *overfitting* yang disebabkan oleh pengaturan *hyperparameter* yang tidak optimal. Oleh karena itu, diperlukan optimasi *hyperparameter* dalam GBM (Zhang, dkk, 2017). Pemilihan *hyperparameter* optimal dapat dilakukan menggunakan *grid search*. Beberapa metode algoritma yang melibatkan *hyperparameter* tersebut antara lain ialah GBM, *Random Forest*, dan *Support Vector Machine* (SVM).

Penelitian ini membahas bagaimana mengoptimalkan model GBM untuk menentukan kelayakan pemberian kredit. Dengan menggunakan teknik ROS dalam menangani *imbalance class* kemudian mengoptimalkan model GBM dengan optimasi *hyperparameter* seperti *Grid Search*, diharapkan model GBM dapat dioptimalkan untuk memberikan hasil klasifikasi yang lebih akurat dan tepat dalam menentukan kelayakan pemberian kredit.

1.2 Rumusan Masalah

Berdasarkan latar belakang didapat masalah utama yang ingin dipecahkan adalah:

1. Bagaimana tingkat akurasi model GBM dengan menggunakan teknik optimasi parameter *Grid Search*.
2. Bagaimana analisis hasil klasifikasi untuk menentukan kelayakan pemberian kredit menggunakan model terbaik.

1.3 Tujuan Penelitian

Dari rumusan masalah didapat tujuan dari penelitian ini ialah untuk:

1. Menentukan tingkat akurasi model GBM menggunakan teknik optimasi parameter *Grid Search*.
2. Menganalisis hasil klasifikasi kelayakan pemberian kredit menggunakan model terbaik.

1.4 Batasan Masalah

Berdasarkan tujuan penelitian didapat beberapa batasan masalah yaitu:

1. Penelitian menggunakan data sekunder, yaitu data historis nasabah pada periode 01 Januari 2022 hingga 31 Desember 2023 yang diambil dari *website* www.kaggle.com.
2. Pengaplikasian *software* R Studio sebagai alat bantu dalam mengimplementasikan dan mencari nilai akurasi pada model GBM.

1.5 Tinjauan Pustaka

Suryana, Warsito, dan Suparti (2021) melakukan penelitian yang membahas tentang penerapan *Gradient Boosting* dengan *Hyperopt* untuk memprediksi keberhasilan telemarketing Bank. Data dalam penelitian ini meliputi 20 variabel independen dan variabel dependen *biner* yang memiliki kelas 'ya' dan 'tidak'. Dari hasil analisis, disimpulkan bahwa *Hyperopt* dapat efektif digunakan untuk menyetel *hyperparameter* dengan rentang pencarian yang luas. Penggunaan metode *gradient boosting* untuk memprediksi keberhasilan telemarketing dengan *hyperparameter* yang di *tuning* menghasilkan akurasi sebesar 90,39%, presisi sebesar 94,91%, dan nilai AUC sebesar 0,939 pada data uji. Presentase ini sangat memuaskan, menunjukkan bahwa model dapat secara efektif mengklasifikasikan dan memprediksi kedua kelas dengan baik.

Wardhana, dkk, (2022) melakukan klasifikasi untuk kacang kering dengan metode *Gradient Boosting Machine*, *Random Forest* dan *Light* GBM untuk melihat model yang efektif. Pada saat *training* terbaik untuk model *Random Forest* dengan parameter pohon keputusan 50 dan *depth* 10, model *Gradient Boosting Machine* pada *learning rate* 1 dan model *Light* GBM pada *learning rate* 0,5 dan estimator 500. Hasil prediksi klasifikasi kelas untuk *Random Forest*, GBM dan *Light* GBM menghasilkan akurasi yang hampir sama. *Random Forest* memiliki akurasi *training* 96% dan GBM dengan akurasi 93%.

Handayani (2022) membahas tentang Model Algoritma *Gradient Boosting* untuk penentuan kelayakan pemberian kredit koperasi. Dari analisis ini, diperlukan sebuah model yang dapat melakukan klasifikasi dan prediksi terhadap debitur yang berpotensi mengalami masalah dalam pembayaran kredit. Ternyata, penggunaan model pohon keputusan mampu meningkatkan tingkat akurasi dalam menilai kelayakan kredit yang diajukan oleh calon debitur. Semakin banyak informasi atau

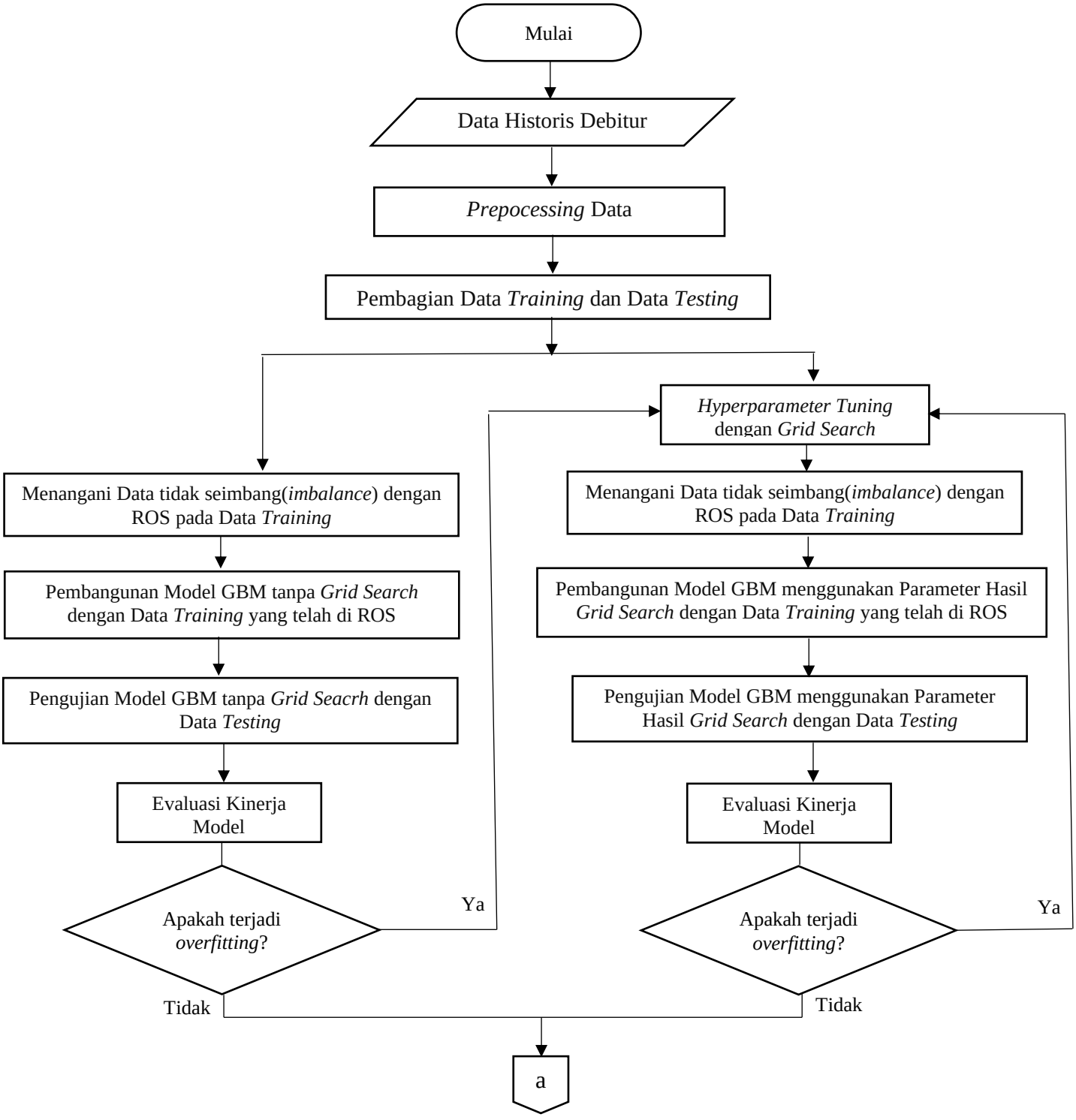
pengetahuan yang terkandung dalam data latihan, semakin tinggi pula akurasi yang dapat dicapai oleh model pohon keputusan. Hasil penelitian menunjukkan bahwa algoritma *Gradient Boosting* memiliki tingkat akurasi yang lebih tinggi daripada analisis yang dilakukan oleh analis kredit. Dapat dibuktikan dengan evaluasi penelitian yang menunjukkan bahwa algoritma *Gradient Boosting* mampu mengidentifikasi debitur bermasalah dan tidak bermasalah dengan tingkat akurasi mencapai 86,63%.

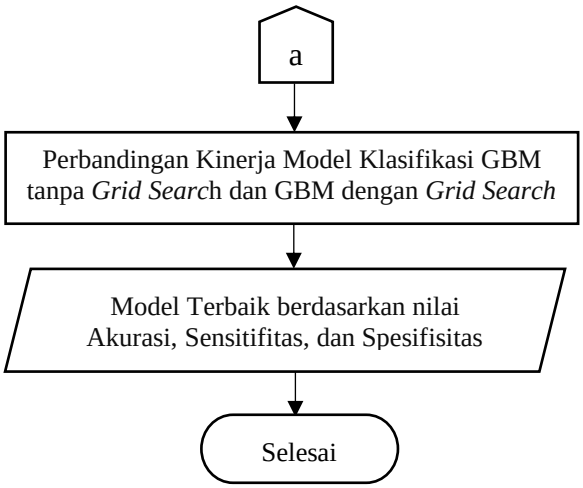
Fitri (2023) membahas tentang prediksi harga rumah dengan analisis perbandingan metode Regresi Linier, *Random Forest R*egression dan *Gradient Boosted Trees Regression*. Dalam penelitian ini, diperoleh hasil pengujian yang menunjukkan bahwa pada *Random Forest Regression*, nilai RMSE adalah 0,440, sementara pada model *Linear Regression*, nilai RMSE adalah 0,515, dan pada *Gradient Boosted Trees Regression*, nilai RMSE adalah 0,508. Data yang diuji terdiri dari 2011 *record*, dengan pembagian 80% untuk data pelatihan dan 20% untuk data pengujian. Data tersebut mencakup enam atribut, yaitu harga rumah, luas tanah , luas bangunan, jumlah kamar mandi, jumlah kamar tidur, dan jumlah garasi. Hasil prediksi menggunakan metode *Random Forest Regression* menunjukkan akurasi tertinggi, mencapai 81,5%, jika dibandingkan dengan model *Linear Regression* dan *Gradient Boosted Trees Regression.*

1.6 Metodologi Penelitian

Penelitian ini menggunakan data sekunder yaitu data historis nasabah yang diperoleh dari dataset pada *website* www.kaggle.com, yang mencakup berbagai variabel seperti umur, jumlah tanggungan, pendapatan bulanan dan faktor-faktor relevan lainnya. Setelah data diperoleh terlebih dahulu dilakukan *preprocessing* data pembersihan data (data *cleaning*) untuk mengatasi nilai-nilai yang hilang atau tidak valid. Selanjutnya, melakukan transformasi data jika diperlukan, seperti normalisasi atau pengkodean variabel kategorik. Langkah berikutnya, membagi informasi menjadi dua bagian, yakni dataset pelatihan (*training* data) dan dataset pengujian (*testing* data). Kemudian menangani *imbalance class* dengan ROS, membuat model GBM awal dengan menggunakan parameter default. Setelah itu digunakan *grid search* untuk mencari kombinasi parameter terbaik yang meningkatkan performa model GBM. Data pengujian dipakai untuk mengevaluasi performa model GBM yang telah dioptimalkan

dan dilakukan pengukuran kinerja model menggunakan metrik yang relevan. Selanjutnya, menganalisis hasil evaluasi untuk memahami sejauh mana model GBM yang dioptimalkan mampu meningkatkan klasifikasi kelayakan pemberian kredit. Kemudian, menarik kesimpulan tentang model terbaik berdasarkan akurasi dari teknik optimasi model GBM dengan *Grid Search* dalam menentukan kelayakan pemberian kredit. Langkah-langkah penelitian ini disajikan dalam *flowchart* pada Gambar 1.1.





Gambar 1.1 *Flowchart* penelitian