# KLASIFIKASI KELULUSAN MAHASISWA FMIPA UNTAN DENGAN *LIGHT GRADIENT BOOSTING MACHINE*

**UTIN AZWA SAYHANI SADIKIN**
**H1091211036**

**SKRIPSI**

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

# KLASIFIKASI KELULUSAN MAHASISWA FMIPA UNTAN DENGAN *LIGHT GRADIENT BOOSTING MACHINE*

**UTIN AZWA SAYHANI SADIKIN**
**H1091211036**

**SKRIPSI**

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

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

## LEMBAR PENGESAHAN

| | |
|---|---|
| Judul Tugas Akhir | : Klasifikasi Kelulusan Mahasiswa FMIPA UNTAN dengan *Light Gradient Boosting Machine* |
| Nama Mahasiswa | : Utin Azwa Sayhani Sadikin |
| NIM | : H1091211036 |
| Jurusan/Program Studi | : Matematika/Statistika |
| Tanggal Lulus | : 17 Juni 2025 |
| SK Pembimbing | : No.2613/UN22.8/TD.06/2024/Tanggal 2 September 2024 |
| SK Penguji | : No.1732/UN22.8/TD.06/2025/Tanggal 10 Juni 2025 |

Dosen Pembimbing

| Pembimbing I | Pembimbing II |
|---|---|
| Shantika Martha, M.Si. | Hendra Perdana, M.Sc. |
| NIP. 198403082008122003 | NIP. 198810102019031020 |

Dosen Penguji

| Ketua Penguji | Anggota Penguji |
|---|---|
| Nurfitri Imro'ah, M.Si | Dr. Yundari, M.Sc |
| NIP. 198907182019032021 | NIP. 198310202008012012 |
| Pimpinan Sidang (merangkap anggota penguji) | Sekretaris Sidang (merangkap anggota penguji) |
| Shantika Martha, M.Si. | Hendra Perdana, M.Sc. |
| NIP. 198403082008122003 | NIP. 198810102019031020 |

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

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

## PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama : Utin Azwa Sayhani Sadikin

NIM : H1091211036

Program Studi/ Jurusan : Statistik/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,

Utin Azwa Sayhani S.

H1091211036

**Klasifikasi Kelulusan Mahasiswa FMIPA UNTAN dengan *Light Gradient Boosting Machine***

## Abstrak

Kelulusan mahasiswa tepat waktu merupakan salah satu indikator keberhasilan dalam pendidikan tinggi. Pencapaian ini dipengaruhi oleh berbagai faktor, seperti performa akademik, latar belakang pendidikan, serta aspek sosial dan ekonomi. Oleh karena itu, diperlukan metode klasifikasi yang efektif untuk membantu pihak program studi maupun fakultas dalam menganalisis faktor-faktor yang berkontribusi terhadap kelulusan mahasiswa dan mengambil langkah strategis yang tepat. Salah satu metode yang dapat digunakan adalah, *Light Gradient Boosting Machine* (LightGBM), yaitu sebuah algoritma berbasis *Gradient Boosting Decision Tree* yang dioptimalkan efisiensi dan kecepatannya. Penelitian ini bertujuan untuk mengklasifikasikan kelulusan mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Tanjungpura (UNTAN) menggunakan metode *Light Gradient Boosting Machine* (LightGBM) dan mengevaluasi hasil klasifikasi tersebut dengan nilai akurasi. Dataset yang digunakan mencakup berbagai variabel seperti faktor akademik yaitu, Indeks Prestasi Semester, dan faktor ekonomi keluarga. Proses pembangunan model dimulai dengan tahap pra-pemrosesan data, termasuk pembersihan data, transformasi variabel. Model LightGBM kemudian dibangun menggunakan hyperparameter yang dioptimalkan untuk memperoleh performa terbaik. Setelah pembuatan model selesai, model disimpan dan dimuat kembali untuk melakukan klasifikasi pada data. Hasil penelitian tersebut didapatkan kesimpulan bahwa mahasiswa yang mendapatkan nilai IP $< 3$ pada 4 semester awal berpotensi tidak lulus tepat waktu. Akurasi yang dihasilkan menggunakan model yang telah dibangun menggunakan metode LightGBM mendapatkan hasil sebesar 75,3%, hasil akurasi tersebut menunjukkan sebesar 75,3% dari data berhasil di klasifikasi dengan benar.

Kata kunci: LightGBM, *Machine Learning*, GBDT, *Boosting*.

***Classification Of Student Graduation Status at FMIPA UNTAN Using Light Gradient Boosting Machine***

## Abstract

*Timely graduation is one of the key indicators of success in higher education. However, this achievement is influenced by various factors, such as academic performance, educational background, and socio-economic aspects. Therefore, an effective classification method is needed to assist study programs and faculties in analyzing the factors contributing to student graduation and making appropriate strategic decisions. One method that can be used is Light Gradient Boosting Machine (LightGBM), an algorithm based on the Gradient Boosting Decision Tree that is optimized for efficiency and speed. Thus, this study aims to develop a classification model using LightGBM to predict student graduation at the Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Tanjungpura (Untan), and evaluate its predictive performance based on accuracy. The dataset used includes various variables such as academic factors, namely Semester Grade Point Average (GPA), and family economic factors. The model development process begins with data preprocessing, including data cleaning, variable transformation, and splitting the dataset into training and testing subsets. The LightGBM model is then built using optimized parameters to achieve the best performance. After the model is developed, it is stored and reloaded for classification on new data. The results indicate that the LightGBM model achieves a good classification accuracy of 75.3%. By utilizing important features identified by the model, the faculty can gain deeper insights into students who are at risk of delayed graduation, allowing for more effective intervention strategies.*

***Keywords:*** *LightGBM, Machine Learning, GBDT, Boosting.*

# PRAKATA

Puji dan syukur atas kehadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya, sehingga penulis diberikan kesempatan serta kekuatan untuk menyelesaikan skripsi yang berjudul **"Klasifikasi Kelulusan Mahasiswa FMIPA UNTAN Dengan *Light Gradient Boosting Machine*"**.

Tujuan penulisan skripsi ini adalah untuk memenuhi syarat dalam memperoleh gelar Sarjana Statistika pada Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura. Dalam penyusunan skripsi ini, penulis menyampaikan ucapat terima kasih kepada :

1. Allah SWT yang telah memberikan kesempatan, kemudahan dan petunjuk kepada penulis dalam menyelesaikan skripsi ini.
2. Keluarga tercinta, terutama orang tua penulis yaitu Mama Eny, Bapak Hasanus, Ami Filda serta Adik-Adik penulis Faris, Gamma, Jasmine, dan Uwais yang selama ini telah memberikan kepercayaan, pengorbanan, mendoakan dan mendukung penulis sehingga dapat menyelesaikan skripsi dengan maksimal.
3. Ibu Naomi Nessyana Debataraja, S.Si., M.Si selaku Dosen Pembimbing Akademik yang telah membimbing, meluangkan waktu, memberi masukan dan semangat selama perkuliahan berjalan
4. Ibu Shantika Martha, M.Si selaku Dosen Pembimbing Pertama yang telah membimbing, meluangkan waktu, memberi masukan dan semangat selama penulisan skripsi ini
5. Bapak Hendra Perdana, M.Sc selaku Dosen Pembimbing Kedua yang telah membimbing, meluangkan waktu, memberi masukin dan semangat selama penulisan skripsi ini.
6. Ibu Nurfitri Imro'ah, M.Si selaku Dosen Penguji Pertama yang telah memberikan masukan dan semangat selama penulisan skripsi ini.

7. Ibu Dr. Yundari, M.Sc selaku Dosen Penguji Kedua yang telah memberikan masukan dan semangat selama penulisan skripsi ini.
8. Seluruh Bapak/Ibu Dosen Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura yang telah memberikan ilmu dan pengalaman selama perkuliahan
9. Teruntuk Viren Marcellya Clarenda Siboro, Maria Meilinda Indriani dan Vriska Dwi Riovani yang membersamai penulis dari awal perkuliahan sampai saat ini. Terima kasih sudah selalu memberikan motivasi, semangat dan masukan dalam keadaan apapun.
10. Teruntuk Daffa Abhinaya Putra yang selalu menemani dan mendukung penulis dalam menyelesaikan skripsi ini. Terima kasih sudah menjadi bagian penting dalam perjalanan penulis hingga saat ini.
11. Teruntuk Riva, Sipa dan Kelly yang selalu memberikan dukungan kepada penulis dan menemani penulis dalam menyelesaikan skripsi ini. Terima kasih sudah menjadi sahabat penulis dalam berproses dan bertumbuh dari SMA hingga sekarang.
12. Teruntuk Immy, Aista, Diana dan Fisca yang selalu menemani dan membersamai di segala kondisi. Terima kasih sudah menjadi sahabat dalam berproses dan bertumbuh di kota ini serta mendukung dan membantu penulis.
13. Teman-teman Statistika Universits Tanjungpura Angkatan 2021 yang telah membersamai selama masa perkuliahan serta memberikan semangat dalam penulisan ini
14. Semua pihak yang telah memberikan dukungan dan bantuannya baik dari segi materil maupun non material yang tidak bisa disebutkan satu persatu.

Pontianak,  Juni 2025

Penulis

# DAFTAR ISI

**LEMBAR PENGESAHAN** ..... **ii**
**PERNYATAAN INTEGRITAS AKADEMIK** ..... **iii**
**Abstrak** ..... **iv**
**Abstract** ..... **v**
**PRAKATA** ..... **vi**
**DAFTAR ISI** ..... **viii**
**DAFTAR TABEL** ..... **x**
**DAFTAR GAMBAR** ..... **xi**
**DAFTAR SIMBOL** ..... **xii**
**BAB 1 PENDAHULUAN** ..... **1**
1.1 Latar Belakang ..... 1
1.2 Rumusan Masalah ..... 3
1.3 Tujuan Penelitian ..... 4
1.4 Batasan Masalah ..... 4
1.5 Tinjauan Pustaka ..... 4
1.6 Metodologi Penelitian ..... 6
**BAB II LANDASAN TEORI** ..... **10**
2.1 Kelulusan Mahasiswa ..... 10
2.2 *Machine Learning* ..... 11
2.3 Klasifikasi ..... 12
2.4 *Gradient Boosting Decision Tree* (GBDT) ..... 12
2.5 *Binary Cross Entropy Loss Function* ..... 13
**BAB III *LIGHT GRADIENT BOOSTING MACHINE*** ..... **15**
3.1. *Classification and Regression Trees* (CART) ..... 15
3.2. *Light Gradient Boosting Machine* ..... 18

3.3. Evaluasi Model Klasifikasi ..... 24
**BAB IV HASIL DAN PEMBAHASAN ..... 26**
4.1 Data Penelitian ..... 26
4.2 Statistik Deskriptif ..... 28
4.3 Pembentukan Pohon Klasifikasi CART ..... 30
4.4 Membangun Model Light Gradient Boosting Machine ..... 36
**BAB V KESIMPULAN DAN SARAN ..... 42**
5.1 Kesimpulan ..... 42
5.2 Saran ..... 43
**DAFTAR PUSTAKA ..... 44**

# DAFTAR TABEL

**Tabel 3.1** *Cost Matrix* ..... **21**
**Tabel 3.2** *Confussion Matrix* ..... **24**
**Tabel 4.1** Klasifikasi Variabel Indenden ..... **26**
**Tabel 4.2** Statistik Deskriptif Data ..... **29**
**Tabel 4.3** Data Calon Simpul Kiri dan Simpul Kanan ..... **31**
**Tabel 4.4** Perhitungan Probabilitas Simpul ..... **32**
**Tabel 4.5** Hasil Perhitungan Informasi Gain ..... **33**
**Tabel 4.6** Hasil Perhitungan Indeks Gini ..... **34**
**Tabel 4.7** Hasil Perhitungan *Goodness Of Split* ..... **34**
**Tabel 4.8** *Hyperparameter* Model LightGBM ..... **37**
**Tabel 4.9** Tampilan Data Awal ..... **38**
**Tabel 4.10** Hasil Perhitungan *Gradien* dan *Hessian* ..... **39**
**Tabel 4.11** Hasil Perhitungan *Gradien* dan *Hessian* Total ..... **39**
**Tabel 4.12** Hasil Perbandingan Data Aktual dan Prediksi pada *Confussion Matrix* .. **41**

# DAFTAR GAMBAR

**Gambar 1.1** *Flowchart* LightGBM ..... **9**
**Gambar 3.1** *Leaf-Wise Tree Growth in LightGBM* ..... **19**
**Gambar 3.3** *Level-Wise Tree Growth in Other Boosting Algorithm* ..... **19**
**Gambar 4.1** Struktur Pohon Keputusan ..... **36**

# DAFTAR SIMBOL

| | |
|---|---|
| $\hat{F}$ | : Fungsi model prediksi dalam LightGBM |
| $\Psi(y, F(x))$ | : Fungsi *Loss* |
| $y$ | : Label target |
| $F(x)$ | : Fungsi dari data sampel |
| $\Gamma_t$ | : Fungsi Objektif pada iterasi ke-$t$ |
| $F_{t-1}(x_i)$ | : Nilai prediksi model di iterasi sebelumnya pada data ke-$i$ |
| $f_t(x_i)$ | : *leaf-wise tree* baru |
| $\Omega(f)$ | : Fungsi regulasi |
| $\omega$ | : Nilai bobot *leaf-wise tree* |
| $\gamma$ dan $\lambda$ | : Koefisien dengan nilai $\gamma = 0{,}01$ dan $\lambda = 1$ |
| $g_i$ | : Turunan orde pertama dari fungsi *loss* terhadap $F_{t-1}(x_i)$ |
| $h_i$ | : Turunan orde kedua dari fungsi *loss* terhadap $F_{t-1}(x_i)$ |
| $F_t(x)$ | : Nilai prediksi baru pada iterasi-$t$ |
| $\eta$ | : *Learning rate* |
| $C_F$ | : Kelas *fault* (kesalahan) |
| $C_N$ | : Kelas normal |
| $R(C_i\|x)$ | : Risiko memprediksi kelas $C_i$ untuk data $x$ |
| $P(C_j\|x)$ | : Probabilitas terjadinya kelas $C_j$ untuk data x |
| $P(c = F\|x_i)$ | : Probabilitas posterior sampel $x_i$ yang termasuk dalam *fault class* |
| $P(c = N\|x_i)$ | : Probabilitas sampel yang termasuk dalam kelas normal |

# BAB 1
# PENDAHULUAN

## 1.1 Latar Belakang

Peguruan tinggi merupakan lembaga yang menyediakan pendidikan akademi bagi mahasiswa. Tingkat keberhasilan mahasiswa dalam menyelesaikan studi menjadi indikator kualitas sebuah perguruan tinggi. Mahasiswa juga menjadi acuan dalam mengevaluasi dan meningkatkan rencana pembelajaran di perguruan tinggi. Menurut Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Nomor 44 Tahun 2015, batas waktu studi maksimal untuk program sarjana adalah tujuh tahun, dengan persyaratan IPK (Indeks Prestasi Kumulatif) minimal 2,0 dan jumlah SKS (Satuan Kredit Semester) yang harus diselesaikan minimal 144 (Crismayella et al., 2023). Masalah tingkat kelulusan menjadi tantangan bagi banyak perguruan tinggi.

Ketidakmampuan dalam memprediksi tingkat kelulusan membuat pengelolaan upaya pencegahan penurunan kelulusan menjadi sulit, yang pada akhirnya dapat berdampak pada akreditas perguruan tinggi tersebut (Yuda et al., 2022). Oleh karena itu, prediksi tingkat kelulusan menjadi hal yang penting dilakukan. Ada banyak faktor yang berkontribusi pada kualitas sebuah perguruan tinggi, salah satunya adalah keberhasilan mahasiswa menyelesaikan studi tepat waktu. Seorang mahasiswa dikatakan lulus tepat waktu jika berhasil menyelesaikan studinya dalam jangka waktu maksimal empat tahun. Jika masa studinya melebihi empat tahun, maka mahasiswa tersebut dianggap lulus tepat waktu. Jika masa studinya melebihi empat tahun, maka mahasiswa tersebut dianggap lulus tidak tepat waktu (Aipassa et al., 2024).

Semakin banyak mahasiswa yang lulus sesuai jadwal, semakin positif dampaknya terhadap akreditasi program studi di universitas tersebut. Untuk meningkatkan tingkat kelulusan, diperlukan analisis terhadap berbagai faktor yang memengaruhinya, seperti rendahnya kemampuan akademis, kualitas program

perkuliahan, indeks prestasi, dan faktor lainnya, sehingga kebijakan yang dihasilkan dapat membantu semua mahasiswa lulus tepat waktu (Zeniarja et al., 2022).

Universitas Tanjungpura (UNTAN) merupakan perguruan tinggi negeri yang terletak di Kota Pontianak, Kalimantan Barat. Salah satu Fakultas yang terdapat di UNTAN adalah Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA). Berdasarkan buku panduan akademik FMIPA UNTAN tahun 2022/2023, keberhasilan studi mahasiswa dievaluasi melalui empat tahap, yaitu: pada akhir setiap semester, akhir empat semester pertama, akhir semester delapan serta akhir program pendidikan. Oleh karena itu, hasil evaluasi pada akhir empat semester pertama dapat digunakan untuk memprediksi klasifikasi mahasiswa FMIPA UNTAN yang berpotensi menyelesaikan studi tepat waktu maupun yang tidak tepat waktu (Aipassa et al., 2024). Namun, tidak semua mahasiswa dapat menyelesaikan studi tepat waktu sesuai dengan batas yang ditentukan. Sebagian mahasiswa bahkan melampaui batas maksimal masa studi S1 atau menghadapi resiko *drop out.* Saat ini, keterlambatan kelulusan masih menjadi tantangan bagi Program Studi, yang disebabkan oleh keterbatasan dalam memprediksi masa studi mahasiswa sejak awal. Dengan adanya prediksi klasifikasi kelulusan mahasiswa, program studi dapat merancang strategi perencanaan, mengawasi perkembangan studi, serta memberikan bimbingan yang lebih intensif kepada mahasiswa yang terindikasi akan terlambat lulus atau berpotensi *drop out* (Orpa et al., 2019).

*Machine learning* merupakan penerapan algoritma matematika dan komputasi yang mempelajari pola dari data untuk menghasilkan prediksi di masa mendatang. Masalah klasifikasi termasuk di dalam metode *machine learning* berbasis *supervised learning*. *Supervised learning* adalah jenis metode dalam *machine learning* di mana model dilatih menggunakan data yang sudah diberi label, yaitu data yang memiliki pasangan input dan output yang diketahui (Roihan et al., 2020). Klasifikasi melibatkan pengelompokan data yang memiliki kelas atau target tertentu untuk diklasifikasikan. Dengan demikian, algoritma dapat diterapkan untuk menyelesaikan masalah klasifikasi yang termasuk dalam kategori *supervised learning* (Apridiansyah et al., 2021).

Sehingga, dengan memanfaatkan *machine learning* pada data pendidikan, sebuah perguruan tinggi dapat mengakses informasi yang bermanfaat untuk melakukan perbaikan dan meningkatkan kualitas program studi dan fakultas (Novianto et al., 2023). Data yang dapat dilihat pada lulusan FMIPA hanya data lulusan setiap periode wisuda. Oleh karena itu, perlu melakukan prediksi klasifikasi kelulusan mahasiswa FMIPA UNTAN menggunakan data dari nilai akhir empat semester pertama, agar dapat di evaluasi.

Metode *Light Gradient Boosting* Machine (LightGBM), yang dikembangkan oleh *Microsoft*, adalah algoritma yang mengoptimalkan teknik *Gradient Boosting*. LightGBM adalah salah satu metode ensamble yang menggabungkan prediksi dari beberapa *decision tree* dengan menambahkan setiap tree secara bertahap. LightGBM termasuk dalam kategori *Gradient Boosting Decision Tree* (GBDT) terbaru yang dirancang untuk mengelola big data serta sejumlah fitur yang besar. Algoritma ini telah banyak diterapkan dalam berbagai aplikasi klasifikasi dan regresi, menunjukkan hasil yang sangat baik, dan membuktikan keefektifannya sebagai algoritma pengklasifikasian (Rizky et al., 2022). Sehingga, dengan menerapkan *machine learning* berbasis *supervised learning* menggunakan metode LightGBM diharapkan dapat memprediksi klasifikasi mahasiswa yang lulus tepat waktu maupun yang tidak tepat waktu di FMIPA UNTAN dengan mengharapkan hasil yang dapat menjadi informasi bagi Fakultas maupun Program Studi dalam mengevaluasi mahasiswa yang berpotensi tidak lulus tepat waktu.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka permasalahan yang akan dibahas dalam penelitian ini adalah:

1. Bagaimana model yang dibangun menggunakan metode *Light Gradient Boosting Machine* dalam mengklasifikasikan kelulusan mahasiswa FMIPA UNTAN?
2. Bagaimana hasil evaluasi klasifikasi kelulusan mahasiswa FMIPA UNTAN dengan metode *Light Gradient Boosting Machine*?

## 1.3 Tujuan Penelitian

Tujuan dari penelitian ini, adalah:

1. Mengklasifikasikan kelulusan mahasiswa FMIPA UNTAN dengan metode LightGBM.
2. Mengevaluasi hasil klasifikasi kelulusan mahasiswa FMIPA UNTAN dengan nilai akurasi.

## 1.4 Batasan Masalah

Batasan ruang lingkup dalam penelitian ini adalah, sebagai berikut:

1. Data yang digunakan adalah data kelulusan mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Periode 2 Tahun Ajaran 2017/2018 sampai Periode 1 Tahun Ajaran 2023/2024, dengan jumlah data sebanyak 1893.
2. Atribut yang digunakan antara lain Status Lulus Tepat Waktu dan Tidak Tepat Waktu, IP semester 1 sampai dengan semester 4, Asal Provinsi, Jalur Masuk Perguruan Tinggi, Pekerjaan ayah, Pekerjaan Ibu, Pendidikan terakhir Ayah, Pendidikan terakhir Ibu, Range Gaji Ayah, dan Range Gaji Ibu,
3. Pengaplikasian *software* R Studio sebagai alat bantu dalam mengimplementasikan dan mencari nilai akurasi pada model Light GBM.

## 1.5 Tinjauan Pustaka

Suniantara et al., (2020) Melakukan penelitian klasifikasi kelulusan mahasiswa dengan meningkatkan akurasi menggunakan metode *Boosting Neural Network*. Penelitian ini bertujuan untuk mengembangkan model klasifikasi *Boosting Feedforward Neural Network* (FFNN) yang dapat diimplementasikan dan digunakan sebagai bentuk model klasifikasi yang menghasilkan akurasi lebih baik. Hasil akurasi yang didapatkan adalah sebesar 72,93%. Setelah diterapkan *boosting* pada FFNN menghasilkan tingkat akurasi yang paling baik, yaitu 74,44%.

*Kurniadi* & Larasati (2022) melakukan penelitian dengan menggunakan metode *Light Gradient Boosting Machine* (LightGBM) dan membandingkannya dengan metode SVM dan metode *Random Forest*. Penelitian ini bertujuan untuk mengklasifikasi dalam mendeteksi penyakit Stroke. Hasil penelitian menggunakan metode LightGBM didapatkan nilai akurasi mencapai 98%. Hasil akurasi menggunakan metode SVM dan *Random Forest* juga mendapatkan hasil yang sama. Namun, penelitian tersebut menggunakan dua *scenario* yang akan dicoba, pertama akan dibahas mengenai hasil yang didapatkan menggunakan proses fitur seleksi dan *scenario* kedua merupakan hasil yang didapatkan setelah mendapatkan hasil dari fitur seleksi. Pada hasil akurasi dari ketiga metode tersebut didapatkan hasil yang sama dari kedua scenario, namun hasil *precision* dan *recall* yang berbeda dari ketiga model dapat diambil kesimpulan bahwa adanya *imbalanced* data yang tidak ditangani dan akan membuat bias.

Diantika (2023) melakukan penelitian dengan menggunakan metode *Light Gradient Boosting Machine* (LightGBM) dengan menerapkan teknik *Random Oversampling* dalam mengatasi *Imbalance Class*. Penelitian tersebut menggunakan data *website phishing detector* yang berisi kumpulan URL situs web untuk 11000+ situs web. Hasil penelitian tersebut didapatkan bahwa model Light GBM memiliki nilai akurasi tertinggi yaitu 99,7%. Dari hasil penelitian menunjukkan bahwa model yang digunakan memiliki performa yang lebih unggul dibandingkan dengan beberapa model lain yang telah diuji dalam penelitian tersebut.

Handayani & Erni (2023) melakukan penelitian menggunakan metode *Light Gradient Boosting Machine* (LightGBM) untuk memprediksi Rasio Klik Tayang. Dalam memprediksi rasio klik tayang penelitian ini menggunakan *dataset social network*. Pada penelitian ini dilakukan pengujian resampling untuk menganalisis dampak penggunaan dataset *social network* yang memiliki ketidakseimbangan. Kemudian, melakukan evaluasi model untuk mengetahui kinerja dari hasil uji model LightGBM dengan *resampling.* Sehingga, memperoleh hasil akurasi 91,25%, *recall* 93,10%, *precision* 84,38%, *Fi-score* 88,52% dan AUC 0,92. Hasil yang diperoleh ini

tanpa *resampling* dan menggunakan *resampling* ROS. Penggunaan *resampling* dalam pengujian *dataset social network* dengan menggunakan LightGBM tidak menghasilkan performa yang berbeda.

Latifah & Erda (2024) melakukan penelitian menggunakan metode *Light Gradient Boosting Machine* (LightGBM) dalam klasifikasi Emisi Gas Rumah Kaca. Data yang digunakan terdiri dari 39 variabel independen yang mempengaruhi perubahan iklim di dunia dan variabel dependennya adalah total emisi gas rumah kaca. Pemodelan dilakukan menggunakan LightGBM dengan klasifikasi biner, kemudian melakukan prediksi pada data pengujian dan menghitung nilai akurasi, *sensitivity*, *specificity*, dan MAE. Hasil yang diperoleh dari proses optimasi parameter pada klasifikasi adalah akurasi sebesar 98,72%, sensitivitas 99,42%, spesifisitas 97,49% dan MAE bernilai 0,0128. Berdasarkan empat nilai evaluasi kinerja yang diperoleh, dapat disimpulkan bahwa metode LightGBM sangat optimal untuk digunakan dalam klasifikasi emisi gas rumah kaca.

## 1.6 Metodologi Penelitian

Penelitian ini menggunakan data primer dan data sekunder. Data yang digunakan adalah Kelulusan Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Periode 2 Tahun Ajaran 2017/2018 hingga Periode 1 Tahun Ajaran 2023/2024. Proses analisis dilakukan dengan diawali pengumpulan data yang terdiri atas variabel dependen yaitu Status dan 13 variabel independen yaitu IP Semester 1 sampai 4, Jalur Masuk Perguruan Tinggi, Asal Provinsi, Jurusan SLTA, Pekerjaan Ayah, Pekerjaan Ibu, Pendidikan Ayah, Pendidikan Ibu, Range Gaji Ayah, dan Range Gaji Ibu. Langkah-langkah analisis pada penelitian ini adalah, sebagai berikut:

Proses analisis diawali dengan penginputan data yang akan digunakan dalam penelitian, kemudian melakukan *pre-processing*. *Pre-processing* dilakukan dengan menghapus data yang hilang (*missing values*). Setelah data siap, dilakukan analisis statistik deskriptif untuk memahami karakteristik data dan untuk melihat pola dalam

data. Selanjutnya, melakukan normalisasi data yang digunakan untuk menyelaraskan skala variabel sehingga model dapat bekerja secara optimal.

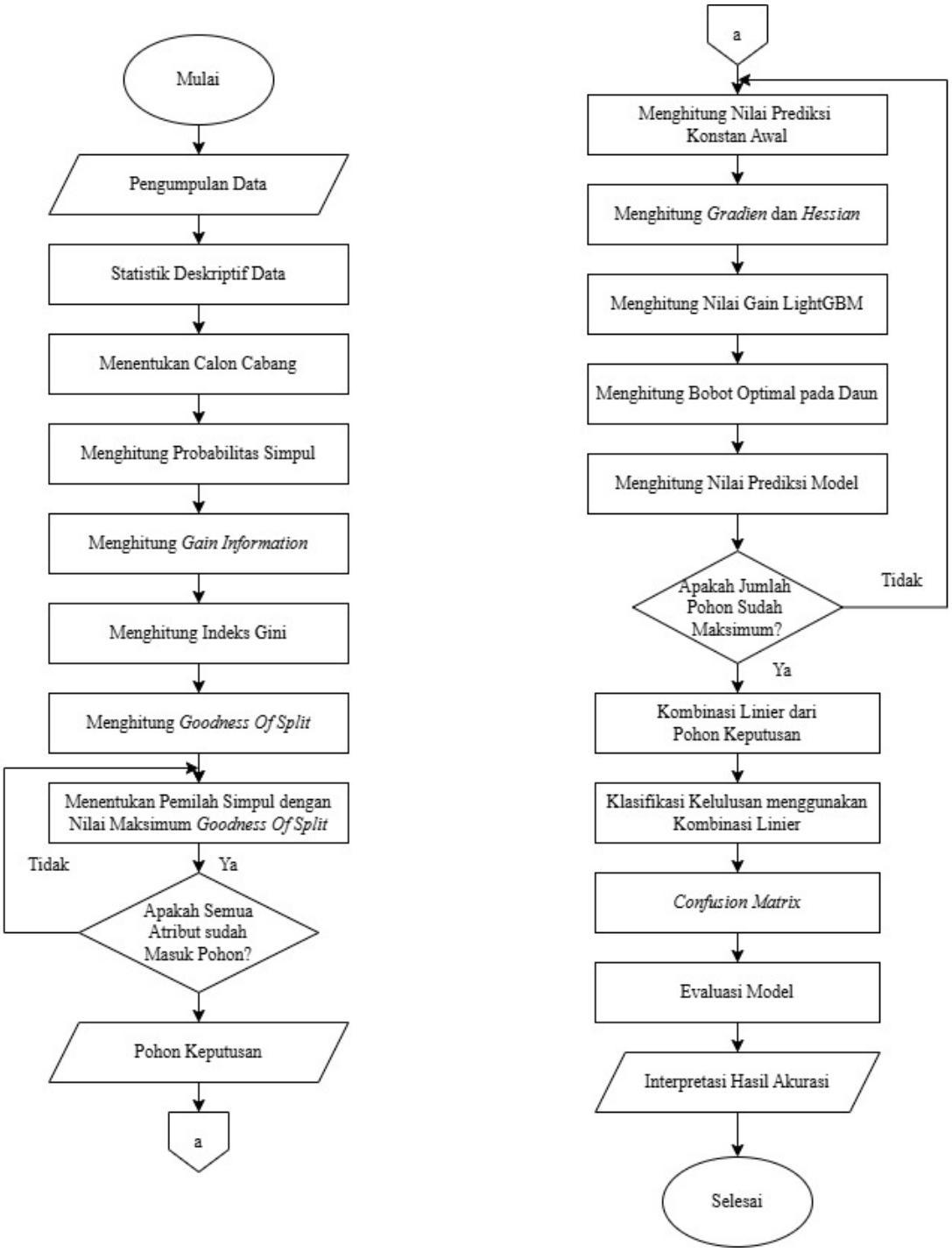
Kemudian langkah selanjutnya adalah menentukan calon cabang (*split candidates*) untuk setiap fitur. Setiap fitur dalam dataset dianalisis sebagai kandidat potensial untuk membagi simpul pada pohon keputusan. Kemudian, menghitung probabilitas setiap kelas dalam simpul dihitung berdasarkan distribusi label target. Nilai probabilitas ini digunakan dalam perhitungan metrik pemilihan atribut terbaik untuk pemisahan simpul. Langkah sleanjutnya adalah menghitung Gain informasi, nilai gain informasi dihitung untuk setiap calon pemisah (*split*) dalam pohon keputusan. Gain informasi menggambarkan seberapa banyak ketidakpastian dalam data yang dapat dikurangi dengan menggunakan atribut tertentu sebagai pemilah. Selain gain informasi, indeks Gini dihitung untuk menilai seberapa baik suatu atribut membagi data dalam simpul menjadi kelompok yang lebih homogen. Semakin rendah nilai indeks Gini, semakin baik atribut tersebut dalam memisahkan kelas.

Selanjutnya adalah menghitung *Goodness of Split*, nilai *goodness of split* merupakan ukuran yang menggabungkan Gain Informasi dan Indeks Gini untuk menentukan efektivitas suatu atribut dalam membagi simpul menjadi dua kelompok yang lebih murni. Kemudian, dari seluruh atribut yang diuji, atribut dengan nilai maksimum *Goodness of Split* dipilih sebagai pemilah terbaik untuk membagi simpul pada tahap tertentu dalam proses pembentukan pohon keputusan. Selanjutnya, atribut terbaik dimasukkan ke dalam struktur pohon keputusan, untuk membentuk percabangan baru. Proses ini dilakukan secara berulang untuk membentuk pohon secara optimal. Setelah beberapa iterasi, pohon keputusan terbentuk, dengan setiap simpul mencerminkan atribut terbaik yang membagi data, dan setiap daun mewakili prediksi berdasarkan distribusi kelas target.

Dalam LightGBM, model pertama kali memulai dengan menghitung nilai prediksi konstan awal, yang biasanya merupakan rata-rata atau log-odds dari fitur target dalam kasus klasifikasi biner. Setelah prediksi awal ditentukan, residual dihitung sebagai selisih antara nilai aktual dan nilai prediksi. Residual ini menjadi dasar bagi

model untuk melakukan perbaikan pada iterasi berikutnya. Kemudian menghitung nilai Gain untuk setiap fitur yang digunakan sebagai kandidat *split* dalam pohon berikutnya. Proses ini memungkinkan model untuk menentukan pemisahan terbaik berdasarkan dampaknya terhadap penurunan kesalahan prediksi. Selanjutnya adalah menghitung obot optimal pada setiap daun dalam pohon dihitung menggunakan pendekatan optimasi berbasis gradien. Bobot ini menentukan kontribusi setiap daun dalam menghasilkan prediksi akhir. Setelah bobot optimal diperoleh, nilai prediksi akhir dihitung dengan menggabungkan kontribusi dari seluruh pohon yang telah dibangun dalam iterasi sebelumnya. Kemudian, proses pembangunan pohon dilanjutkan hingga jumlah maksimum pohon yang telah ditentukan tercapai atau hingga tidak ada perbaikan signifikan dalam performa model. Setelah pohon-pohon terbentuk, model LightGBM yang telah dilatih digunakan untuk melakukan prediksi terhadap data uji.

Langkah selanjutnya, dari model yang sudah dibangun menghasilkan prediksi berdasarkan pola yang dipelajari selama proses pembangunan daun dari pohon. Kemudian melakukan evaluasi model dengan menggunakan *confusion matrix* untuk mengukur performa model dalam mengklasifikasikan data dengan benar, dengan memperhatikan nilai akurasi. Setelah evaluasi dilakukan, hasil tersebut dianalisis untuk memahami faktor-faktor yang berpengaruh dalam klasifikasi. Kemudian mengambil kesimpulan berdasarkan performa model dan interpretasi terhadap fitur yang paling berkontribusi dalam prediksi. Dengan selesainya analisis ini, model LightGBM telah berhasil digunakan untuk klasifikasi data dengan metode yang efisien dan optimal. Adapun *flowchart* penelitian disajikan pada Gambar 1.1 berikut:




**Gambar 1.1** *Flowchart* CART dan LightGBM