



**IMPLEMENTASI *LIGHT GRADIENT BOOSTING MACHINE* PADA
PREDIKSI PENYAKIT LIVER DENGAN MENGGUNAKAN *INDIAN
PATIENT LIVER DATASET***

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

Yudhitya Rahman

NIM 1711016310033

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

MEI 2024



**IMPLEMENTASI *LIGHT GRADIENT BOOSTING MACHINE*
PADA PREDIKSI PENYAKIT LIVER DENGAN MENGGUNAKAN
*INDIAN PATIENT LIVER DATASET***

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

YUDHITYA RAHMAN

NIM 1711016310033

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
MEI 2024**

SKRIPSI

IMPLEMENTASI *LIGHT GRADIENT BOOSTING MACHINE* PADA PREDIKSI PENYAKIT LIVER DENGAN MENGGUNAKAN *INDIAN PATIENT LIVER DATASET*

Oleh:

YUDHITYA RAHMAN

NIM 1711016310033

Telah dipertahankan di depan Dosen Penguji pada tanggal 21 Juni 2024.

Susunan Dosen Penguji:

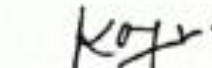
Pembimbing I



Muliadi, S.Kom., M.Cs.

NIP. 197804222010121002

Dosen Penguji I



Dwi Kartini, S.Kom., M.Kom

NIP. 198704212012122003

Pembimbing II



Friska Ahadi, S.Kom., M.Kom

NIP. 198809132023211010

Dosen Penguji II



Mohammad Reza Faisal, S.Si., S.T.M.T PhD

NIP. 197612202008121001



21 Juni 2024
Program Studi Ilmu Komputer

Yudhitya Rahman, S.T., M.Kom

NIP. 197703252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, Mei 2024
Yang Menyatakan,



Yudhitya Rahman
NIM.1711016310033

ABSTRAK

IMPLEMENTASI *LIGHT GRADIENT BOOSTING MACHINE* PADA PREDIKSI PENYAKIT LIVER DENGAN MENGGUNAKAN *INDIAN PATIENT LIVER DATASET*

(Oleh : Yudhitya Rahman; Pembimbing: Muliadi S.Kom, M.Cs dan Friska Abadi S.Kom., M.Kom.; 2023; 46 halaman)

Penyakit liver, terutama sirosis hati dan hepatitis non-alkoholik, merupakan masalah kesehatan global yang signifikan dengan tingkat morbiditas dan mortalitas yang tinggi. Dalam upaya untuk meningkatkan deteksi dini dan manajemen penyakit liver, penelitian ini menggunakan *dataset Indian Liver Patient* yang mencakup fitur klinis yang relevan untuk memprediksi kehadiran atau absennya penyakit liver. Dataset ini merupakan dataset yang tidak seimbang yang mana pasien yang terdiagnosa penyakit liver berjumlah 414 dan pasien yang tidak terdiagnosa penyakit liver berjumlah 165. Kami menerapkan metode *LightGBM*, sebuah algoritma pembelajaran mesin yang efisien dan kuat, untuk membangun model prediksi. Evaluasi kinerja model dilakukan menggunakan *Area Under the Curve (AUC)* dari kurva *Receiver Operating Characteristic (ROC)*, yang memberikan ukuran keseluruhan dari kemampuan model dalam memisahkan kelas positif dan negatif. Hasil eksperimen kami menunjukkan bahwa model *LightGBM* mencapai *AUC* sebesar 0,71, menunjukkan kinerja yang cukup baik dalam memprediksi penyakit liver. Temuan ini menunjukkan potensi *LightGBM* sebagai alat bantu dalam deteksi dini dan manajemen penyakit liver, serta menyoroti pentingnya penggunaan metode pembelajaran mesin dalam pemrosesan data medis yang kompleks.

Kata Kunci : Klasifikasi; *Light Gradient Boosting Machine*; Akurasi; *AUC*;

ABSTRACT

IMPLEMENTATION OF LIGHTGBM IN PREDICTING LIVER DISEASES USING THE INDIAN LIVER PATIENT DATASET

(By : Yudhitya Rahman; Supervisor: Muliadi S.Kom, M.Cs and Friska Abadi S.Kom., M.Kom.; 2022; 46 pages)

Liver diseases, particularly liver cirrhosis and non-alcoholic hepatitis, represent a significant global health issue with high rates of morbidity and mortality. In an effort to improve early detection and management of liver diseases, this research utilized the Indian Liver Patient dataset containing relevant clinical features to predict the presence or absence of liver disease. This dataset is unbalanced, with 414 patients diagnosed with liver disease and 165 patients without liver disease. We applied the LightGBM method, an efficient and powerful machine learning algorithm, to construct a prediction model. Model performance evaluation was conducted using the Area Under the Curve (AUC) from the Receiver Operating Characteristic (ROC) curve, providing an overall measure of the model's ability to separate positive and negative classes. Our experimental results showed that the LightGBM model achieved an AUC of 0.71, indicating good performance in predicting liver disease. These findings highlight the potential of LightGBM as an aid in early detection and management of liver disease, underscoring the importance of using machine learning methods in processing complex medical data.

Kata Kunci : *Classification; Light Gradient Boosting Machine; Accuracy; AUC;*

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “*Implementasi Light Gradient Boosting Machine Pada Prediksi Penyakit Liver dengan Menggunakan Indian Liver Patient Dataset*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiyamah*.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Keluarga, terutama kedua orang tua tercinta untuk mama dan abah yang dari awal proses perkuliahan selalu memberikan bantuan, semangat, doa dan dukungan hingga sampai pada tahap proses penyelesaian skripsi ini.
2. Bapak selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini.
3. Bapak . selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan Staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman yang telah bersedia membantu memecahkan kebingungan-kebingungan dan memberikan saran, dukungan, semangat, serta motivasi, selama proses penyelesaian skripsi.
7. Teman-teman keluarga Ilmu Komputer angkatan 2017 yang memberikan dukungan dan bantuan semasa kuliah dan selama proses penyelesaian skripsi.
8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 27 Mei 2024

Yudhitya Rahman

DAFTAR ISI

ABSTRAK	4
ABSTRACT	5
PRAKATA.....	6
DAFTAR ISI.....	8
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
4.1 Tujuan.....	2
4.2 Manfaat Penelitian	2
BAB II TINJAUAN PUSTAKA.....	3
2.1 Kajian Terdahulu	3
2.2 Landasan Teori.....	8
2.2.1. Penyakit Liver.....	8
2.2.2. Indian Liver Patient Dataset.....	8
2.2.3. Ketidakseimbangan Data	9
2.2.4. Preprocessing	10
2.2.5. Missing Value	11
2.2.6. Label Encoder	12
2.2.7. Klasifikasi.....	14
2.2.3. Decission Tree.....	15
2.2.4. Ensemble Learning	17
2.2.5. Boosting	18
2.2.6. LightGBM.....	20
2.2.9. Confusion Matrix	26
BAB III METODE PENELITIAN.....	29
3.1 Bahan Penelitian	29
3.2 Alat Penelitian.....	29
3.3 Variabel Penelitian.....	29
3.4 Prosedur Penelitan	29
BAB IV HASIL DAN PEMBAHASAN	33
4.1 Hasil	33
4.1.1 Pengumpulan Dataset.....	33
4.1.2 Preprocessing Data.....	35
4.1.2.1 Label Encoder	36
4.1.3.1 Missing Value	38
4.1.3 Klasifikasi <i>Light Gradient Boosting Machine</i>	39
4.2 Pembahasan.....	43
BAB V PENUTUP.....	45
5.1 Kesimpulan.....	45
5.2 Saran.....	45
DAFTAR PUSTAKA	50

DAFTAR TABEL

Tabel 1. Keaslian Penelitian.....	7
Tabel 2. Atribut dan Outcome.....	9
Tabel 3. Parameter LightGBM.....	23
Tabel 4. Stratified 10-Fold Cross Validation.....	25
Tabel 5. Klasifikasi AUC.....	28
Tabel 6. Indian Liver Patient Dataset.....	31
Tabel 7. Fitur untuk prediksi Indian liver patient dataset.....	33
Tabel 8. Contoh Indian liver patient dataset.....	34
Tabel 9. Jumlah Distribusi Kelas.....	34
Tabel 10. Dataset sebelum dilakukan label encoder.....	36
Tabel 11. Proses Label Encoder.....	36
Tabel 12. dataset setelah dilakukan label encoder.....	37
Tabel 13. Distribusi Kelas Sebelum pengisian dengan nilai rata-rata.....	38
Tabel 14. Distribusi Kelas Setelah pengisian dengan nilai rata-rata pada missing Value.....	39
Tabel 15. Parameter Default LightGBM.....	39
Tabel 16. Confusion matrix Model LightGBM.....	40
Tabel 17. Mengubah nilai parameter default.....	41
Tabel 18. Confusion matrix klasifikasi Light gradient boosting machine.....	42

DAFTAR GAMBAR

Gambar 1 Struktur Decision Tree	17
Gambar 2. Ilustrasi Ensemble Learning	17
Gambar 3. Boosting Ensemble Learning	19
Gambar 4. Teknik Pembelahan Node daun pada Model LightGBM	20
Gambar 5. Tahapan Light Gradient Boosting Machine	21
Gambar 6. Contoh AUC Curve	27
Gambar 7. Alur Penelitian	30
Gambar 8. Perbandingan Presentase Pasien Liver dan Bukan Pasien Liver	35
Gambar 9. Alur Preprocessing	35
Gambar 10. Ilustrasi Pembagian Dataset	37
Gambar 11. Fitur yang terdapat missing value	38
Gambar 12. Fitur A/G setelah dilakukan missing value	39
Gambar 13. ROC AUC dengan model Light Gradient Boosting Machine	41

DAFTAR LAMPIRAN

Lampiran 1 Source Code.....	54
Lampiran 2. <i>Source Code Preprocessing Missing Value</i>	55
Lampiran 3. <i>Source Code Preprocessing Label Encoder</i>	56
Lampiran 4. Splitting Data	56
Lampiran 5. AUC.....	57