

## EVALUASI KINERJA MODEL GRADIENT BOOSTED TREES UNTUK PREDIKSI STATUS KOMORBIDITAS PADA PASIEN BATU EMPEDU

Fasya Gilar Amali<sup>1)</sup>, Hasbi Firmansyah<sup>2)</sup>, Wahyu Asriani<sup>3)</sup>

Universitas Pancasakti Tegal

Email: fasyagilar80@gmail.com<sup>1)</sup>, hasbifirmansyah@upstegal.ac.id<sup>2)</sup>, asriani1409@gmail.com<sup>3)</sup>

### ABSTRAK

Penelitian ini bertujuan mengevaluasi kinerja model Gradient Boosted Trees (GBT) dalam memprediksi status komorbiditas (ada/tidak ada) pada pasien batu empedu. Data berasal dari UCI Machine Learning Repository dan mencakup 319 individu dengan 38 fitur klinis. Proses pemodelan meliputi pembersihan data, transformasi fitur, pembagian data dengan rasio 70:300 untuk pelatihan dan pengujian, pelatihan model GBT dengan skema 10-fold cross-validation, serta evaluasi menggunakan empat metrik utama, yaitu akurasi, classification error, weighted mean precision, dan weighted mean recall. Hasil eksperimen menunjukkan bahwa model GBT mencapai akurasi 82,29%, classification error 17,71%, weighted mean precision 89,63%, dan weighted mean recall 72,58%. Analisis lebih lanjut mengungkap precision sempurna (100,00%) pada kelas komorbiditas dan recall sempurna (100,00%) pada kelas tanpa komorbiditas, sementara recall kelas komorbiditas hanya 45,16%, yang mengindikasikan bias model terhadap kelas mayoritas dan keterbatasan dalam mendeteksi pasien berkomorbid. Temuan ini menunjukkan bahwa GBT merupakan pendekatan menjanjikan sebagai alat bantu keputusan untuk mengonfirmasi komorbiditas, namun masih kurang ideal sebagai alat skrining awal. Penelitian lanjutan disarankan menerapkan teknik penanganan ketidakseimbangan kelas, optimasi hyperparameter, penyesuaian threshold, serta perbandingan dengan algoritma ensemble lain guna meningkatkan sensitivitas dan generalisasi model.

**Kata Kunci :** Komorbiditas, Batu Empedu, Gradient Boosted Trees, Klasifikasi, RapidMiner, Weighted Mean Precision

### ABSTRACT

*This study aims to evaluate the performance of a Gradient Boosted Trees (GBT) model for predicting comorbidity status (present/absent) in gallstone patients. The dataset, obtained from the UCI Machine Learning Repository, comprises 319 individuals with 38 clinical features. The modelling pipeline includes data cleaning, feature transformation, a 70:30 train-test split, GBT training with 10-fold cross-validation, and evaluation using four metrics: accuracy, classification error, weighted mean precision, and weighted mean recall. Experimental results show that the GBT model achieves an accuracy of 82.29%, a classification error of 17.71%, a weighted mean precision of 89.63%, and a weighted mean recall of 72.58%. Further analysis reveals perfect precision (100.00%) for the comorbidity class and perfect recall (100.00%) for the non-comorbidity class, while recall for the comorbidity class is only 45.16%. This pattern indicates bias toward the majority class and limited sensitivity for detecting comorbid patients. These findings suggest that GBT is promising as a decision-support tool to confirm the presence of comorbidity, but less suitable as a primary screening tool. Future research should apply class-imbalance handling, hyperparameter optimisation, threshold adjustment, and comparison with alternative ensemble algorithms to improve sensitivity and generalisability.*

**Keyword :** Comorbidity, Gallstones, Gradient Boosted Trees, Classification, RapidMiner, Weighted Mean Precision

## I. PENDAHULUAN

### 1.1 Latar Belakang

Penyakit batu empedu, atau kolelitiasis, merupakan salah satu gangguan sistem hepatobilier yang paling umum di seluruh dunia. Di negara berkembang, prevalensi kondisi ini diperkirakan mencapai 10 hingga 15 persen dari populasi dewasa. Batu empedu terbentuk akibat ketidakseimbangan komponen empedu, seperti kolesterol, garam empedu, dan bilirubin, serta terganggunya motilitas kantung empedu. Proses pembentukan batu bersifat multifaktorial dan melibatkan interaksi faktor genetik, metabolik, hormonal, serta gaya hidup. Dalam banyak kasus, batu empedu tidak menimbulkan gejala hingga terjadi komplikasi, sehingga diagnosis dan penanganan kerap terlambat.

Selain gejala utama yang terkait langsung dengan batu empedu, pasien sering kali juga memiliki komorbiditas, yaitu kehadiran satu atau lebih kondisi medis lain yang terjadi bersamaan. Komorbiditas seperti diabetes melitus, hiperlipidemia, penyakit jantung koroner, hipertensi, dan hipotiroidisme sering ditemukan pada penderita batu empedu. Kehadiran komorbiditas ini meningkatkan kompleksitas manajemen klinis dan memperburuk luaran jangka panjang, termasuk meningkatnya risiko komplikasi pascaoperasi, perubahan respons terhadap terapi, perpanjangan durasi rawat inap, serta penurunan kualitas hidup pasien.

Pendekatan klinis tradisional dalam menilai komorbiditas umumnya masih bergantung pada intuisi dan pengalaman praktisi medis. Kompleksitas data klinis modern yang mencakup parameter demografis, bioimpedansi, dan laboratorium menjadikan penilaian subjektif berpotensi tidak konsisten dan sulit direplikasi. Oleh karena itu, dibutuhkan pendekatan yang mampu mengelola volume dan variabilitas data secara efisien sekaligus memberikan penilaian yang lebih objektif dan terstandar. Dalam konteks ini, pemanfaatan teknologi machine learning menjadi sangat relevan untuk mendukung pengambilan keputusan klinis.

Salah satu algoritma yang menonjol di domain ini adalah Gradient Boosted Trees (GBT), yaitu metode ensemble yang membangun sekumpulan pohon keputusan secara bertahap dengan meminimalkan kesalahan model sebelumnya. GBT dikenal mampu menangani data kategorikal maupun numerik, menangkap hubungan non-linear antar variabel, dan umumnya memberikan kinerja prediktif yang tinggi pada data tabular klinis. Berbagai studi telah menunjukkan keberhasilan GBT dalam tugas-tugas prediktif di bidang medis, seperti deteksi kanker, prediksi risiko diabetes, estimasi risiko rawat inap, dan klasifikasi penyakit kronis. Selain itu, analisis feature importance pada GBT dapat memberikan wawasan mengenai fitur klinis yang paling berpengaruh, yang penting bagi transparansi dan penerimaan di lingkungan klinis.

Namun demikian, sebagian besar penelitian sebelumnya lebih berfokus pada prediksi keberadaan batu empedu atau luaran klinis lain yang spesifik, bukan pada identifikasi status komorbiditas yang menyertai pasien batu empedu. Padahal, pemahaman terhadap pola komorbiditas pada populasi ini sangat penting untuk merancang intervensi dini dan strategi penanganan jangka panjang yang lebih efektif. Penelitian ini hadir untuk menjembatani kesenjangan tersebut dengan mengembangkan dan mengevaluasi kinerja model Gradient Boosted Trees untuk memprediksi status komorbiditas (ada/tidak ada) pada pasien batu empedu. Dataset klinis yang digunakan dalam penelitian ini merupakan dataset publik dari UCI Machine Learning Repository, yang mencakup 319 individu dengan 38 fitur demografis, bioimpedansi, dan laboratorium.

Dengan memanfaatkan dataset klinis komprehensif dari UCI tersebut, penelitian ini bertujuan membangun model klasifikasi berbasis GBT serta mengevaluasi kinerjanya menggunakan metrik akurasi, *classification error*, *weighted mean precision*, dan *weighted mean recall*. Selain itu, penelitian ini juga berupaya mengidentifikasi fitur-fitur klinis utama yang berkontribusi terhadap status komorbiditas. Diharapkan, hasil

penelitian dapat mendukung tenaga medis dalam pengambilan keputusan yang lebih cepat, tepat, dan personal, serta memberikan kontribusi pada pengembangan sistem pendukung keputusan medis berbasis kecerdasan buatan. pendukung keputusan klinis.

## 1.2 Analisa Permasalahan

Permasalahan penelitian ini berfokus pada kebutuhan identifikasi status komorbiditas pada pasien batu empedu karena keberadaan komorbid (misalnya diabetes, hipertensi, hiperlipidemia, penyakit jantung, dan gangguan endokrin) dapat memperberat kondisi klinis, meningkatkan risiko komplikasi, serta memengaruhi pemilihan terapi dan luaran pasien. Namun, penilaian komorbiditas dalam praktik masih sering bertumpu pada intuisi dan pengalaman klinisi, sementara data klinis modern bersifat kompleks sehingga pendekatan subjektif berpotensi tidak konsisten dan sulit direplikasi. Di sisi lain, penelitian terkait batu empedu lebih banyak menitikberatkan prediksi kejadian batu empedu, sehingga pemetaan status komorbiditas yang menyertai pasien batu empedu masih relatif kurang dieksplorasi, padahal penting untuk intervensi dini dan perencanaan penanganan jangka panjang. Selain itu, evaluasi model menunjukkan tantangan utama pada sensitivitas deteksi pasien berkomorbid (recall 45,16%) meskipun presisi tinggi, yang mengindikasikan risiko false negative sehingga sebagian pasien berkomorbid dapat tidak teridentifikasi dan tidak memperoleh perhatian klinis tambahan yang diperlukan.

## 1.3 Ruang Lingkup Penelitian

Penelitian ini mencakup evaluasi kinerja algoritma Gradient Boosted Trees (GBT) untuk memprediksi status komorbiditas pada pasien batu empedu menggunakan dataset publik *UCI Machine Learning Repository* yang berisi 319 individu dengan 38 fitur klinis. Data yang dianalisis dibatasi pada fitur tabular yang dikelompokkan menjadi demografis (mis. usia, jenis kelamin, BMI), bioimpedansi (mis. *total body water*, *visceral fat area*, *hepatic fat accumulation*), dan parameter laboratorium (mis. glukosa, profil lipid, enzim hati, kreatinin/GFR, CRP, hemoglobin, vitamin D), tanpa melibatkan data pencitraan maupun intervensi klinis. Variabel target adalah atribut “*Comorbidity*” yang merepresentasikan tingkat komorbiditas, lalu pada tahap pemodelan dilakukan transformasi agar sesuai kebutuhan klasifikasi (termasuk konversi ke format binomial untuk skema prediksi status). Implementasi dilakukan di RapidMiner dengan alur kerja pembersihan/transformasi data, pembagian data train-test menggunakan *stratified sampling*, pelatihan GBT (dengan *10-fold cross-validation*), serta evaluasi menggunakan *akurasi*, *classification error*, *weighted mean precision*, dan *weighted mean recall*.

## 1.4 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah: (1) bagaimana kinerja model Gradient Boosted Trees dalam memprediksi status komorbiditas pada pasien batu empedu berdasarkan fitur demografis, bioimpedansi, dan parameter laboratorium pada dataset yang digunakan; (2) sejauh mana performa model tersebut jika diukur menggunakan metrik evaluasi akurasi, *classification error*, *precision*, dan *recall*, serta (3) apakah model memiliki kemampuan yang memadai dalam mengidentifikasi pasien yang benar-benar berkomorbid (mengurangi false negative) sehingga hasil prediksi dapat mendukung kebutuhan pemetaan risiko komorbiditas secara lebih objektif dan konsisten.

## 1.5 Urgensi Penelitian

Urgensi penelitian ini terletak pada kebutuhan klinis untuk mengidentifikasi status komorbiditas pada pasien batu empedu secara lebih dini dan objektif, karena komorbid seperti diabetes, hipertensi, hiperlipidemia, penyakit jantung, maupun gangguan endokrin dapat memperberat kondisi pasien, meningkatkan risiko komplikasi, serta memengaruhi keputusan terapi dan luaran perawatan. Di praktik, penilaian komorbiditas masih sering bertumpu pada intuisi dan pengalaman klinisi, padahal data klinis yang tersedia bersifat kompleks sehingga berisiko menimbulkan ketidakkonsistenan dan sulit direplikasi tanpa pendekatan yang terstandar.

Selain itu, evaluasi awal menunjukkan tantangan penting pada kemampuan model dalam menangkap pasien yang benar-benar berkomorbid (*recall* komorbiditas masih rendah), yang berimplikasi pada risiko *false negative* dan potensi pasien berisiko tidak terdeteksi. Karena itu, penelitian ini mendesak untuk menilai secara sistematis kinerja Gradient Boosted Trees serta menegaskan area perbaikan agar prediksi komorbiditas dapat lebih andal sebagai dukungan pengambilan keputusan klinis.

### 1.6 Tujuan Penelitian

Secara umum, penelitian ini bertujuan untuk mengevaluasi kinerja model klasifikasi yang interpretable guna memprediksi status komorbiditas pada pasien batu empedu berbasis data klinis. Secara khusus, tujuan penelitian adalah:

1. Membangun model klasifikasi menggunakan algoritma Gradient Boosted Trees dengan memanfaatkan data demografis, bioimpedansi, dan parameter laboratorium dari dataset publik UCI Machine Learning Repository.
2. Mengevaluasi kinerja model Gradient Boosted Trees dalam memprediksi status komorbiditas menggunakan metrik akurasi, *classification error*, *weighted mean precision*, dan *weighted mean recall*.
3. Menganalisis hasil evaluasi untuk mengidentifikasi kekuatan dan keterbatasan model, khususnya dalam mendeteksi pasien yang benar-benar berkomorbid (meminimalkan *false negative*), sebagai dasar pengembangan rekomendasi perbaikan model agar lebih andal untuk pemetaan risiko klinis.

### 1.7 Manfaat Penelitian

Adapun manfaat dari penelitian ini:

1. Memberikan dukungan objektif dalam identifikasi status komorbiditas pada pasien batu empedu, mengingat komorbiditas dapat memperberat kondisi klinis serta memengaruhi keputusan terapi dan luaran perawatan.
2. Menjadi dasar skrining awal dan prioritasasi pasien yang membutuhkan perhatian klinis tambahan dengan menekan risiko kasus berkomorbid yang tidak terdeteksi (*false negative*), sejalan dengan temuan *recall* komorbiditas yang masih rendah.
3. Menyediakan baseline kinerja model Gradient Boosted Trees pada prediksi status komorbiditas melalui metrik evaluasi (akurasi, *classification error*, *precision*, *recall*) sebagai rujukan untuk pengembangan dan optimasi model pada penelitian lanjutan.
4. Memperkuat kontribusi akademik terkait penerapan machine learning pada data klinis tabular serta mendorong penilaian komorbiditas yang lebih terstandar dan tidak semata bergantung pada intuisi klinisi.

## II. LANDASAN TEORI

### 2.1 Risiko Komorbiditas pada Pasien Batu Empedu

Penyakit batu empedu (kolelitiasis) merupakan gangguan hepatobilier yang pembentukannya bersifat multifaktorial dan melibatkan interaksi faktor genetik, metabolik, hormonal, serta gaya hidup.

Kompleksitas faktor pembentuk tersebut berkorelasi dengan tingginya kemungkinan pasien batu empedu mengalami kondisi medis penyerta (komorbiditas). Pada populasi pasien batu empedu, komorbiditas seperti diabetes melitus, hiperlipidemia, penyakit jantung koroner, hipertensi, dan hipotiroidisme sering ditemukan bersamaan. Keberadaan komorbiditas ini tidak hanya menambah beban klinis, tetapi juga meningkatkan kompleksitas manajemen dan berpotensi memperburuk luaran jangka panjang, termasuk risiko komplikasi pascaoperasi, perubahan respons terapi, perpanjangan durasi rawat inap, serta penurunan kualitas hidup.

Dalam konteks operasional penelitian, risiko komorbiditas pada pasien batu empedu direpresentasikan melalui variabel target “*Comorbidity*” yang mengelompokkan tingkat komorbiditas dan mencakup kondisi seperti penyakit arteri koroner, hipotiroidisme, hiperlipidemia, dan diabetes mellitus, sehingga dapat dianalisis secara terukur dalam pemodelan klasifikasi

## 2.2 Data Mining dan Machine Learning dalam Bidang Medis

Analitik data kesehatan merupakan penerapan konsep data mining untuk mengekstraksi pola bermakna dari data klinis berskala besar, seperti rekam medis elektronik, hasil laboratorium, dan pemeriksaan penunjang. Data mining umumnya mencakup tahapan pemahaman masalah klinis, seleksi dan pembersihan data, transformasi fitur, pemodelan, serta evaluasi dan interpretasi hasil. Dalam konteks medis, pendekatan ini penting karena data pasien bersifat heterogen, multidimensi, dan sering mengandung nilai hilang, variabilitas pengukuran, serta distribusi kelas yang tidak seimbang.

Machine learning adalah bagian dari data mining yang berfokus pada pengembangan algoritma yang dapat “belajar” dari data untuk melakukan klasifikasi atau prediksi. Pada aplikasi klinis, machine learning banyak digunakan untuk deteksi dini penyakit, prediksi risiko komplikasi, estimasi prognosis, dan stratifikasi pasien berbasis risiko. Keunggulan utamanya adalah kemampuan memodelkan hubungan kompleks antarvariabel yang sulit ditangkap oleh pendekatan statistik sederhana. Namun, implementasi di bidang kesehatan menuntut perhatian khusus pada validasi yang ketat, interpretabilitas model, dan aspek etika/privasi data. Oleh karena itu, pemanfaatan machine learning untuk memprediksi status komorbiditas pada pasien batu empedu diposisikan sebagai upaya analitik yang mendukung penilaian risiko secara lebih objektif, konsisten, dan terukur.

## 2.3 Gradient Boosted Trees sebagai Model Klasifikasi

Gradient Boosted Trees (GBT) adalah metode ensemble berbasis boosting yang membangun sejumlah pohon keputusan secara bertahap, di mana setiap pohon baru dilatih untuk memperbaiki kesalahan prediksi dari model sebelumnya melalui optimasi fungsi kerugian (*loss function*). Pendekatan ini efektif untuk data tabular karena mampu menangkap pola non-linear dan interaksi antarfitur yang sering muncul pada data klinis, seperti keterkaitan antara parameter metabolik, inflamasi, dan fungsi organ.

Kinerja GBT dipengaruhi oleh pengaturan hiperparameter, antara lain jumlah pohon (number of trees), kedalaman pohon (*max depth*), *learning rate*, dan mekanisme regularisasi. Walaupun GBT merupakan *ensemble* (lebih kompleks daripada pohon tunggal), model ini tetap dapat dianalisis secara interpretatif melalui importance fitur, analisis kontribusi fitur terhadap prediksi, serta metode penjelasan model (misalnya SHAP atau *partial dependence*). Karakteristik ini penting dalam aplikasi medis karena klinisi memerlukan dasar penalaran yang dapat ditelusuri untuk memahami faktor apa yang paling berperan dalam prediksi status komorbiditas.

## 2.4 Evaluasi Kinerja Model Klasifikasi

Evaluasi kinerja model klasifikasi dilakukan untuk menilai ketepatan prediksi dan kemampuan model membedakan kelas target. Evaluasi umumnya merujuk pada *confusion matrix* yang terdiri dari *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Metrik yang digunakan meliputi:

### 1. Akurasi (*Accuracy*)

Mengukur proporsi prediksi yang benar terhadap seluruh prediksi. Akurasi memberikan gambaran umum performa model, namun dapat kurang informatif bila distribusi kelas tidak seimbang.

### 2. Classification Error

Merupakan komplement dari akurasi, yaitu proporsi prediksi yang salah terhadap seluruh prediksi. Nilai ini menggambarkan ruang perbaikan model secara global.

### 3. Precision

Precision per kelas mengukur seberapa besar proporsi prediksi positif pada kelas tertentu yang benar-benar berasal dari kelas tersebut. Precision yang tinggi menunjukkan bahwa model jarang memberikan “alarm palsu” (*false positive*) untuk kelas tersebut.

### 4. Recall (Sensitivity)

*Recall* per kelas mengukur kemampuan model dalam mengenali seluruh sampel aktual pada kelas tersebut. Nilai recall yang tinggi penting terutama pada kelas risiko tinggi, karena berkaitan dengan kemampuan model mendeteksi pasien yang benar-benar membutuhkan pemantauan intensif.

#### 5. Weighted Mean Precision dan Weighted Mean Recall

Pada klasifikasi multi-kelas, *precision* dan *recall* per kelas dapat dirata-ratakan dengan bobot tertentu (misalnya bobot yang sama untuk tiap kelas) untuk menghasilkan gambaran performa rata-rata model secara lebih seimbang antar kelas.

### 2.5 Interpretabilitas Model dan Implikasi Klinis Prediksi

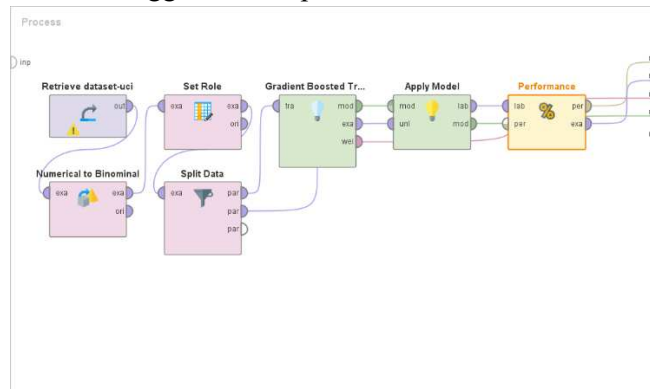
Dalam penerapan machine learning di bidang kesehatan, interpretabilitas menjadi krusial karena keluaran model tidak hanya dinilai dari akurasi, tetapi juga dari sejauh mana prediksi dapat dijelaskan secara rasional dan dapat dipertanggungjawabkan dalam konteks klinis. Model dengan kinerja tinggi tetapi sulit dijelaskan berisiko menurunkan tingkat kepercayaan pengguna, terutama ketika hasil prediksi digunakan untuk skrining, penentuan prioritas pemantauan, atau penyusunan tindak lanjut pasien. Karena itu, evaluasi prediksi status komorbiditas perlu disertai penjelasan mengenai faktor klinis yang paling berkontribusi terhadap hasil prediksi agar klinisi dapat menilai kesesuaian prediksi dengan pengetahuan dan praktik klinis.

Pada model Gradient Boosted Trees, interpretasi dapat diperoleh melalui analisis pentingnya fitur untuk mengidentifikasi variabel yang paling berpengaruh, serta metode penjelasan pada tingkat individu pasien untuk menunjukkan kontribusi masing-masing variabel terhadap keluaran prediksi. Selain itu, implikasi klinis juga berkaitan dengan pengaturan ambang keputusan yang tepat, terutama bila tujuan utama adalah meminimalkan pasien berkomorbid yang tidak terdeteksi. Dengan demikian, prediksi komorbiditas sebaiknya diposisikan sebagai dukungan analitik yang terukur, dengan penekanan pada keterjelasan alasan prediksi, konsistensi hasil, dan relevansi klinis dari variabel yang digunakan.

## III. METODOLOGI PENELITIAN

### 3.1 Desain Penelitian dan Alur Kerja Pemodelan

Penelitian ini menggunakan pendekatan kuantitatif berbasis data mining untuk mengevaluasi kinerja model Gradient Boosted Trees (GBT) dalam mengklasifikasikan status komorbiditas pada pasien dengan dataset Gallstone-1. Implementasi pemodelan dilakukan menggunakan RapidMiner, yang memungkinkan seluruh tahapan pengolahan data dilakukan secara sistematis mulai dari pengambilan dataset, prapemrosesan, pembagian data, pelatihan model, hingga evaluasi performa.



Gambar III.1 Alur Pemodelan

Secara umum, tahapan *workflow* pemodelan terdiri atas: (1) pengambilan dataset Gallstone-1, (2) transformasi data numerik menjadi biner, (3) penetapan atribut target (label), (4) pembagian data menjadi data latih dan data uji, (5) pelatihan model menggunakan algoritma GBT, (6) penerapan model pada data uji untuk menghasilkan prediksi, dan (7) evaluasi performa berdasarkan metrik klasifikasi.

### 3.2 Data dan Sumber Data

Dataset yang digunakan berasal dari UCI Machine Learning Repository dengan nama Gallstone-1. Dataset ini dikumpulkan dari *Internal Medicine Outpatient Clinic, Ankara VM Medical Park Hospital* pada periode Juni 2022–Juni 2023. Secara keseluruhan, dataset mencakup 319 individu, di mana 161 pasien terdiagnosis batu empedu. Data disajikan dalam format tabular, sehingga setiap baris merepresentasikan satu individu/pasien, sedangkan setiap kolom merepresentasikan atribut yang menggambarkan kondisi klinis pasien. Dataset memiliki 38 atribut yang mencakup variabel demografis, indikator bioimpedansi, serta parameter laboratorium. Kombinasi atribut tersebut digunakan untuk menangkap kondisi kesehatan pasien secara komprehensif, termasuk informasi dasar (misalnya usia/jenis kelamin bila tersedia), status klinis, serta indikator pemeriksaan penunjang yang dapat berhubungan dengan komorbiditas. Dataset telah melalui proses pembersihan pada sumbernya dan dilaporkan tidak mengandung nilai hilang (no missing values), sehingga penelitian dapat berfokus pada pemodelan tanpa memerlukan imputasi nilai kosong. Selain itu, pengumpulan data dilaporkan mengikuti ketentuan etik dari *Ankara City Hospital Ethics Committee (E2-23-4632)*.

Tabel III.1 Ringkasan Dataset Gallstone

Parameter	Nilai
Sumber	UCI Machine
Nama Dataset	Differentiated Thyroid Cancer Recurrence
Jumlah Instance	319
Jumlah Atribut	38 Atribut (Demografis, bioimpedansi, laboratorium)
Tipe Data	Tabular (Campuran numerik & Kategorikal)
Task	Klasifikasi Komordibitas
Missing Value	Tidak Ada
Tools Pemodelan	Rapid Miner

Ringkasan pada Tabel 3.1 menunjukkan karakteristik dataset yang digunakan sebagai dasar tahap prapemrosesan dan pemodelan pada subbab berikutnya.

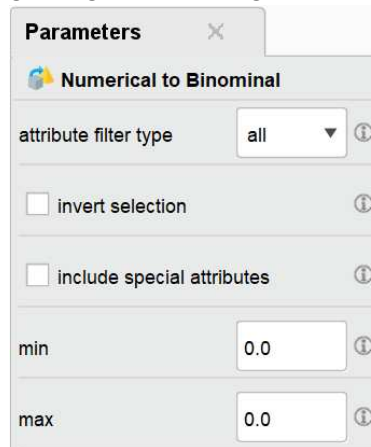
### 3.3 Prapemrosesan Data

Tahap prapemrosesan bertujuan memastikan format data sesuai dengan kebutuhan pemodelan klasifikasi di RapidMiner serta memastikan peran atribut (prediktor dan label) telah ditetapkan dengan benar. Prapemrosesan pada penelitian ini difokuskan pada pembentukan label biner komorbiditas dan penetapan peran atribut, sehingga alur pelatihan dan evaluasi model berjalan konsisten.

Langkah prapemrosesan yang dilakukan adalah sebagai berikut.

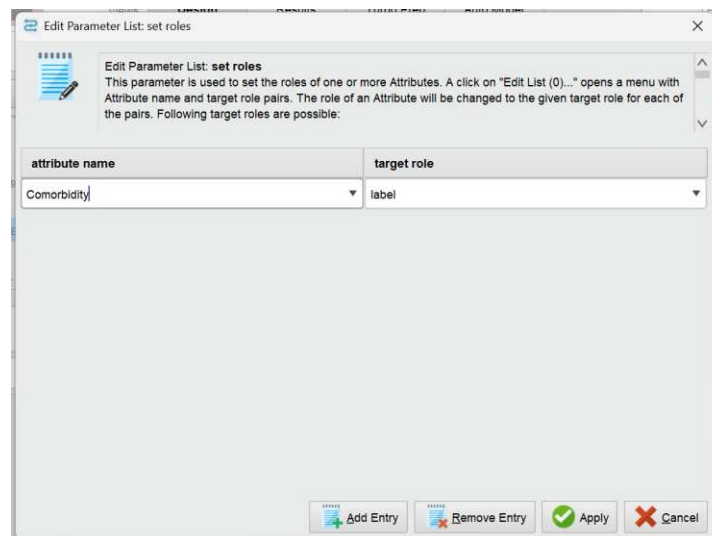
1. Transformasi atribut target menjadi biner menggunakan operator *Numerical to Binominal*. Operator ini mengubah atribut *Comorbidity* yang semula bernilai numerik (0–3) menjadi label

*binominal* dengan konfigurasi  $min = 0$  dan  $max = 0$ . Dengan konfigurasi tersebut, nilai 0 dipisahkan dari nilai selain 0, sehingga terbentuk dua kelas: tidak ada komorbiditas (0) dan ada komorbiditas (1–3). Transformasi ini penting agar target sesuai dengan skenario klasifikasi biner.



Gambar III.2 Konfigurasi operator *Numerical to Binomial*

2. Penetapan peran atribut menggunakan operator *Set Role*, dengan menetapkan hasil transformasi komorbiditas sebagai label (target klasifikasi) dan atribut lainnya sebagai *predictor (regular attributes)*. Penetapan ini memastikan RapidMiner menggunakan atribut yang tepat sebagai keluaran yang diprediksi dan atribut lainnya sebagai masukan model.



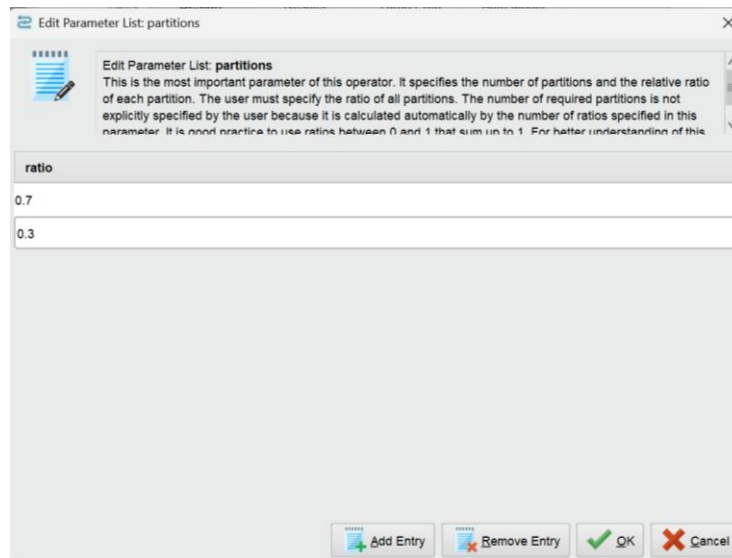
Gambar III.3 Penetapan Targer Role penelitian

3. Pemeriksaan tipe data dan konsistensi nilai, yaitu memastikan tipe atribut telah sesuai untuk diproses oleh operator GBT (misalnya atribut numerik dipertahankan sebagai numerik, dan atribut kategorikal tetap terbaca sebagai nominal/binominal). Tahap ini juga memastikan tidak ada atribut yang tidak terdefinisi dengan benar sebagai prediktor.

Output tahap prapemrosesan adalah dataset yang telah memiliki label biner komorbiditas dan struktur atribut yang siap digunakan pada tahap pembagian data, pelatihan, pengujian, serta evaluasi performa.

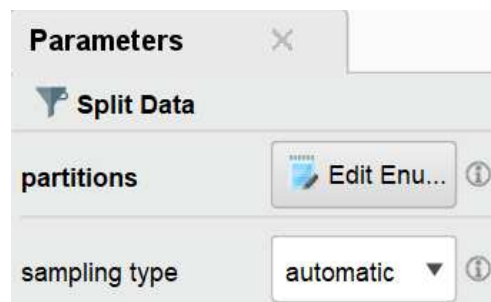
### 3.4 Pembagian Data dan Pembangunan Model

Setelah prapemrosesan, dataset dibagi menjadi data latih (*training*) dan data uji (*testing*) menggunakan skema *hold-out* dengan rasio 70% : 30%. Pembagian dilakukan menggunakan operator *Split Data* pada RapidMiner dengan parameter *sampling type = automatic* serta partisi 0.7 dan 0.3 sebagaimana ditunjukkan pada Gambar III.4. Penggunaan *sampling type* otomatis bertujuan untuk membagi data secara praktis sesuai parameter partisi sehingga proses pelatihan dan pengujian dapat dilakukan tanpa campur tangan manual pada pemilihan *record*.



Gambar III.4 Partisi splitt data model

Penggunaan *sampling type* otomatis bertujuan untuk membagi data secara praktis sesuai parameter partisi sehingga proses pelatihan dan pengujian dapat dilakukan tanpa campur tangan manual pada pemilihan *record*.



Gambar III.5 Partisi splitt data model

Pembagian 70:30 dipilih karena dataset yang digunakan adalah dataset medis yang biasanya tidak seimbang, selain itu juga untuk mencegah terjadinya *overfitting* atau model lambat belajar.

#### 3.4.1 Pembangunan Model Gradient Boosted Trees

Model klasifikasi dibangun menggunakan algoritma Gradient Boosted Trees (GBT). GBT merupakan metode *ensemble boosting* yang membangun model secara bertahap melalui penambahan pohon keputusan (*decision trees*) secara berurutan. Inti dari boosting adalah bahwa setiap pohon baru dilatih untuk

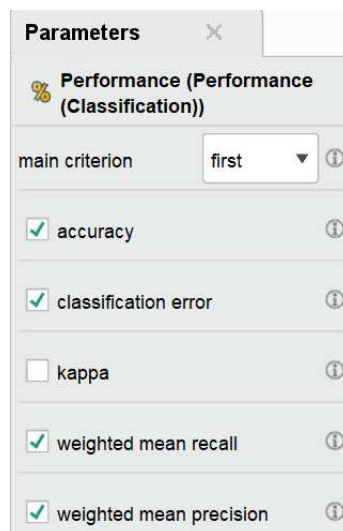
memperbaiki kesalahan prediksi dari model sebelumnya. Dengan pendekatan ini, GBT mampu menangkap hubungan non-linear antar atribut, meningkatkan akurasi, dan umumnya menghasilkan performa yang lebih stabil dibandingkan model pohon tunggal. Secara umum, proses GBT dapat dipahami melalui tahapan: (1) membangun model awal, (2) menghitung kesalahan prediksi model terhadap label, (3) melatih pohon berikutnya untuk memodelkan pola kesalahan tersebut, dan (4) menggabungkan kontribusi seluruh pohon menjadi model akhir. Dalam implementasi RapidMiner, model GBT dilatih pada data latih hingga menghasilkan model terlatih (output model) yang kemudian digunakan untuk melakukan prediksi pada data uji.

### 3.4.2 Penerapan Model data uji

Model terlatih diterapkan pada data uji menggunakan operator *Apply Model*. Tahap ini menghasilkan dua keluaran utama: (1) label prediksi komorbiditas untuk setiap record data uji, dan (2) nilai *confidence* untuk tiap kelas yang menggambarkan tingkat keyakinan model terhadap prediksi. Keluaran ini selanjutnya menjadi masukan pada tahap evaluasi performa. Tahapan pelatihan dan penerapan model tersebut tercakup dalam *workflow* utama sebagaimana ditampilkan pada Gambar 3.1.

### 3.5 Matrik Evaluasi Model

Evaluasi performa dilakukan untuk menilai kemampuan model GBT dalam mengklasifikasikan status komorbiditas pada data uji. Evaluasi dilakukan menggunakan operator Performance (*Classification*) pada RapidMiner dengan parameter sebagaimana ditampilkan pada Gambar 3.4. Metrik yang digunakan mencakup *accuracy*, *classification error*, *weighted mean recall*, dan *weighted mean precision*. Pemilihan metrik berbobot dilakukan untuk memberikan ringkasan performa yang lebih representatif, terutama ketika distribusi kelas tidak sepenuhnya seimbang.



Gambar III.6 Parameter Classification

Parameter evaluasi pada operator Performance (*Classification*), (*accuracy*, *classification error*, *weighted mean recall*, *weighted mean precision*). Perhitungan metrik didasarkan pada *confusion matrix* yang terdiri atas komponen:

Secara matematis, akurasi didefinisikan sebagai perbandingan antara jumlah prediksi yang benar dan total jumlah prediksi pada data pengujian.

### 3.5.1 Metrik Akurasi

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

#### Keterangan :

TP (True Positives) adalah jumlah sampel yang sebenarnya positif dan diprediksi model sebagai positif dengan benar

TN (True Negatives) adalah jumlah sampel yang sebenarnya negatif dan diprediksi model sebagai negatif dengan benar

FP (False Positives) adalah jumlah sampel yang sebenarnya negatif tetapi diprediksi model secara keliru sebagai positif

FN (False Negatives) adalah jumlah sampel yang sebenarnya positif tetapi diprediksi model secara keliru sebagai negative

### 3.5.2 Metrik Classification Error

$$CE = \frac{E}{N} \times 100\% \quad (2)$$

#### Keterangan :

CE (Classification Error) : Tingkat kesalahan klasifikasi model (proporsi prediksi yang salah).

E : Jumlah prediksi salah

N : Jumlah data uji

Nilai klasifikasi *error* menunjukkan seberapa besar proporsi prediksi yang masih keliru dan menjadi indikator ruang perbaikan bagi model.

### 3.5.3 Metrik Weighted Mean Precision

Precision dihitung terlebih dahulu untuk setiap kelas pada label komorbiditas (misalnya tidak ada komorbiditas dan ada komorbiditas). *Precision* menunjukkan proporsi prediksi pada suatu kelas yang benar, yaitu perbandingan antara jumlah data yang benar diprediksi sebagai kelas tersebut dengan seluruh data yang diprediksi sebagai kelas tersebut. Untuk kelas ke-*i*, precision didefinisikan sebagai:

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \quad (3)$$

#### Keterangan :

$TP_i$  : True Positives untuk kelas ke-*i*

$FP_i$  : False Positives untuk kelas ke-*i*

Dengan  $TP_i$  menyatakan jumlah *true positive* pada kelas ke- $i$ , dan  $FP_i$  menyatakan jumlah *false positive* pada kelas ke- $i$ . Nilai precision untuk setiap kelas komorbiditas kemudian dirata-ratakan dengan bobot tertentu, sehingga menghasilkan *weighted mean precision*.

$$\text{Weighted Mean Precision} = \frac{\sum_{i=1}^k w_i \cdot \text{Precision}_i}{\sum_{i=1}^k w_i} \quad (4)$$

Keterangan :

*Weighted mean precision* dihitung sebagai rata-rata precision dari seluruh kelas dengan mempertimbangkan bobot masing-masing kelas. Pada perhitungan ini,  $k$  menyatakan jumlah kelas pada label komorbiditas dan  $w_i$  merupakan bobot untuk kelas ke- $i$ . Dalam pengaturan RapidMiner yang digunakan, bobot setiap kelas dapat diatur sama, misalnya  $W_i = 1$  untuk semua kelas, sehingga nilai *weighted mean precision* menjadi *ekuivalen* dengan rata-rata *precision* per kelas.

### 3.5.4 Metrik Weighted Mean Recall

*Recall* per kelas adalah perbandingan antara jumlah data yang benar diprediksi berada pada kelas ke- $i$  dengan seluruh data yang secara aktual berasal dari kelas tersebut:

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (5)$$

Nilai *recall* per kelas kemudian dirata-ratakan dengan bobot tertentu sehingga diperoleh *weighted mean recall*:

$$\text{Weighted Mean Recall} = \sum_{i=1}^k w_i \cdot \text{Recall}_i \quad (6)$$

Keterangan :

$k$  : Jumlah kelas atau kategori pada dataset.

$w_i$  : Bobot untuk kelas ke- $i$ , biasanya proporsional terhadap jumlah data aktual pada kelas tersebut.

Jika bobot  $w_i$  ditetapkan sama untuk setiap kelas, maka *weighted mean recall* merepresentasikan rata-rata *recall* antar kelas tanpa dipengaruhi perbedaan jumlah data pada tiap kelas (setiap kelas memiliki kontribusi yang setara). Sebaliknya, jika bobot  $w_i$  ditentukan berdasarkan jumlah data aktual pada masing-masing kelas, maka *weighted mean recall* menjadi ringkasan performa yang lebih mencerminkan distribusi kelas pada dataset.

## IV. ANALISIS DAN HASIL PERANCANGAN

Penerapan algoritma Gradient Boosted Trees (GBT) pada dataset Gallstone-1 menghasilkan model klasifikasi biner untuk memprediksi status komorbiditas, yaitu *false* (tidak ada komorbiditas) dan *true* (ada komorbiditas). Kinerja model dievaluasi menggunakan data uji (testing set) yang tidak dilibatkan dalam proses pelatihan. Evaluasi performa ditampilkan oleh RapidMiner dalam empat keluaran utama, yaitu *accuracy*, *classification error*, *weighted mean precision*, dan *weighted mean recall*, yang masing-masing disertai *confusion matrix* sebagai dasar perhitungan metrik. Oleh karena itu, hasil pengujian pada penelitian

ini dirangkum pada Tabel 1 sampai dengan Tabel 4, yang memuat nilai metrik serta matriks klasifikasi untuk memudahkan interpretasi performa model.

*Confusion matrix* pada seluruh tabel memiliki format yang sama, dengan baris merepresentasikan kelas prediksi (*pred. false* dan *pred. true*) dan kolom menunjukkan kelas aktual (*true false* dan *true true*). Berdasarkan *confusion matrix* tersebut, total data uji berjumlah 96 data, dengan komposisi 65 data pada kelas aktual false dan 31 data pada kelas aktual *true*. Prediksi yang tepat berada pada diagonal utama, yaitu 65 data *false* yang diprediksi false dan 14 data *true* yang diprediksi *true*. Sementara itu, kesalahan klasifikasi tampak pada nilai di luar diagonal, yaitu terdapat 17 data aktual *true* yang diprediksi sebagai *false*, sedangkan kesalahan tipe sebaliknya tidak ditemukan (data aktual *false* yang diprediksi *true* bernilai 0). Dengan demikian, komponen *confusion matrix* dapat diringkas sebagai  $TN = 65$ ,  $FP = 0$ ,  $FN = 17$ , dan  $TP = 14$ .

Berdasarkan hasil tersebut, nilai *accuracy* yang diperoleh adalah 82,29% dengan *classification error* 17,71%, yang menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar. Dari sisi ketepatan prediksi per kelas, nilai *class precision* untuk prediksi *true* mencapai 100% (tidak ada false positive), sedangkan *precision* untuk prediksi *false* sebesar 79,27%. Namun demikian, kemampuan model dalam menangkap kasus *true* masih terbatas, tercermin dari *class recall true* = 45,16%, sementara *class recall false* = 100%. Nilai weighted mean precision 89,63% dan weighted mean recall 72,58% (dengan bobot kelas 1,1) merepresentasikan rata-rata kinerja antar kelas dan menjadi indikator ringkas untuk membandingkan performa model secara keseluruhan pada subbab pembahasan berikutnya.

#### 4.1 Hasil Akurasi Model

Tabel IV.1 menampilkan *confusion matrix* model GBT beserta nilai akurasi keseluruhan sebesar 82,29%. Pada kelas tanpa komorbiditas (*true false*), terdapat 65 sampel yang diprediksi benar sebagai tidak berkomorbid (*pred. false*) dan tidak ada satupun sampel kelas ini yang salah diprediksi sebagai berkomorbid (*pred. true* = 0). Hal ini menghasilkan *class recall* sebesar 100,00% untuk kelas tanpa komorbiditas, yang berarti seluruh pasien yang benar-benar tidak memiliki komorbiditas berhasil dikenali dengan tepat oleh model. Pada kelas dengan komorbiditas (*true true*), terdapat 17 sampel yang salah diprediksi sebagai tidak berkomorbid (*pred. false*) dan 14 sampel yang diprediksi benar sebagai berkomorbid (*pred. true*). *Recall* untuk kelas komorbiditas tercatat sebesar 45,16%, menunjukkan bahwa kurang dari setengah pasien yang memiliki komorbiditas berhasil dideteksi oleh model.

**Tabel IV.1 Hasil model akurasi**

Akurasi: 82.29%	True false	True true	Class precision
Pred. false	65	17	79.27%
Pred. true	0	14	100.00%
Class recall	100.00%	45.16%	

Dari sisi *precision*, kelas tanpa komorbiditas memiliki nilai *precision* sebesar 79,27%, yang berarti sekitar 79% dari seluruh prediksi “tidak berkomorbid” memang benar tidak memiliki komorbiditas. Sementara itu, *precision* untuk kelas komorbiditas mencapai 100,00%, menandakan bahwa setiap kali model memprediksi adanya komorbiditas, prediksi tersebut selalu benar. Kombinasi *recall* yang tinggi pada kelas tanpa komorbiditas dan *precision* yang sempurna pada kelas komorbiditas menghasilkan akurasi keseluruhan 82,29%. Nilai ini menunjukkan performa model yang cukup baik secara agregat, namun masih meninggalkan ruang perbaikan terutama dalam sensitivitas (*recall*) terhadap pasien yang benar-benar memiliki komorbiditas.

#### 4.2 Hasil Classification Error

Tabel IV.2 menggambarkan metrik *classification error* yang diperoleh sebesar 17,71%. Nilai ini merupakan komplement dari akurasi dan merepresentasikan proporsi sampel yang diklasifikasikan secara keliru oleh model. Dengan struktur *confusion matrix* yang sama seperti pada Tabel 1, *classification error* 17,71% berarti bahwa sekitar satu dari enam prediksi model masih salah. Kesalahan ini seluruhnya berasal dari pasien dengan komorbiditas (*true true*) yang diklasifikasikan sebagai tidak berkomorbid (17 kasus), sedangkan tidak ditemukan kasus sebaliknya.

Tabel IV.2 Hasil model classification error

<b>Classification Error :</b> <b>17.71%</b>	<b>True false</b>	<b>True true</b>	<b>Class precision</b>
Pred. false	65	17	79.27%
Pred. true	0	14	100.00%
Class recall	100.00%	45.16%	

Secara klinis, pola *error* seperti ini menunjukkan bahwa model cenderung “konservatif” dalam mendeteksi komorbiditas: model sangat jarang memberikan *false positive* (memvonis ada komorbiditas padahal tidak), tetapi cukup sering menghasilkan *false negative* (menyatakan tidak ada komorbiditas padahal ada). Dari perspektif praktik klinis, hal ini perlu dicermati karena pasien berkomorbid yang tidak teridentifikasi berpotensi tidak mendapatkan perhatian tambahan yang seharusnya. Oleh karena itu, meskipun *classification error* relatif rendah, penting untuk mempertimbangkan metrik lain yang lebih sensitif terhadap kelas minoritas.

#### 4.3 Hasil Weighted mean precision

Tabel IV.3 menyajikan nilai *weighted mean precision* sebesar 89,63%. Metrik ini merupakan rata-rata precision untuk masing-masing kelas yang dibobot berdasarkan proporsi jumlah sampel di setiap kelas. Dengan *precision* kelas tanpa komorbiditas 79,27% dan precision kelas komorbiditas 100,00%, nilai *weighted mean precision* yang tinggi menunjukkan bahwa secara umum model memiliki ketepatan prediksi yang sangat baik, terutama ketika menyatakan bahwa seorang pasien memiliki komorbiditas.

Tabel IV.3 Hasil model weighted mean precision

<b>Weighted Mean Precision:</b> <b>89.63% 1,1</b>	<b>True false</b>	<b>True true</b>	<b>Class precision</b>
Pred. false	65	17	79.27%
Pred. true	0	14	100.00%
Class recall	100.00%	45.16%	

Dalam konteks prediksi klinis, *weighted mean precision* yang tinggi berarti bahwa sebagian besar hasil prediksi yang dikeluarkan model, baik “berkomorbid” maupun “tidak berkomorbid”, dapat dipercaya dari sisi kebenaran label. Secara khusus, *precision* 100% pada kelas komorbiditas mengindikasikan tidak adanya pasien yang sebenarnya tidak berkomorbid tetapi salah diklasifikasikan sebagai berkomorbid. Hal ini penting untuk menghindari *over-treatment* atau kekhawatiran yang tidak perlu pada pasien yang sebenarnya sehat dari sisi komorbiditas. Namun, metrik *precision* yang tinggi ini perlu tetap diseimbangkan dengan *recall*, karena model bisa saja hanya mengidentifikasi sebagian kecil pasien berkomorbid tetapi dengan keyakinan sangat tinggi..

#### 4.4 Hasil Weighted mean recall

Tabel IV.4 menampilkan nilai *weighted mean recall* sebesar 72,58%. Metrik ini merupakan rata-rata *recall* per kelas yang dibobot berdasarkan proporsi jumlah sampel di setiap kelas. Dengan *recall* kelas tanpa komorbiditas 100,00% dan *recall* kelas komorbiditas 45,16%, nilai *weighted mean recall* 72,58% mencerminkan bahwa secara keseluruhan model cukup baik dalam menangkap sampel dari kedua kelas, tetapi masih kurang sensitif terhadap pasien yang benar-benar berkomorbid.

SS

Tabel IV.4 Hasil model weighted mean recall

Weighted Mean Recall: 72.58% 1,1	True false	True true	Class precision
Pred. false	65	17	79.27%
Pred. true	0	14	100.00%
Class recall	100.00%	45.16%	

Secara praktis, *recall* 100% pada kelas tanpa komorbiditas menunjukkan bahwa tidak ada pasien tanpa komorbiditas yang salah diklasifikasikan sebagai berkomorbid (tidak ada *false positive* dari sisi klinis). Namun *recall* 45,16% pada kelas komorbiditas berarti bahwa lebih dari setengah pasien yang memiliki komorbiditas tidak terdeteksi oleh model dan diklasifikasikan sebagai tidak berkomorbid. Bagi pengambilan keputusan klinis, kondisi ini mengindikasikan bahwa model lebih cocok digunakan sebagai alat bantu untuk mengkonfirmasi komorbiditas ketika model memprediksi “*true*”, tetapi kurang ideal jika digunakan sebagai satu-satunya alat skrining untuk mendeteksi semua pasien berkomorbid.

#### 4.5 Interpretasi Kesalahan

Secara keseluruhan, kombinasi akurasi 82,29%, *classification error* 17,71%, *weighted mean precision* 89,63%, dan *weighted mean recall* 72,58% menunjukkan bahwa model Gradient Boosted Trees memiliki performa yang cukup baik dalam memprediksi komorbiditas pada pasien batu empedu. Model sangat andal dalam menghindari *false positive* untuk komorbiditas (*precision* 100% pada kelas komorbiditas) dan tidak pernah salah mengklasifikasikan pasien tanpa komorbiditas sebagai berkomorbid (*recall* 100% pada kelas tanpa komorbiditas). Namun, kelemahan utama model adalah kemampuan yang masih terbatas dalam menangkap seluruh pasien yang benar-benar berkomorbid (*recall* kelas komorbiditas 45,16%).

Dari sudut pandang klinis, hasil ini mengindikasikan bahwa model lebih tepat digunakan sebagai alat bantu untuk mengidentifikasi pasien dengan komorbiditas yang “jelas” secara pola data (ketika model memprediksi ada komorbiditas, kemungkinan besar benar), daripada sebagai alat skrining yang diharapkan mampu menangkap semua kasus komorbiditas. Untuk penelitian lanjutan, beberapa strategi yang dapat dipertimbangkan untuk meningkatkan sensitivitas antara lain penyesuaian *threshold* keputusan, penggunaan *class weighting* atau resampling untuk menyeimbangkan distribusi kelas, serta eksplorasi konfigurasi parameter GBT yang lebih agresif terhadap kelas minoritas. Dengan demikian, model diharapkan dapat mencapai kompromi yang lebih baik antara *precision* dan *recall* pada kelas komorbiditas tanpa mengorbankan akurasi keseluruhan secara signifikan.

## V. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Penelitian ini mengimplementasikan algoritma Gradient Boosted Trees untuk memprediksi status komorbiditas pada pasien batu empedu menggunakan dataset klinis dari *UCI Machine Learning Repository* yang mencakup 319 individu dengan 38 fitur. Proses metodologis dilakukan secara terstruktur meliputi prapemrosesan data, pembagian data dengan rasio 70 banding 30, serta evaluasi menggunakan empat metrik utama. Hasil evaluasi menunjukkan akurasi 82,29 persen, *classification error* 17,71 persen, *weighted mean precision* 89,63 persen, dan *weighted mean recall* 72,58 persen.

Berdasarkan confusion matrix, model memiliki keunggulan pada ketepatan prediksi kelas komorbiditas karena *precision* pada kelas komorbiditas mencapai 100,00 persen, serta kemampuan mengenali kelas tanpa komorbiditas yang sangat baik karena *recall* pada kelas tanpa komorbiditas mencapai 100,00 persen. Namun, kelemahan utama terdapat pada *recall* kelas komorbiditas yang hanya 45,16 persen, yang menunjukkan bahwa lebih dari separuh pasien yang benar benar berkomorbid tidak terdeteksi dan diprediksi sebagai tidak berkomorbid. Pola ini mengindikasikan adanya kecenderungan bias terhadap kelas mayoritas dan keterbatasan sensitivitas pada kelas komorbiditas.

Secara metodologis, penelitian ini menunjukkan bahwa Gradient Boosted Trees layak digunakan untuk klasifikasi komorbiditas pada data tabular klinis karena mampu menangani kombinasi fitur dan menangkap hubungan non linear antar variabel, serta memungkinkan identifikasi fitur berpengaruh melalui analisis *feature importance*. Dari perspektif klinis, model lebih tepat diposisikan sebagai alat bantu untuk mengonfirmasi adanya komorbiditas ketika model memprediksi komorbiditas, dibandingkan sebagai alat skrining utama yang menuntut deteksi menyeluruh pada seluruh pasien berkomorbid.

### 5.2 Saran

Penelitian lanjutan disarankan memprioritaskan peningkatan sensitivitas deteksi komorbiditas dengan menerapkan teknik penanganan ketidakseimbangan kelas seperti SMOTE atau class *weighting*. Selain itu, penyesuaian *threshold* keputusan perlu dipertimbangkan agar model dapat mencapai kompromi yang lebih baik antara *precision* dan *recall* pada kelas komorbiditas sesuai kebutuhan klinis.

Dari sisi optimasi, disarankan melakukan penalaan hyperparameter Gradient Boosted Trees secara lebih ekstensif, misalnya melalui *grid search* atau *Bayesian optimization*, sehingga konfigurasi model tidak hanya mengoptimalkan akurasi keseluruhan tetapi juga memperbaiki kinerja pada kelas komorbiditas. Untuk memperkuat posisi metode, penelitian berikutnya juga perlu membandingkan performa dengan algoritma ensemble alternatif seperti XGBoost, LightGBM, atau CatBoost.

Untuk meningkatkan relevansi praktis, disarankan menambahkan analisis interpretabilitas yang lebih sistematis agar kontribusi fitur terhadap prediksi mudah ditelusuri oleh pengguna klinis. Selain itu, validasi pada data dari sumber berbeda atau periode yang berbeda akan membantu menilai generalisasi model dan mengurangi risiko bias yang mungkin muncul akibat karakteristik dataset tunggal. Jika model akan diarahkan ke penggunaan klinis, studi lanjutan sebaiknya menilai dampak kesalahan *false negative* terhadap alur layanan dan menyesuaikan konfigurasi evaluasi serta *threshold* berdasarkan prioritas keselamatan pasien.

## DAFTAR PUSTAKA

- [1] I. Esen, H. Arslan, S. A. Esen, M. Gülşen, N. Kültekin, and O. Özdemir, “Early prediction of gallstone disease with a machine learning-based method from bioimpedance and laboratory data,” *Medicine*, vol. 103, 2024. DOI: 10.1097/MD.00000000000043567
- [2] E. Omar, H. Mat, A. Z. Abd Karim, R. Sanaudi, F. Ibrahim, M. A. Omar, et al., “Comparative analysis of logistic regression, gradient boosted trees, SVM, and random forest algorithms for prediction of acute kidney injury requiring dialysis after cardiac surgery,” *Int. J. Nephrol. Renovasc. Dis.*, vol. 17, pp. 197–204, 2024. DOI: 10.2147/IJNRD.S438932
- [3] J. Zhou, S. Lee, Y. Liu, T. Liu, G. Tse, and Q. Zhang, “Predicting stroke and mortality in mitral regurgitation: a gradient boosting approach,” *Front. Cardiovasc. Med.*, 2021. DOI: 10.1101/2021.01.04.21249215
- [4] M. Chen, Z. Wang, Z. Zhao, W. Zhang, X. Guo, and J. Shen, “Task-wise split gradient boosting trees for multi-center diabetes prediction,” in *Proc. ACM SIGKDD Conf.*, 2021. DOI: 10.1145/3447548.3467402
- [5] F. Yagin, İ. Cicek, and Z. Kucukakcali, “Classification of stroke with gradient boosting tree using SMOTE-based oversampling method,” *Medicine Science Int. Med. J.*, vol. 10, no. 2, pp. 567–573, 2021. DOI: 10.5455/medscience.2021.03.015
- [6] J. Gutierrez, M. Volkovs, T. Poutanen, T. Watson, and L. Rosella, “Risk stratification for COVID-19 hospitalization: a multivariable model based on gradient-boosting decision trees,” *CMAJ Open*, vol. 9, no. 4, pp. E1223–E1231, 2021. DOI: 10.9778/cmajo.20210036
- [7] G. V. Aiosa, M. Palesi, and F. Sapuppo, “Explainable AI for decision support to obesity comorbidities diagnosis,” *IEEE Access*, vol. 11, pp. 107767–107782, 2023. DOI: 10.1109/ACCESS.2023.3320057
- [8] E. Etu, L. Monplaisir, S. Arslanturk, S. Masoud, C. Aguwa, I. Markevych, and J. Miller, “Prediction of length of stay in the emergency department for COVID-19 patients: A machine learning approach,” *IEEE Access*, vol. 10, pp. 42229–42237, 2022. DOI: 10.1109/ACCESS.2022.3168045
- [9] M.-L. Nielsen, T. Petersen, J. Maul, J. J. Wu, M. Rasmussen, T. Bertelsen, et al., “Multivariable predictive models to identify the optimal biologic therapy for treatment of patients with psoriasis at the individual level,” *JAMA Dermatol.*, vol. 158, no. 12, pp. 1380–1388, 2022. DOI: 10.1001/jamadermatol.2022.3171
- [10] R. Aznar-Gimeno, L. Esteban, G. Labata-Lezaun, R. del-Hoyo-Alonso, D. Abadía-Gallego, J. Paño-Pardo, et al., “A clinical decision web to predict ICU admission or death for patients hospitalised with COVID-19 using machine learning algorithms,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 16, 2021. DOI: 10.3390/ijerph18168677
- [11] A. W. Boerman, M. Schinkel, L. Meijerink, E. S. van den Ende, L. C. Pladet, M. Scholtemeijer, et al., “Using machine learning to predict blood culture outcomes in the emergency department: a single-centre, retrospective, observational study,” *BMJ Open*, vol. 12, no. 2, 2022. DOI: 10.1136/bmjopen-2021-053332
- [12] X. Zhang, Z. Xi, Y. Zhao, and F. Liu, “Forecasting hospitalization for adult asthma patients in emergency departments based on multiple environmental and clinical factors,” *J. Asthma Allergy*, vol. 18, pp. 861–876, 2025. DOI: 10.2147/JAA.S512405
- [13] R. Kasambara and M. Kamndaya, “Application of Extreme Gradient Boosting to predict NCD-HIV/AIDS comorbidity in young adults in Malawi,” *medRxiv preprint*, 2025. DOI: 10.1101/2025.08.05.25333013