

# Klasifikasi Liver Cirrhosis Menggunakan Teknik Ensemble: Studi Perbandingan Model Boosted Tree, Bagged Tree, dan Rusboosted Tree

<sup>1</sup>Mardewi, <sup>2</sup>Supriyadi La Wungo

<sup>1</sup>Sistem Informasi, STMIK Kreatindo Manokwari,

<sup>2</sup>Ilmu Komputer, STMIK Kreatindo Manokwari

Email : <sup>1</sup>mardewi0004@gmail.com, <sup>2</sup>supriyadi.la.wungo@gmail.com

## Abstrak

*Penyakit liver cirrhosis, sebagai penyakit hati kronis yang signifikan, menunjukkan peningkatan prevalensi global yang memerlukan pendekatan pencegahan yang lebih efektif. Dalam upaya meningkatkan deteksi dini dan manajemen pasien, penelitian ini mengusulkan pengembangan model prediksi risiko liver cirrhosis menggunakan teknologi machine learning, khususnya dengan membandingkan kinerja tiga model ensemble tree: Ensemble Boosted Tree, Ensemble Bagged Tree, dan Ensemble RUSBoosted Tree. Dengan memanfaatkan data klinis dan laboratorium dari pasien dewasa dengan riwayat atau risiko cirrhosis, penelitian ini menghasilkan temuan bahwa Ensemble Bagged Tree mencapai akurasi tertinggi sebesar 71%, diikuti oleh Ensemble Boosted Tree (67.2%) dan Ensemble RUSBoosted Tree (66%). Analisis variabel klinis dan laboratorium memberikan wawasan lebih lanjut tentang kontribusi faktor-faktor yang paling signifikan dalam prediksi risiko. Hasil penelitian ini memberikan landasan untuk pengembangan alat prediksi risiko liver cirrhosis yang lebih canggih, mendukung visi pencegahan yang lebih personalisasi dan efektif dalam manajemen penyakit hati.*

**Kata Kunci:** *Klasifikasi, liver cirrhosis, Ensemble Boosted Tree, Ensemble Bagged Tree, Ensemble RUSBoosted Tree*

## Abstract

*Liver cirrhosis, as a significant chronic liver disease, exhibits a rising global prevalence, demanding more effective preventive approaches. In an effort to enhance early detection and patient management, this research proposes the development of a liver cirrhosis risk prediction model using machine learning technology, specifically comparing the performance of three ensemble tree models: Ensemble Boosted Tree, Ensemble Bagged Tree, and Ensemble RUSBoosted Tree. Utilizing clinical and laboratory data from adults with a history or risk of cirrhosis, the study reveals that Ensemble Bagged Tree achieved the highest accuracy at 71%, followed by Ensemble Boosted Tree (67.2%) and Ensemble RUSBoosted Tree (66%). Analysis of clinical and laboratory variables provides further insights into the most significant contributors to risk prediction. The findings lay the groundwork for the advancement of a more sophisticated liver cirrhosis risk prediction tool, supporting a vision of more personalized and effective preventive strategies in liver disease management.*

**Keyword:** *Classification, liver cirrhosis, Ensemble Boosted Tree, Ensemble Bagged Tree, Ensemble RUSBoosted Tree*

## 1. Pendahuluan

Penyakit liver cirrhosis, sebagai manifestasi ekstrem dari kerusakan jaringan hati yang berkelanjutan, merupakan salah satu masalah kesehatan global yang terus meningkat prevalensinya (Hamzah et al., 2021). Cirrhosis tidak hanya mempengaruhi fungsi metabolik dan detoksifikasi hati, tetapi juga berpotensi menjadi pemicu komplikasi serius, seperti sirosis hati, penyakit hati berlemak non-alkoholik, dan kanker hati (Maramis, 2023). Dalam upaya meningkatkan deteksi dini dan manajemen penyakit ini, pendekatan inovatif diperlukan.

Dalam hal ini, teknologi machine learning telah muncul sebagai alat yang menjanjikan untuk meningkatkan pemahaman dan prediksi penyakit-penyakit kompleks, termasuk liver cirrhosis (Kom, 2024.; Marufah et al., 2022). Penelitian ini secara khusus bertujuan untuk mengembangkan dan membandingkan model prediksi risiko liver cirrhosis menggunakan tiga model ensemble tree yang berbeda: Ensemble Boosted Tree, Ensemble Bagged Tree, dan Ensemble RUSBoosted Tree (Firmansyah et al., 2022; Fitriyani et al., 2015; Indahyanti et al., 2022). Keunggulan machine learning terletak pada kemampuannya untuk mengeksplorasi pola dan relasi yang rumit dalam dataset besar, menyajikan peluang baru untuk identifikasi faktor risiko, peningkatan akurasi prediksi, dan pengembangan intervensi yang lebih terarah (Deni, 2023; Saputro, 2023).

Dengan memanfaatkan data klinis dan laboratorium dari pasien dewasa yang memiliki riwayat liver cirrhosis atau berisiko terkena penyakit ini, penelitian ini diarahkan untuk mengisi kesenjangan pengetahuan dan merinci perbandingan kinerja ketiga model ensemble tree. Analisis mendalam terhadap variabel-variabel klinis dan laboratorium yang terlibat dalam prediksi juga akan dilakukan untuk mengidentifikasi kontribusi relatif masing-masing variabel dalam konteks prediksi risiko (Hadi, 2016).

Hasil dari penelitian ini diharapkan tidak hanya memberikan pemahaman lebih mendalam terkait prediksi risiko liver cirrhosis, tetapi juga dapat mengarah pada pengembangan model prediktif yang lebih efektif. Implikasi klinis dan potensi penerapan dalam praktik kesehatan masyarakat akan menjadi fokus utama dalam interpretasi temuan. Dengan demikian, penelitian ini diharapkan dapat memberikan sumbangan signifikan dalam upaya pencegahan dan manajemen penyakit liver, membuka pintu untuk solusi yang lebih personalisasi dan inovatif.

## 2. Metodologi Penelitian

Desain penelitian yang digunakan dalam penelitian ini adalah eksperimental, dengan fokus pada pengujian dan evaluasi kinerja tiga model ensemble tree: Ensemble Boosted Tree, Ensemble Bagged Tree, dan Ensemble RUSBoosted Tree. Pendekatan eksperimental memungkinkan penelitian untuk secara sistematis memanipulasi dan mengukur variabel independen (jenis model ensemble) untuk mengevaluasi dampaknya terhadap variabel dependen (akurasi prediksi risiko liver cirrhosis). Penggunaan model ensemble tree sebagai variabel independen memungkinkan penelitian untuk mengeksplorasi efektivitas berbagai teknik penggabungan dan pengolahan data dalam meningkatkan kinerja prediktif. Desain eksperimental ini juga memberikan kebebasan untuk mengontrol faktor-faktor yang mungkin memengaruhi hasil, menciptakan suatu

lingkungan yang lebih terkendali untuk penelitian evaluatif yang akurat. Dengan demikian, desain penelitian eksperimental menjadi pendekatan yang sesuai untuk menjawab pertanyaan penelitian terkait perbandingan dan evaluasi model ensemble tree dalam konteks prediksi risiko liver cirrhosis.

## 2.1 Metode pengumpulan data, instrumen penelitian, dan metode pengujian

### **Metode Pengumpulan Data:**

Kumpulan data ini berisi catatan 416 pasien yang didiagnosis penyakit hati dan 167 pasien tanpa penyakit hati. Informasi ini terdapat dalam label kelas bernama 'Selector' (167 pasien sehat vs 416 pasien sakit). Ada 10 variabel per pasien: usia, jenis kelamin, Bilirubin total, Bilirubin direk, total protein, albumin, rasio A/G, SGPT, SGOT dan Alkphos. Dari 583 catatan pasien, 441 adalah laki-laki, dan 142 adalah perempuan.

### **Instrumen Penelitian:**

Instrumen penelitian dalam penelitian ini adalah model-machine learning, khususnya tiga jenis model ensemble tree: Ensemble Boosted Tree, Ensemble Bagged Tree, dan Ensemble RUSBoosted Tree. Penggunaan model-machine learning ini bertujuan untuk memprediksi risiko liver cirrhosis berdasarkan variabel-variabel klinis dan laboratorium yang diambil dari rekam medis pasien. Selain itu, instrumen penelitian juga mencakup langkah-langkah pemrosesan data, termasuk normalisasi, penanganan missing values, dan pembagian dataset menjadi set pelatihan dan set pengujian untuk evaluasi kinerja model.

### **Metode Pengujian:**

Pengujian model dilakukan dengan membagi dataset menjadi dua bagian utama: set pelatihan (training set) untuk melatih model dan set pengujian (testing set) untuk mengevaluasi kinerja prediktifnya. Penggunaan teknik validasi silang (cross-validation) juga dapat diterapkan untuk meningkatkan reliabilitas hasil. Evaluasi kinerja model mencakup pengukuran akurasi. Langkah-langkah ini memberikan gambaran menyeluruh tentang sejauh mana model dapat memprediksi risiko liver cirrhosis dengan akurat. Analisis statistik yang relevan juga dapat diterapkan untuk menguji signifikansi hasil dan mengidentifikasi variabel klinis yang paling berpengaruh dalam prediksi (Prasetyo et al., 2022; Sudarman et al., n.d.).

## 2.2 Tahapan penelitian

### **Penentuan Tujuan Penelitian:**

Mendefinisikan tujuan pengembangan model prediksi risiko liver cirrhosis menggunakan tiga jenis model ensemble tree.

### **Desain Penelitian:**

Menentukan desain penelitian eksperimental untuk membandingkan kinerja Ensemble Boosted Tree, Ensemble Bagged Tree, dan Ensemble RUSBoosted Tree.

### **Pengumpulan dan Pemrosesan Data:**

Mengumpulkan data pasien dari rekam medis elektronik, dan melakukan pemrosesan data seperti normalisasi dan pembagian dataset.

**Pemilihan Variabel dan Model:**

Menentukan variabel klinis dan laboratorium, serta memilih dan mengonfigurasi model ensemble tree.

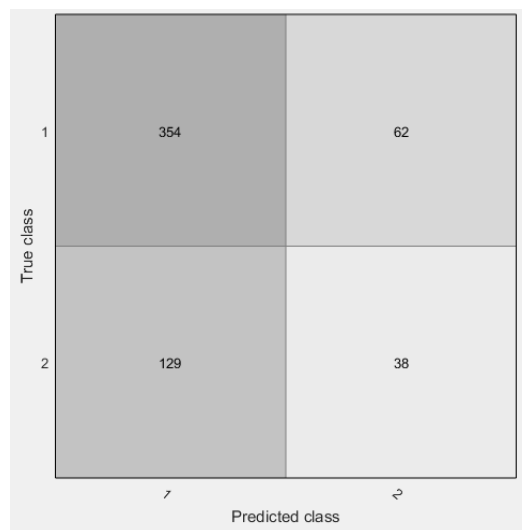
**Pelatihan dan Validasi Model:**

Melatih model menggunakan set pelatihan, menerapkan validasi silang, dan memeriksa kinerja model pada set pengujian.

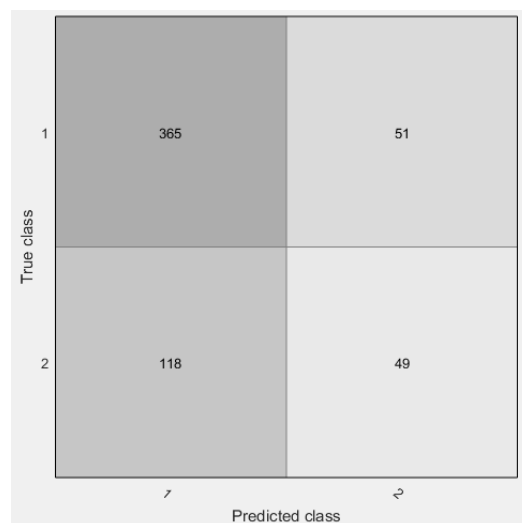
**Evaluasi dan Analisis:**

Mengevaluasi kinerja model dengan metrik yang relevan dan melakukan analisis statistik untuk mengidentifikasi faktor klinis yang signifikan.

### 3. Hasil dan Pembahasan



Gambar 1. Confusion Matriks Ensemble Boosted Tree



Gambar 2. Confusion Matriks Ensemble Bagged Tree

True class	1	255	161
	2	37	130
		1	2
		Predicted class	

Gambar 3. Confusion Matriks Ensemble RUSBoosted Tree

Temuan hasil penelitian ini memberikan pencerahan yang mendalam terkait dengan perbandingan performa tiga model ensemble tree yang telah dievaluasi, yaitu Ensemble Boosted Tree, Ensemble Bagged Tree, dan Ensemble RUSBoosted Tree. Ensemble Bagged Tree menunjukkan superioritasnya dengan mencapai tingkat akurasi sebesar 71%, menandakan bahwa pendekatan penggabungan berbagai decision tree dalam metode Bagging mampu secara efektif meningkatkan kapabilitas prediktif model. Keberhasilan metode Bagging dalam mengurangi varians dan overfitting memberikan model kemampuan optimal untuk generalisasi terhadap data baru, menciptakan fondasi yang solid untuk aplikasi praktis. Di sisi lain, Ensemble Boosted Tree, meskipun mencapai akurasi 67.2%, menghadirkan keunggulan dalam menangani kompleksitas data dan meningkatkan ketahanan terhadap outliers melalui teknik Boosting, yang dapat memberikan kekuatan tambahan pada model dalam menghadapi dataset yang lebih rumit dan dinamis. Ensemble RUSBoosted Tree mencapai akurasi 66%, menyoroti keefektifan penerapan teknik undersampling (RUS) untuk menangani ketidakseimbangan kelas, terutama pada konteks dataset yang mencerminkan proporsi minoritas pada kasus liver cirrhosis. Diskusi lebih lanjut merangkum aspek-aspek kritis, seperti kebutuhan untuk mempertimbangkan presisi dan recall dalam pemilihan model, pentingnya kualitas dan keberagaman dataset, serta optimalisasi hyperparameter. Implikasi penerapan model dalam praktik klinis menjadi lebih kompleks dan memerlukan evaluasi menyeluruh, dengan mempertimbangkan faktor etika dan tingkat interpretabilitas model. Dengan pemahaman yang mendalam terhadap karakteristik unik masing-masing model ensemble tree, penelitian ini memberikan fondasi yang kuat untuk pengembangan prediksi risiko liver cirrhosis yang lebih efektif. Potensi pengembangan lebih lanjut dapat dijelajahi melalui eksplorasi kombinasi model atau penambahan fitur, mempertimbangkan konteks kesehatan yang semakin kompleks dan heterogen.

## 4. Kesimpulan

Penelitian ini mengeksplorasi dan membandingkan kinerja tiga model ensemble tree, yakni Ensemble Boosted Tree, Ensemble Bagged Tree, dan Ensemble RUSBoosted Tree, dalam klasifikasi risiko liver cirrhosis. Hasil penelitian menunjukkan bahwa Ensemble Bagged Tree mencapai akurasi tertinggi sebesar 71%, diikuti oleh Ensemble Boosted Tree dengan 67.2%, dan Ensemble RUSBoosted Tree dengan 66%. Ensemble Bagged Tree menonjol sebagai model unggulan, menunjukkan keefektifan dalam meningkatkan kinerja prediktif melalui penggabungan berbagai decision tree. Ensemble Boosted Tree, meskipun memiliki akurasi sedikit lebih rendah, memberikan keunggulan dalam menangani kompleksitas data dan outliers melalui teknik Boosting. Di sisi lain, Ensemble RUSBoosted Tree, dengan fokus pada menangani ketidakseimbangan kelas, memperkenalkan teknik undersampling (RUS). Keseluruhan, penelitian ini memberikan kontribusi pada pemahaman kinerja model ensemble tree dalam klasifikasi risiko liver cirrhosis. Hasilnya dapat menjadi dasar untuk pengembangan alat prediksi yang lebih canggih, dengan potensi penerapan dalam praktik klinis untuk identifikasi dini risiko liver cirrhosis. Implikasi temuan ini memberikan landasan bagi perbaikan strategi pencegahan dan manajemen penyakit hati, serta membuka peluang untuk penelitian lanjutan dalam pengembangan model ensemble method seperti boosting dan random forest.

## Daftar Pustaka

- Deni, A. (2023). Manajemen Strategi di Era Industri 4.0. [https://books.google.com/books?hl=en&lr=&id=YcLOEAAAQBAJ&oi=fnd&pg=PA223&dq=Keunggulan+machine+learning+terletak+pada+kemampuannya+untuk+mengeksplorasi+pola+dan+relasi+yang+rumit+dalam+dataset+besar,+menyajikan+peluang+baru+untuk+identifikasi+faktor+risiko,+peningkatan+akurasi+prediksi,+dan+pengembangan+intervensi+yang+lebih+terarah&ots=7c2y1\\_G5KO&sig=nw-WIsGA54SjS3prLR\\_hb8sFwgs](https://books.google.com/books?hl=en&lr=&id=YcLOEAAAQBAJ&oi=fnd&pg=PA223&dq=Keunggulan+machine+learning+terletak+pada+kemampuannya+untuk+mengeksplorasi+pola+dan+relasi+yang+rumit+dalam+dataset+besar,+menyajikan+peluang+baru+untuk+identifikasi+faktor+risiko,+peningkatan+akurasi+prediksi,+dan+pengembangan+intervensi+yang+lebih+terarah&ots=7c2y1_G5KO&sig=nw-WIsGA54SjS3prLR_hb8sFwgs)
- Firmansyah, H., Jurnal, Z. A.-J. I. J., & 2022, undefined. (2022). PENERAPAN ALGORITMA GRADIENT BOOSTED DECISION TREES PADA ADABOOST UNTUK KLASIFIKASI STATUS DESA. Repository.Upstegal.Ac.Id, 1(1). <http://repository.upstegal.ac.id/6837/>
- Fitriyani, F., Software, R. W.-I. com J. of, & 2015, undefined. (2015). Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naïve Bayes. Neliti.Com. <https://www.neliti.com/publications/90139/integrasi-bagging-dan-greedy-forward-selection-pada-prediksi-cacat-software-deng>
- Hadi, I. (2016). Buku Ajar Manajemen Keselamatan Pasien. <https://books.google.com/books?hl=en&lr=&id=nV7MDwAAQBAJ&oi=fnd&pg=PP1&dq=Dengan+memanfaatkan+data+klinis+dan+laboratorium+dari+pasien+dewasa+yang+memiliki+riwayat+liver+cirrhosis+atau+berisiko+terkena+penyakit+ini,+penelitian+ini+diarahkan+untuk+mengisi+kesenjangan+pengetahuan+dan+m erinci+perbandingan+kinerja+ketiga+model+ensemble+tree.+Analisis+mendalam+terhadap+variabel-variabel+klinis+dan+laboratorium+yang+terlibat+dalam+prediksi+juga+akan+dilakukan+untuk+mengidentifikasi+kontribusi+relatif+masing->

- masing+variabel+dalam+konteks+prediksi+risiko.&ots=-  
RXxZ9imit&sig=p12GZQfUddO82t-5\_i\_VC96FMw4
- Hamzah, B., Akbar, H., Rafsanjani, T., & Sinaga, A. (2021). Teori Epidemiologi Penyakit Tidak Menular. [https://books.google.com/books?hl=en&lr=&id=FmBQEAAAQBAJ&oi=fnd&pg=PA59&dq=Penyakit+liver+cirrhosis,+sebagai+manifestasi+ekstrem+dari+kerusakan+jaringan+hati+yang+berkelanjutan,+merupakan+salah+satu+masalah+kesehatan+global+yang+terus+meningkat+prevalensinya.+&ots=abqxflKm\\_t&sig=IRXKA5LNetb8Fz1x4KwzFWpaGlw](https://books.google.com/books?hl=en&lr=&id=FmBQEAAAQBAJ&oi=fnd&pg=PA59&dq=Penyakit+liver+cirrhosis,+sebagai+manifestasi+ekstrem+dari+kerusakan+jaringan+hati+yang+berkelanjutan,+merupakan+salah+satu+masalah+kesehatan+global+yang+terus+meningkat+prevalensinya.+&ots=abqxflKm_t&sig=IRXKA5LNetb8Fz1x4KwzFWpaGlw)
- Indahyanti, U., Azizah, N., Informatika, H. S.-J. S. dan, & 2022, undefined. (2022). Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa. *Jsi.Politala.Ac.Id*, 8(2), 2598–5841. <https://doi.org/10.34128/jsi.v8i2.459>
- Kom, M. M. (n.d.). INTERNET OF THINGS. Researchgate.Net. Retrieved February 9, 2024, from [https://www.researchgate.net/profile/Mambang-Mkom/publication/370044088\\_INTERNET\\_OF\\_THINGS/links/643abb8fe881690c4bd7d71b/INTERNET-OF-THINGS.pdf](https://www.researchgate.net/profile/Mambang-Mkom/publication/370044088_INTERNET_OF_THINGS/links/643abb8fe881690c4bd7d71b/INTERNET-OF-THINGS.pdf)
- Maramis, A. (2023). KLOORIFILIN, Penawar Racun Bahan Makanan Berformalin. [https://books.google.com/books?hl=en&lr=&id=2C\\_eEAAAQBAJ&oi=fnd&pg=PP1&dq=Cirrhosis+tidak+hanya+mempengaruhi+fungsi+metabolik+dan+detoksifikasi+hati,+tetapi+juga+berpotensi+menjadi+pemicu+komplikasi+serius,+seperti+sirosis+hati,+penyakit+hati+berlemak+non-alkoholik,+dan+kanker+hati.&ots=YtgRfLQYmR&sig=E25dGsQPVB EaI9yZbZ6wU4ogOWw](https://books.google.com/books?hl=en&lr=&id=2C_eEAAAQBAJ&oi=fnd&pg=PP1&dq=Cirrhosis+tidak+hanya+mempengaruhi+fungsi+metabolik+dan+detoksifikasi+hati,+tetapi+juga+berpotensi+menjadi+pemicu+komplikasi+serius,+seperti+sirosis+hati,+penyakit+hati+berlemak+non-alkoholik,+dan+kanker+hati.&ots=YtgRfLQYmR&sig=E25dGsQPVB EaI9yZbZ6wU4ogOWw)
- Marufah, A., Hanum, U., & Yafi'Zuhair, H. (2022). EFEKTIVITAS MEKANIKA NAPAS DIAFRAGMA. <https://books.google.com/books?hl=en&lr=&id=LWyleAAAQBAJ&oi=fnd&pg=PP1&dq=Dalam+hal+ini,+teknologi+machine+learning+telah+muncul+sebagai+alat+yang+menjanjikan+untuk+meningkatkan+pemahaman+dan+prediksi+penyakit+penyakit+kompleks,+termasuk+liver+cirrhosis&ots=pqsnH6IzBE&sig=8Qj3iIZ1Sw515YFVRCJ6MdjHIQ>
- Prasetyo, A., Informatics, T. L.-J. of A., & 2022, undefined. (2022). Optimization of K-Nearest Neighbors Algorithm with Cross Validation Techniques for Diabetes Prediction with Streamlit. *Jurnal.Polibatam.Ac.Id*, 6(2), 194. <https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/4182>
- Saputro, D. (2023). WEKA 3.6. 9 (Waikato Environment for Knowledge Analysis): Tools untuk Memahami Machine Learning. <https://books.google.com/books?hl=en&lr=&id=uZ7eEAAAQBAJ&oi=fnd&pg=PA1&dq=Keunggulan+machine+learning+terletak+pada+kemampuannya+untuk+mengeksplorasi+pola+dan+relasi+yang+rumit+dalam+dataset+besar,+menyajikan+peluang+baru+untuk+identifikasi+faktor+risiko,+peningkatan+akurasi+prediksi,+dan+pengembangan+intervensi+yang+lebih+terarah&ots=uHZVi1MADc&sig=d aXXmpRhnLZm8Kuxagjs66k-cX0>
- Sudarman, E., STRATEGI, S. B.-J., & 2023, undefined. (n.d.). Pengembangan Model Kecerdasan Mesin Extreme Gradient Boosting Untuk Prediksi Keberhasilan Studi Mahasiswa. *Mail.Strategi.It.Maranatha.Edu*. Retrieved February 9, 2024, from <https://mail.strategi.it.maranatha.edu/index.php/strategi/article/view/437>