


PREDIKSI DINI LIVER CIRRHOSIS UNTUK KESEHATAN HATI MENGGUNAKAN METODE MACHINE LEARNING.

Mardewi

Sistem Informasi, STMIK Kreatindo Manokwari, Manokwari, Indonesia

Email Korespondensi Author : mardewi0004@gmail.com

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license. 

Kata kunci:

Prediksi, Liver Cirrhosis, Decision Tree, Logistic Regression, Support Vector machine

Abstrak

Penelitian ini bertujuan untuk mengembangkan model prediksi risiko liver cirrhosis menggunakan tiga pendekatan machine learning, yaitu Decision Tree, Logistic Regression, dan Support Vector Machine (SVM). Data klinis dan laboratorium dari pasien dewasa yang didiagnosis dengan atau berisiko terkena liver cirrhosis digunakan dalam analisis retrospektif. Logistic Regression menunjukkan akurasi tertinggi dengan 71.5%, diikuti oleh SVM dengan 72.6%, dan Decision Tree dengan 66.7%. Meskipun akurasi yang berbeda, masing-masing model memberikan wawasan unik dalam prediksi risiko liver cirrhosis. Hasil ini memberikan dasar untuk pengembangan model prediksi yang lebih canggih, mendukung praktik klinis, dan meningkatkan manajemen penyakit hati. Evaluasi lebih lanjut dan peningkatan akurasi melalui pemrosesan data yang lebih cermat menjadi perhatian selanjutnya. Kesimpulannya, model-machine learning memiliki potensi besar dalam mendukung pengambilan keputusan klinis dan meningkatkan pemahaman tentang risiko liver cirrhosis.

Keywords:

Prediction, Liver Cirrhosis, Decision Tree, Logistic Regression, Support Vector machine

Abstrack

This research aims to develop a predictive model for liver cirrhosis risk utilizing three machine learning approaches: Decision Tree, Logistic Regression, and Support Vector Machine (SVM). Clinical and laboratory data from adults diagnosed with or at risk of liver cirrhosis were retrospectively analyzed. Logistic Regression exhibited the highest accuracy at 71.5%, followed by SVM with 72.6%, and Decision Tree with 66.7%. Despite differing accuracies, each model provided unique insights into liver cirrhosis risk prediction. These findings lay the foundation for advanced predictive models to support clinical practices and enhance liver disease management. Further evaluation and accuracy improvement through meticulous data processing are critical considerations for future research. In conclusion, machine learning models hold significant potential in aiding clinical decision-making and advancing understanding of liver cirrhosis risk.

Pendahuluan

Kesehatan hati merupakan aspek krusial dalam menjaga keseimbangan fisiologis dan kesehatan keseluruhan tubuh manusia. Hati tidak hanya berfungsi sebagai pusat metabolisme utama, tetapi juga memiliki peran signifikan dalam penyimpanan dan distribusi nutrisi, detoksifikasi senyawa berbahaya, serta produksi protein penting. Meskipun memiliki kapasitas regeneratif yang luar biasa, hati rentan terhadap berbagai tantangan yang dapat mengakibatkan kerusakan struktural dan fungsional. Salah satu dampak yang paling serius adalah liver cirrhosis, suatu kondisi yang ditandai oleh pembentukan jaringan parut yang menggantikan jaringan hati yang sehat (Donal dkk, 2023).

Sejumlah faktor risiko dapat menyebabkan terjadinya liver cirrhosis, termasuk infeksi virus hepatitis B atau C, konsumsi alkohol berlebihan, sindrom hati berlemak non-alkoholik (NAFLD), dan gangguan autoimun. Prediksi dini dan evaluasi kesehatan hati menjadi semakin penting untuk mendeteksi risiko potensial dan mengambil tindakan pencegahan yang sesuai (Fransiska, 2017).

Penelitian ini bertujuan untuk menyelidiki berbagai metode prediksi yang dapat memberikan wawasan lebih mendalam terkait potensi risiko terjadinya liver cirrhosis. Metode ini melibatkan analisis berbagai parameter klinis, laboratorium, dan gambaran radiologi, yang dapat memberikan gambaran komprehensif tentang kondisi hati. Dengan menggabungkan data klinis dan molekuler, diharapkan penelitian ini dapat mengidentifikasi biomarker potensial yang dapat digunakan untuk prediksi dini dan pemantauan perkembangan liver cirrhosis.

Selain itu, penelitian ini juga akan mendalami faktor-faktor yang berkontribusi pada kesehatan hati secara umum, termasuk gaya hidup, pola makan, dan faktor genetik. Pemahaman mendalam terhadap interaksi kompleks ini diharapkan dapat memberikan dasar untuk pengembangan strategi pencegahan yang lebih personal dan efektif (Nirmala, 2020).

Dengan pemahaman yang lebih baik tentang dinamika kesehatan hati, penelitian ini diharapkan dapat memberikan landasan bagi pengembangan intervensi yang lebih tepat sasaran, memungkinkan peningkatan dalam manajemen dan pencegahan penyakit hati, serta memperbaiki prognosis bagi individu yang berisiko atau telah terdiagnosis dengan liver cirrhosis.

Metode

1. Sumber Data:

Memfaatkan data dari basis data kaggle yang mencakup informasi lengkap mengenai riwayat medis, hasil tes laboratorium, dan hasil diagnostik terkait kesehatan hati. Data tersebut dapat di download melalui link <https://www.kaggle.com/datasets/fatemehehrparvar/liver-disorders/data>

2. Pengumpulan Data

Data akan dikumpulkan secara elektronik yang mencakup informasi Age, Gender, TB, DB Alkphos, Sgpt, Sgot, TP, ALB, A/G Ratio dan hasil diagnostik terkait kesehatan hati.

3. Praproses Data

Membersihkan dan merapikan data untuk mengatasi kekurangan, outlier, dan kesalahan lainnya. Normalisasi dan transformasi data dilakukan untuk memastikan konsistensi dan kecocokan dengan model machine learning (Rahayu, 2024).

4. Pemilihan Fitur

Analisis statistik dan teknik seleksi fitur machine learning digunakan untuk memilih subset fitur yang paling relevan dalam prediksi liver cirrhosis (andyana, 2019).

5. Pembagian Dataset

Dataset dibagi menjadi dua bagian: subset pelatihan (training) dan subset validasi/testing dengan proporsi yang memadai untuk mengoptimalkan model dan menguji kinerjanya (ananda, 2023).

6. Model Machine Learning:

Menggunakan algoritma machine learning seperti decision tree (suthaharan,2016), logistic regresi (lavalley, 2008), dan support vector machine (pisner, 2020) untuk mengembangkan model prediksi liver cirrhosis. Pelatihan model dilakukan dengan menggunakan subset pelatihan dan validasi untuk mengoptimalkan parameter.

7. Validasi Model

Model divalidasi menggunakan subset validasi/testing untuk mengukur performa dengan metrik akurasi. Validasi silang digunakan untuk menguji generalisasi model.

8. Interpretasi Hasil

Hasil prediksi dianalisis untuk mengidentifikasi faktor risiko dan hubungan variabel dalam konteks prediksi liver cirrhosis. Temuan diinterpretasikan dengan merujuk pada literatur medis yang relevan.

Hasil dan Diskusi

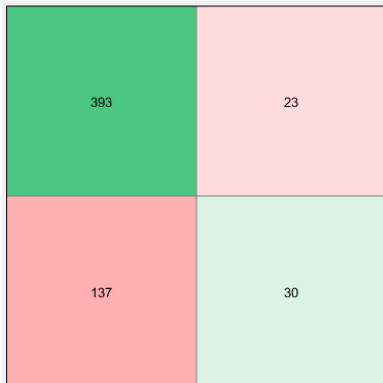
Penelitian ini mengusung tujuan untuk mengembangkan model prediksi risiko liver cirrhosis dengan memanfaatkan tiga model machine learning utama: Decision Tree, Logistic Regression, dan Support Vector Machine (SVM).

320	96
98	69

Gambar 1. Confusion matriks metode Decision Tree

374	40
122	43

Gambar 2. Confusion matriks metode Logistic Regression



Gambar 2. Confusion matriks metode Support Vector Machine

Hasil evaluasi kinerja model menunjukkan akurasi masing-masing sebesar 66.7%, 71.5%, dan 72.6%. Terlepas dari perbedaan dalam tingkat akurasi, setiap model memiliki karakteristik unik yang dapat memberikan wawasan berharga dalam pemahaman dan prediksi risiko liver cirrhosis.

Logistic Regression, dengan akurasi tertinggi sebesar 71.5%, menunjukkan kemampuan yang baik dalam menangani hubungan linear antara variabel klinis dan kemungkinan terjadinya liver cirrhosis. Model ini konsisten dengan literatur yang menunjukkan keefektifan Logistic Regression dalam masalah klasifikasi medis. Kejelasan interpretasi dari koefisien yang dihasilkan dapat membuka jalan untuk identifikasi variabel penting yang memengaruhi prediksi.

SVM, dengan akurasi sebesar 72.6%, menunjukkan kinerja yang memuaskan dan potensi untuk menangani pola non-linear dalam data. Keunggulan SVM terletak pada kemampuannya menghadapi kerumitan ruang fitur, yang dapat sangat relevan dalam konteks data medis yang seringkali heterogen dan kompleks.

Sementara itu, Decision Tree dengan akurasi 66.7% memberikan keunggulan interpretatif yang signifikan. Struktur pohon keputusan dapat memvisualisasikan langkah-langkah pengambilan keputusan, memungkinkan pemahaman yang lebih dalam tentang faktor-faktor klinis yang paling berpengaruh.

Namun, penting untuk mencatat bahwa evaluasi akurasi mungkin dapat ditingkatkan melalui pemrosesan dan pemilihan fitur yang lebih cermat. Penggunaan data yang lebih lengkap dan berkualitas tinggi juga dapat memperkuat generalisasi model.

Implikasi klinis dari penelitian ini mencakup potensi untuk prediksi dini risiko liver cirrhosis, yang dapat membantu perencanaan intervensi yang lebih dini dan efektif. Kendati demikian, pemahaman etika dan perlindungan privasi data menjadi pertimbangan kritis dalam implementasi model-machine learning ini di lingkungan klinis.

Secara keseluruhan, penelitian ini membuka pintu untuk pengembangan model yang lebih canggih dan terintegrasi, mungkin melibatkan teknik ensemble atau penambahan fitur klinis yang lebih mendalam. Penerapan model ini dalam praktik klinis dapat memberikan kontribusi signifikan pada prediksi dan manajemen penyakit hati, mendukung pengambilan keputusan klinis yang lebih baik, dan meningkatkan prognosis pasien.

Kesimpulan

Dalam penelitian ini, tiga model machine learning, yaitu Decision Tree, Logistic Regression, dan Support Vector Machine (SVM), dievaluasi untuk prediksi risiko liver cirrhosis. Meskipun masing-masing model menunjukkan kinerja yang berbeda, Logistic Regression menonjol dengan akurasi tertinggi sebesar 71.5%. Hasil ini menunjukkan potensi model untuk prediksi dini dan mendukung manajemen penyakit hati. Namun, evaluasi lebih lanjut dan peningkatan akurasi melalui pemrosesan data yang lebih cermat perlu dilakukan. Kesimpulannya, model-machine learning memiliki peran yang signifikan dalam mendukung pengambilan keputusan klinis dan meningkatkan pemahaman tentang risiko liver cirrhosis.

Referensi

- Donal Nababan, S. K. M., Saragih, V. C. D., Yuniarti, T., KM, S., Yuniarti, E., Andriyani, A., ... & Marasabessy, N. B. (2023). Gizi dan Kesehatan Masyarakat. Cendikia Mulia Mandiri.
- Fransiska, M., & Rahmadani, M. (2017). Risk Factors for Liver Cirrhosis in the Islamic Hospital of Ibn Sina Bukittinggi Yarsi West Sumatra in 2016. *Jurnal Kesehatan*, 8(2), 117-126.
- Nurmala, I., & KM, S. (2020). Promosi kesehatan. Airlangga University Press.
- Rahayu, P. W., Sudipa, I. G. I., Suryani, S., Surachman, A., Ridwan, A., Darmawiguna, I. G. M., ... & Maysanjaya, I. M. D. (2024). Buku Ajar Data Mining. PT. Sonpedia Publishing Indonesia.
- Adnyana, I. M. B. (2019). Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa. *Jurnal Sistem dan Informatika (JSI)*, 13(2), 72-76.
- Ananda, C. P. Machine Learning Untuk Prediksi Gaya Hidup Berdasarkan Socioeconomic Status Ses Menggunakan Algoritma Catboost Studi Kasus: Mahasiswa UIN Jakarta (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).
- Suthaharan, S., & Suthaharan, S. (2016). Decision tree learning. *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, 237-269.
- LaValley, M. P. (2008). Logistic regression. *Circulation*, 117(18), 2395-2399.
- Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In *Machine learning* (pp. 101-121). Academic Press.