

IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR UNTUK MENINGKATKAN MODEL KLASIFIKASI PENYAKIT LIVER

Muhammad Fathurohim¹, Cep Lukman Rohmat², Heliyanti Susana³, Mulyawan⁴.

Program Studi Teknik Informatika¹³
Program Studi Rekayasa Perangkat Lunak²
Program Studi Sistem Informasi⁴

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
mochamadfathurohim45@gmail.com

(*) Corresponding Author : mochamadfathurohim45@gmail.com
Published : 30 Mei 2026

Abstract— Liver disease is a major health problem with a relatively high prevalence and is often difficult to detect at an early stage due to nonspecific symptoms. Delayed diagnosis can cause the disease to progress to a more severe stage and increase the risk of mortality. Therefore, a data-driven approach is needed to support the classification of liver disease objectively and efficiently based on patients' medical data. This study aims to implement the K-Nearest Neighbor (KNN) algorithm to improve a liver disease classification model using clinical patient data. The research methodology follows the Knowledge Discovery in Database (KDD) framework, which consists of data selection, data preprocessing, exploratory data analysis (EDA), modeling, and evaluation stages. The dataset used in this study is the Indian Liver Patient Dataset (ILPD), consisting of 583 patient records with 10 feature attributes and 1 target attribute. Data preprocessing includes handling missing values using median imputation, encoding categorical data, normalizing features using StandardScaler, and splitting the dataset into training and testing data with an 80:20 ratio. Exploratory Data Analysis is conducted to analyze data characteristics, class distribution, outliers, and feature correlations. The classification model is built using the KNN algorithm with $k = 5$ and Euclidean Distance as the distance metric. The results show that the KNN algorithm can be successfully applied to classify liver disease, achieving an accuracy of 63.25%. Based on the classification report, the model performs relatively well in detecting liver disease patients, with a precision of 0.74 and a recall of 0.73 for the liver class. However, the performance for the non-liver class remains relatively low due to class imbalance in the dataset. The confusion matrix analysis also indicates the presence of false positive and false negative predictions. Overall, this study demonstrates that the KNN algorithm can serve as a baseline method for liver disease classification, although further optimization is required to improve model performance.

Keywords: Liver Disease, K-Nearest Neighbor, Data Mining, Classification, Machine Learning

Abstrak— Penyakit liver merupakan salah satu masalah kesehatan yang memiliki tingkat prevalensi cukup tinggi dan sering kali sulit dideteksi pada tahap awal karena gejala yang tidak spesifik. Keterlambatan diagnosis dapat menyebabkan penyakit berkembang ke tahap yang lebih serius dan berisiko mengancam jiwa. Oleh karena itu, diperlukan suatu pendekatan berbasis data yang mampu membantu proses klasifikasi penyakit liver secara cepat dan objektif berdasarkan data medis pasien. Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Nearest Neighbor (KNN) dalam meningkatkan model klasifikasi penyakit liver menggunakan data klinis pasien. Metode penelitian yang digunakan mengacu pada kerangka Knowledge Discovery in Database (KDD), yang meliputi tahapan seleksi data, pra-pemrosesan data, exploratory data analysis (EDA), pemodelan, dan evaluasi. Dataset yang digunakan adalah Indian Liver Patient Dataset (ILPD) yang terdiri dari 583 data pasien dengan 10 atribut fitur dan 1 atribut target. Tahap pra-pemrosesan data meliputi penanganan missing value menggunakan imputasi median, encoding data kategorikal, normalisasi data menggunakan StandardScaler, serta pembagian data menjadi data latih dan data uji dengan rasio 80:20. Analisis EDA dilakukan untuk memahami karakteristik data, distribusi kelas target, keberadaan outlier, serta korelasi antar fitur. Model klasifikasi dibangun menggunakan algoritma KNN dengan parameter $k = 5$ dan metode pengukuran jarak Euclidean Distance. Hasil penelitian menunjukkan bahwa algoritma KNN berhasil diterapkan untuk klasifikasi penyakit liver dengan nilai

akurasi sebesar 63,25%. Berdasarkan classification report, model menunjukkan performa yang cukup baik dalam mendeteksi pasien penderita penyakit liver dengan nilai precision sebesar 0,74 dan recall sebesar 0,73. Namun, performa pada kelas pasien non-liver masih relatif rendah akibat ketidakseimbangan kelas dalam dataset. Hasil confusion matrix juga menunjukkan masih adanya kesalahan false positive dan false negative. Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma KNN dapat digunakan sebagai metode dasar dalam klasifikasi penyakit liver, namun masih memerlukan optimasi lanjutan untuk meningkatkan kinerja model.

Kata Kunci: Penyakit Liver, K-Nearest Neighbor, Data Mining, Klasifikasi, Machine Learning

INTRODUCTION

Penyakit liver merupakan salah satu masalah kesehatan global yang terus mengalami peningkatan dari tahun ke tahun dan menjadi penyebab utama morbiditas serta mortalitas di berbagai negara. Organ liver memiliki peran vital dalam metabolisme tubuh, detoksifikasi racun, pengaturan komposisi darah, serta produksi empedu untuk proses pencernaan. Gangguan pada fungsi liver dapat berdampak serius terhadap kesehatan manusia, bahkan berujung pada kematian apabila tidak terdeteksi dan ditangani sejak dini. World Health Organization (WHO) mencatat bahwa penyakit liver berkontribusi signifikan terhadap angka kematian global, terutama yang disebabkan oleh hepatitis kronis, sirosis hati, dan kanker hati. Salah satu permasalahan utama dalam penanganan penyakit liver adalah sifatnya yang sering kali tidak menunjukkan gejala pada tahap awal, sehingga penyakit ini dikenal sebagai silent killer. Kondisi ini menyebabkan masyarakat cenderung terlambat melakukan pemeriksaan medis, sementara kerusakan pada organ hati telah berkembang ke tahap yang lebih parah. Selain itu, gaya hidup tidak sehat seperti konsumsi alkohol berlebihan, pola makan tidak seimbang, kurangnya aktivitas fisik, serta paparan zat berbahaya menjadi faktor risiko utama meningkatnya prevalensi penyakit liver. Tantangan lain yang dihadapi masyarakat adalah keterbatasan akses terhadap layanan kesehatan yang memadai serta ketergantungan pada diagnosis konvensional yang bersifat subjektif dan memerlukan waktu lama. Oleh karena itu, diperlukan suatu pendekatan berbasis teknologi informasi yang mampu membantu proses diagnosis secara lebih cepat, objektif, dan akurat dengan memanfaatkan data medis yang tersedia, sehingga upaya deteksi dini penyakit liver dapat ditingkatkan secara signifikan (Abrar et al., 2023; Habibie et al., n.d.)

Klasifikasi penyakit dalam dunia medis secara tradisional dilakukan melalui pendekatan konvensional yang mengandalkan pengalaman dan keahlian tenaga medis, hasil pemeriksaan

laboratorium, serta interpretasi klinis terhadap gejala pasien. Dalam konteks penyakit liver, proses klasifikasi biasanya didasarkan pada hasil tes fungsi hati seperti kadar bilirubin, enzim transaminase, alkaline phosphatase, albumin, dan rasio albumin-globulin. Data-data tersebut dianalisis secara manual oleh dokter untuk menentukan apakah pasien mengalami gangguan liver atau tidak. Meskipun pendekatan ini telah lama digunakan dan menjadi standar dalam praktik klinis, metode klasifikasi manual memiliki sejumlah keterbatasan. Proses analisis yang bergantung pada interpretasi manusia berpotensi menimbulkan subjektivitas, terutama ketika jumlah data pasien semakin besar dan kompleks. Selain itu, klasifikasi tanpa dukungan algoritma komputasi memerlukan waktu yang relatif lama dan sulit untuk mengidentifikasi pola tersembunyi yang mungkin tidak terlihat secara langsung. Dalam era digital saat ini, volume data medis terus meningkat dan menumpuk, sehingga metode klasifikasi konvensional menjadi kurang efektif untuk mendukung pengambilan keputusan yang cepat dan akurat. Kondisi ini menunjukkan bahwa pendekatan tradisional perlu dikombinasikan dengan metode berbasis data untuk meningkatkan efisiensi dan keandalan proses klasifikasi penyakit liver, khususnya dalam mendukung diagnosis awal dan pencegahan komplikasi yang lebih serius. (323508-Klasifikasi-Pasien-Penderita-Penyakit-Li-496b23e3, n.d.; Zulaikhah et al., 2022)

Penelitian berjudul "Optimasi Algoritma K-Nearest Neighbor (KNN) dengan Normalisasi dan Seleksi Fitur untuk Klasifikasi Penyakit Liver" dilakukan oleh Siti Zulaikhah HR, Abdul Aziz, dan Wahyudi Harianto pada tahun 2022. Penelitian ini membahas permasalahan rendahnya akurasi algoritma K-Nearest Neighbor dalam beberapa studi klasifikasi penyakit liver, yang disebabkan oleh ketidakseimbangan data, keberadaan atribut yang tidak relevan, serta perbedaan rentang nilai antar fitur. Untuk mengatasi permasalahan tersebut, peneliti menerapkan teknik normalisasi data dan seleksi fitur sebagai upaya optimasi model KNN. Metode seleksi fitur yang digunakan meliputi Information Gain, Gain Ratio, dan Symmetrical

Uncertainty, sedangkan normalisasi dilakukan dengan metode Min-Max. Dataset yang digunakan adalah Indian Liver Patient Dataset (ILPD) yang diambil dari UCI Machine Learning Repository. Hasil penelitian menunjukkan bahwa penerapan normalisasi dan seleksi fitur mampu meningkatkan kinerja algoritma KNN secara signifikan, dengan akurasi optimal diperoleh pada nilai K tertentu setelah proses optimasi dilakukan. Penelitian ini menegaskan bahwa tahap prapemrosesan data memiliki peran krusial dalam meningkatkan performa model klasifikasi berbasis KNN pada kasus penyakit liver (Aldana & Sasongko Wibowo, n.d.; Noviriandini et al., 2019).

Meskipun penelitian yang dilakukan oleh Siti Zulaikhah HR dkk. (2022) telah berhasil meningkatkan akurasi algoritma KNN melalui normalisasi dan seleksi fitur, masih terdapat celah penelitian (research gap) yang perlu dikaji lebih lanjut. Penelitian tersebut berfokus pada optimasi internal algoritma KNN tanpa melakukan perbandingan mendalam dengan pendekatan klasifikasi lain yang memiliki karakteristik berbeda, seperti Decision Tree. Selain itu, penelitian terdahulu lebih menekankan pada aspek peningkatan akurasi, namun belum secara komprehensif membahas interpretabilitas model yang sangat penting dalam konteks medis. Model KNN dikenal sebagai lazy learner yang sulit diinterpretasikan oleh tenaga medis karena tidak menghasilkan aturan keputusan yang eksplisit. Di sisi lain, kebutuhan akan sistem pendukung keputusan medis tidak hanya menuntut akurasi tinggi, tetapi juga transparansi dan kemudahan interpretasi hasil klasifikasi. Oleh karena itu, penelitian ini mengambil posisi berbeda dengan tidak hanya mengimplementasikan algoritma KNN sebagai model utama, tetapi juga mengkaji pendekatan klasifikasi lain dalam kerangka Knowledge Discovery in Database (KDD) untuk melihat peluang peningkatan kualitas model secara menyeluruh. Dengan demikian, penelitian ini diharapkan dapat mengisi kesenjangan antara performa akurasi dan kebutuhan interpretabilitas dalam klasifikasi penyakit liver berbasis data mining.

Sebagai upaya penyelesaian permasalahan klasifikasi penyakit liver, penelitian ini mengadopsi kerangka Knowledge Discovery in Database (KDD) yang terdiri dari tahapan seleksi data, prapemrosesan, transformasi data, data mining, dan evaluasi hasil. Dalam kerangka ini, algoritma Decision Tree digunakan sebagai pendekatan pembanding dan pendukung untuk memahami pola klasifikasi penyakit liver secara

lebih interpretatif. Decision Tree memiliki keunggulan dalam menghasilkan aturan keputusan yang mudah dipahami dan divisualisasikan dalam bentuk struktur pohon, sehingga sangat sesuai untuk aplikasi medis. Melalui tahapan KDD, data pasien akan diproses secara sistematis untuk menghilangkan noise, menangani missing value, serta menyeleksi atribut yang relevan sebelum dilakukan proses klasifikasi. Penggunaan Decision Tree dalam penelitian ini bertujuan untuk membantu mengelompokkan data pasien berdasarkan karakteristik klinis yang dominan, sekaligus memberikan wawasan tambahan terhadap hasil klasifikasi yang dihasilkan oleh KNN. Dengan demikian, integrasi KNN dan Decision Tree dalam kerangka KDD diharapkan mampu menghasilkan model klasifikasi yang tidak hanya akurat, tetapi juga informatif dan dapat dijadikan dasar pengambilan keputusan medis.

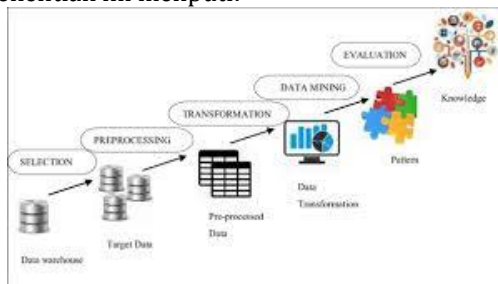
Tujuan dari penelitian ini adalah untuk mengimplementasikan algoritma K-Nearest Neighbor dalam meningkatkan model klasifikasi penyakit liver berbasis data medis, serta mengevaluasi kinerjanya melalui pendekatan sistematis dalam kerangka Knowledge Discovery in Database. Penelitian ini bertujuan untuk mengidentifikasi pola-pola penting dalam data pasien penyakit liver yang dapat digunakan sebagai dasar klasifikasi, sekaligus mengukur tingkat akurasi, presisi, recall, dan performa model secara keseluruhan. Selain itu, penelitian ini juga bertujuan untuk membandingkan hasil klasifikasi KNN dengan pendekatan lain yang lebih interpretatif, sehingga diperoleh pemahaman yang lebih komprehensif terhadap karakteristik data penyakit liver. Dengan tercapainya tujuan tersebut, diharapkan penelitian ini dapat memberikan kontribusi ilmiah dalam pengembangan model klasifikasi penyakit liver yang lebih andal dan aplikatif di bidang kesehatan.

Implikasi dari penelitian ini mencakup aspek akademik dan praktis. Secara akademik, penelitian ini diharapkan dapat memperkaya kajian ilmiah di bidang data mining dan machine learning, khususnya dalam penerapan algoritma KNN untuk klasifikasi penyakit liver. Hasil penelitian ini dapat menjadi referensi bagi peneliti selanjutnya dalam mengembangkan model klasifikasi yang lebih optimal dan sesuai dengan karakteristik data medis. Secara praktis, penelitian ini berpotensi memberikan manfaat bagi dunia kesehatan dengan menyediakan model pendukung keputusan yang dapat membantu tenaga medis dalam melakukan deteksi dini penyakit liver secara lebih cepat dan akurat. Dengan adanya sistem klasifikasi berbasis data mining, diharapkan kualitas layanan kesehatan dapat ditingkatkan, risiko kesalahan diagnosis

dapat diminimalkan, dan peluang kesembuhan pasien dapat ditingkatkan melalui penanganan yang lebih tepat waktu.

MATERIALS AND METHODS

Desain penelitian yang digunakan dalam penelitian ini mengacu pada kerangka Knowledge Discovery in Database (KDD). Kerangka KDD dipilih karena menyediakan tahapan yang sistematis dan terstruktur dalam pengolahan data, mulai dari pemilihan data hingga evaluasi hasil klasifikasi. Tahapan desain penelitian ini meliputi:



Gambar 1 Alur Penelitian

Berdasarkan Gambar 1 tentang Alur Penelitian menerangkan bahwa tahap pertama dalam desain penelitian :

1. **Seleksi Data** Pada tahap ini dilakukan pemilihan data yang relevan dengan tujuan penelitian, yaitu data pasien penyakit liver yang memiliki atribut klinis terkait fungsi hati.
2. **Pra-pemrosesan Data** Tahap ini meliputi pembersihan data, penanganan missing value, dan normalisasi data agar data memiliki kualitas yang baik dan siap digunakan dalam proses klasifikasi.
3. **Transformasi Data** Data yang telah diproses kemudian ditransformasikan ke dalam format yang sesuai untuk algoritma KNN, termasuk penyamaan skala nilai antar atribut.
4. **Data Mining** Pada tahap ini diterapkan algoritma K-Nearest Neighbor (KNN) untuk membangun model klasifikasi penyakit liver berdasarkan data latih.
5. **Evaluasi dan Interpretasi Model** yang dihasilkan dievaluasi menggunakan metrik evaluasi seperti confusion matrix, akurasi, presisi, dan recall untuk menilai kinerja model klasifikasi.

RESULTS AND DISCUSSION

Hasil Akurasi

Tabel 1 Hasil Akurasi

Metrik	Nilai
Akurasi Model KNN	0,6325 (63,25%)

Tabel 2 Confusion matrik

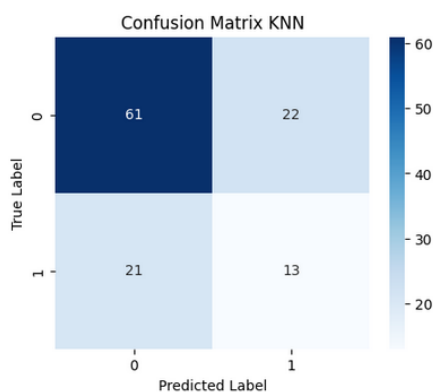
Kelas	Precision	Recall	F1-Score	Support
Kelas 1 (Liver)	0,74	0,73	0,74	83
Kelas 2 (Non-Liver)	0,37	0,38	0,38	34
Accuracy	-	-	0,63	117
Macro Average	0,56	0,56	0,56	117
Weighted Average	0,64	0,63	0,63	117

Berdasarkan tabel 1 dan 2 hasil evaluasi model K-Nearest Neighbor (KNN), diperoleh nilai akurasi sebesar 63,25%, yang menunjukkan bahwa model mampu mengklasifikasikan data uji dengan tingkat ketepatan yang cukup. Nilai akurasi ini memberikan gambaran umum mengenai performa model, namun dalam konteks klasifikasi medis, evaluasi model perlu dianalisis lebih mendalam menggunakan metrik lain seperti precision, recall, dan F1-score.

Pada kelas 1 (pasien penderita penyakit liver), model KNN menghasilkan nilai precision sebesar 0,74 dan recall sebesar 0,73, dengan F1-score 0,74. Nilai precision yang cukup tinggi menunjukkan bahwa sebagian besar prediksi pasien penderita liver yang dihasilkan oleh model memang benar-benar termasuk dalam kelas tersebut. Sementara itu, nilai recall yang juga relatif tinggi menunjukkan bahwa model mampu mendeteksi sebagian besar pasien yang benar-benar menderita penyakit liver. Hal ini menunjukkan bahwa model KNN cukup baik dalam mengenali kelas penderita penyakit liver, yang merupakan aspek penting dalam aplikasi medis.

Sebaliknya, pada kelas 2 (pasien non-liver), nilai precision dan recall relatif rendah, masing-masing sebesar 0,37 dan 0,38, dengan F1-score 0,38. Nilai ini menunjukkan bahwa model kurang optimal dalam mengklasifikasikan pasien yang tidak menderita penyakit liver. Rendahnya performa pada kelas ini dapat disebabkan oleh ketidakseimbangan jumlah data antar kelas, di mana jumlah data pasien non-liver lebih sedikit dibandingkan pasien penderita penyakit liver. Kondisi ini menyebabkan model cenderung bias terhadap kelas mayoritas.

Nilai macro average yang berada pada kisaran 0,56 menunjukkan bahwa jika setiap kelas dianggap memiliki bobot yang sama, maka performa model masih tergolong sedang. Sementara itu, nilai weighted average yang lebih tinggi, yaitu sekitar 0,63, menunjukkan bahwa performa model lebih dipengaruhi oleh kelas mayoritas. Hal ini menguatkan indikasi adanya ketidakseimbangan kelas dalam dataset yang digunakan. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa algoritma KNN dengan parameter $k = 5$ mampu melakukan klasifikasi penyakit liver dengan performa yang cukup baik pada kelas penderita penyakit liver, namun masih memiliki keterbatasan dalam mengklasifikasikan pasien non-liver. Oleh karena itu, diperlukan pengembangan lebih lanjut, seperti optimasi nilai k , penanganan ketidakseimbangan kelas, atau penggunaan metode tambahan untuk meningkatkan kinerja model klasifikasi penyakit liver.



Gambar 2 Confusion Matrix

Gambar 2 tersebut menampilkan Confusion Matrix dari hasil klasifikasi penyakit liver menggunakan algoritma K-Nearest Neighbor (KNN). Confusion matrix digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap label sebenarnya pada data uji. Matriks ini terdiri dari empat komponen utama, yaitu *True Positive*, *True Negative*, *False Positive*, dan *False Negative*, yang masing-masing memberikan informasi detail mengenai kesalahan dan keberhasilan klasifikasi.

Berdasarkan confusion matrix yang ditampilkan, dapat dijelaskan bahwa nilai 61 pada posisi *True Negative (TN)* menunjukkan jumlah data pasien yang sebenarnya tidak menderita penyakit liver dan berhasil diprediksi dengan benar sebagai non-liver oleh model. Nilai ini mencerminkan kemampuan model dalam mengenali pasien sehat dengan cukup baik.

Nilai 13 pada posisi *True Positive (TP)* menunjukkan jumlah data pasien yang benar-

benar menderita penyakit liver dan berhasil diprediksi dengan benar oleh model sebagai penderita liver. Nilai ini menunjukkan bahwa model mampu mendeteksi sebagian pasien yang benar-benar menderita penyakit liver, meskipun jumlahnya masih relatif terbatas dibandingkan total pasien penderita liver.

Nilai 22 pada posisi *False Positive (FP)* menunjukkan jumlah data pasien yang sebenarnya tidak menderita penyakit liver, namun diprediksi oleh model sebagai penderita penyakit liver. Kesalahan jenis ini menunjukkan adanya kecenderungan model untuk memberikan prediksi positif pada pasien yang sebenarnya sehat, yang dapat menyebabkan alarm palsu dalam konteks medis. Sementara itu, nilai 21 pada posisi *False Negative (FN)* menunjukkan jumlah data pasien yang sebenarnya menderita penyakit liver, tetapi diprediksi oleh model sebagai non-liver. Kesalahan ini tergolong kritis dalam konteks medis karena dapat menyebabkan pasien yang sakit tidak terdeteksi oleh sistem klasifikasi, sehingga berpotensi menunda penanganan medis yang diperlukan. Secara keseluruhan, confusion matrix ini menunjukkan bahwa model KNN memiliki kemampuan klasifikasi yang cukup, namun masih menghasilkan jumlah kesalahan yang cukup signifikan, khususnya pada kesalahan *false negative* dan *false positive*. Hal ini sejalan dengan nilai akurasi dan metrik evaluasi lainnya yang menunjukkan bahwa performa model masih perlu ditingkatkan. Oleh karena itu, diperlukan upaya lanjutan seperti optimasi parameter K , penanganan ketidakseimbangan kelas, atau penggunaan metode klasifikasi lain untuk meningkatkan kemampuan model dalam mendeteksi penyakit liver secara lebih akurat.

Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Nearest Neighbor (KNN) dalam melakukan klasifikasi penyakit liver berdasarkan data medis pasien serta mengevaluasi kinerja model yang dihasilkan. Berdasarkan hasil yang diperoleh pada tahap eksperimen, dapat dilakukan pembahasan secara menyeluruh terhadap setiap tahapan penelitian, mulai dari karakteristik data, proses prapemrosesan, hingga hasil evaluasi model klasifikasi.

Berdasarkan tahap seleksi dan prapemrosesan data, dataset yang digunakan berasal dari Indian Liver Patient Dataset (ILPD) dengan total 583 data pasien dan 11 atribut. Atribut-atribut yang digunakan mencerminkan parameter klinis yang relevan dengan fungsi hati, seperti kadar bilirubin, enzim transaminase, alkaline phosphatase, albumin, serta rasio albumin dan globulin. Proses prapemrosesan menunjukkan bahwa sebagian

besar atribut tidak memiliki missing value, kecuali pada atribut Albumin and Globulin Ratio yang memiliki sejumlah kecil data kosong. Missing value tersebut ditangani menggunakan metode imputasi median agar tidak mengurangi jumlah data secara signifikan dan tetap mempertahankan karakteristik distribusi data. Selain itu, atribut kategorikal Gender berhasil diubah menjadi bentuk numerik melalui proses encoding sehingga dapat diproses oleh algoritma KNN. Tahapan normalisasi data menggunakan StandardScaler dilakukan untuk menyamakan skala antar atribut, mengingat algoritma KNN sangat sensitif terhadap perbedaan rentang nilai antar fitur.

Hasil Exploratory Data Analysis (EDA) menunjukkan bahwa dataset memiliki ketidakseimbangan kelas, di mana jumlah pasien penderita penyakit liver lebih banyak dibandingkan pasien non-liver. Kondisi ini berdampak langsung terhadap performa model klasifikasi, karena algoritma cenderung lebih akurat dalam memprediksi kelas mayoritas. Selain itu, EDA juga menunjukkan adanya outlier pada beberapa atribut enzim hati, seperti Alkaline Phosphatase, Alamine Aminotransferase, dan Aspartate Aminotransferase. Keberadaan outlier ini secara klinis masih relevan karena mencerminkan kondisi pasien dengan gangguan liver berat, sehingga tidak dihilangkan dalam penelitian ini. Distribusi data yang tidak normal serta korelasi yang cukup tinggi antar beberapa fitur juga menjadi karakteristik penting yang memengaruhi kinerja model.

Pada tahap pemodelan, algoritma KNN diterapkan dengan nilai parameter $k = 5$ dan menggunakan metode pengukuran jarak Euclidean Distance. Pemilihan nilai k ini didasarkan pada nilai umum yang sering digunakan sebagai titik awal dalam klasifikasi KNN. Hasil pengujian model menunjukkan bahwa algoritma KNN berhasil dilatih dan menghasilkan nilai akurasi sebesar 63,25%. Nilai akurasi ini menunjukkan bahwa model mampu mengklasifikasikan data uji dengan tingkat ketepatan yang cukup, meskipun belum optimal. Hasil ini sejalan dengan beberapa penelitian terdahulu yang menyatakan bahwa KNN tanpa optimasi lanjutan umumnya menghasilkan akurasi pada kisaran sedang dalam klasifikasi penyakit liver.

Analisis lebih lanjut menggunakan classification report menunjukkan bahwa performa model tidak merata pada setiap kelas. Pada kelas pasien penderita penyakit liver, model menunjukkan nilai precision sebesar 0,74

dan recall sebesar 0,73, yang mengindikasikan bahwa model cukup baik dalam mendeteksi pasien yang benar-benar menderita penyakit liver. Hal ini menjadi poin positif karena dalam konteks medis, kemampuan mendeteksi pasien sakit (recall tinggi) sangat penting untuk mendukung deteksi dini. Namun, pada kelas pasien non-liver, nilai precision dan recall relatif rendah, masing-masing sebesar 0,37 dan 0,38. Hal ini menunjukkan bahwa model masih kesulitan dalam mengenali pasien yang tidak menderita penyakit liver secara akurat. Kondisi ini dipengaruhi oleh ketidakseimbangan kelas serta karakteristik data yang kompleks.

Hasil confusion matrix memperkuat temuan tersebut. Model mampu mengklasifikasikan sejumlah besar pasien non-liver secara benar, namun masih menghasilkan cukup banyak kesalahan false positive dan false negative. Kesalahan false negative, yaitu pasien penderita liver yang diprediksi sebagai non-liver, menjadi perhatian penting karena berpotensi menimbulkan risiko klinis apabila model digunakan sebagai sistem pendukung keputusan. Oleh karena itu, meskipun model KNN menunjukkan performa yang cukup, hasil ini mengindikasikan bahwa masih diperlukan pengembangan dan optimasi lebih lanjut untuk meningkatkan keandalan model.

Jika dibandingkan dengan penelitian terdahulu, hasil yang diperoleh dalam penelitian ini masih berada dalam rentang yang wajar. Beberapa penelitian yang menerapkan optimasi parameter K , seleksi fitur, atau metode metaheuristik seperti Particle Swarm Optimization mampu meningkatkan akurasi KNN hingga di atas 70%. Namun, penelitian ini secara sengaja membatasi penggunaan metode optimasi lanjutan untuk menilai performa dasar algoritma KNN dalam kerangka Knowledge Discovery in Database (KDD). Dengan demikian, hasil penelitian ini memberikan gambaran objektif mengenai kemampuan KNN tanpa optimasi kompleks, sekaligus menjadi dasar untuk penelitian lanjutan. Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma K-Nearest Neighbor mampu digunakan untuk klasifikasi penyakit liver dengan tingkat kinerja yang cukup baik, khususnya dalam mendeteksi pasien penderita penyakit liver. Namun, keterbatasan masih terlihat pada kemampuan model dalam mengklasifikasikan kelas minoritas. Oleh karena itu, penelitian ini menegaskan pentingnya proses prapemrosesan data, pemilihan parameter yang tepat, serta evaluasi model yang komprehensif dalam membangun sistem klasifikasi penyakit liver berbasis data mining. Hasil ini juga membuka peluang bagi penelitian selanjutnya untuk melakukan optimasi model atau mengombinasikan

KNN dengan pendekatan lain guna meningkatkan akurasi dan reliabilitas klasifikasi penyakit liver.

CONCLUSION

Berdasarkan hasil penelitian yang telah dilakukan mengenai implementasi algoritma K-Nearest Neighbor (KNN) untuk klasifikasi penyakit liver, dapat disimpulkan beberapa hal penting. Pertama, algoritma KNN berhasil diterapkan dalam mengklasifikasikan penyakit liver menggunakan data medis pasien yang diperoleh dari Indian Liver Patient Dataset (ILPD). Seluruh tahapan penelitian, mulai dari seleksi data, prapemrosesan, exploratory data analysis, hingga pemodelan dan evaluasi, telah dilaksanakan secara sistematis dengan mengacu pada kerangka Knowledge Discovery in Database (KDD).

Kedua, proses prapemrosesan data berperan penting dalam meningkatkan kesiapan dataset untuk proses klasifikasi. Penanganan missing value menggunakan metode imputasi median, encoding data kategorikal, serta normalisasi data berhasil menghasilkan dataset yang bersih dan siap digunakan oleh algoritma KNN. Hasil Exploratory Data Analysis menunjukkan bahwa dataset memiliki karakteristik berupa ketidakseimbangan kelas, distribusi fitur yang tidak normal, serta keberadaan outlier pada beberapa atribut medis, yang secara tidak langsung memengaruhi kinerja model klasifikasi.

Ketiga, hasil pengujian model KNN dengan nilai parameter $k = 5$ menunjukkan bahwa model mampu mengklasifikasikan penyakit liver dengan nilai akurasi sebesar 63,25%. Berdasarkan evaluasi menggunakan classification report dan confusion matrix, model menunjukkan performa yang cukup baik dalam mendeteksi pasien penderita penyakit liver, ditunjukkan oleh nilai precision dan recall yang relatif tinggi pada kelas tersebut. Namun demikian, performa model pada kelas pasien non-liver masih tergolong rendah, yang dipengaruhi oleh ketidakseimbangan kelas dan karakteristik data yang kompleks.

Keempat, hasil penelitian ini menunjukkan bahwa algoritma KNN dapat digunakan sebagai metode dasar dalam klasifikasi penyakit liver, namun masih memiliki keterbatasan dalam menghasilkan performa yang optimal. Oleh karena itu, penelitian ini memberikan gambaran bahwa KNN tanpa optimasi lanjutan mampu menghasilkan performa yang cukup, sekaligus menjadi dasar untuk pengembangan dan penelitian lanjutan di

bidang klasifikasi penyakit liver berbasis data mining.

REFERENCE

- Abrar, I. N., Abdullah, A., & Sucipto, S. (2023). Liver Disease Classification Using the Elbow Method to Determine Optimal K in the K-Nearest Neighbor (K-NN) Algorithm. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 12(2), 218–228.
<https://doi.org/10.32736/sisfokom.v12i2.1643>
- Aldana, S., & Sasongko Wibowo, J. (n.d.). *Penerapan Data Mining Terhadap Klasifikasi Pasien Penderita Penyakit Liver Menggunakan Metode K-Nearest Neighbor*.
- Amelia, A. F. (2025). Association Rule Analysis for Sales Strategy Optimization. *Sistemasi Journal*.
- Anhari Arifin, M., Rezki Hamdani, M., Millano, W., & Efrizoni, L. (2024). PERBANDINGAN ALGORITMA K-NN DAN RANDOM FOREST DALAM MEMREDIKSI PENYAKIT LIVER STMIK AMIK RIAU ABSTRAK. In *Jurnal Kajian Ilmiah Multidisipliner* (Vol. 8, Issue 4).
<https://www.kaggle.com/datasets/jeevannagaraj/indian-liver->
- Aqliyah, Z. H. (2025). FP-Growth Algorithm for Association Model Optimization. *Journal of Advanced Informatics and Engineering Applications*.
- Audiana, W. (2025). Application of Market Basket Analysis for Sales Transaction Optimization. *CSRID Journal*.
- Habibie, E. Y., Oktavianto, H., & A'yun, Q. (n.d.). *PENERAPAN ALOGARITMA K-NEAREST NEIGHBOR UNTUK KLASIFIKASI PENYAKIT LIVER*.
- Novriandini, A., Handayani, P., Bsi, U., & Nusa Mandiri, S. (2019). Prediksi Penyakit Liver Dengan Menggunakan Metode Naïve Bayes Dan K-Nearest Neighbour (KNN). *Seminar Nasional Rekayasa Dan Teknologi*, 27.
<http://archive.ics.uci.edu/ml/>.
- Prindo, B., Putro, S., Handayani, Y., & Hidayat, T. (n.d.). *Optimasi Algoritma K-Nearest Neighbor dengan Particle Swarm Optimization untuk Klasifikasi Penyakit Liver*.
- Zulaikhah, S. H., Aziz, A., & Harianto, W. (2022). OPTIMASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DENGAN NORMALISASI DAN SELEKSI FITUR UNTUK KLASIFIKASI PENYAKIT LIVER. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 6, Issue 2).
<https://archive.ics.uci.edu/ml/index.php>