

Analisis Kinerja Algoritma K-Nearest Neighbor dalam Klasifikasi Penyakit Liver

Ikhwan Fauzi^{1*}, Cendra Harmon²

¹Sistem Informasi, Universitas Pamulang, Jl. Raya Puspitek, Tangerang Selatan 15310, Indonesia

²Teknik Informatika, Universitas Pamulang, Jl. Raya Puspitek, Tangerang Selatan 15310, Indonesia
e-mail: dosen03150@unpam.ac.id* ; dosen02677@unpam.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis kinerja algoritma K-Nearest Neighbor (KNN) dalam klasifikasi penyakit liver guna mendukung deteksi dini dan mengurangi risiko komplikasi akibat keterlambatan diagnosis. Data yang digunakan berasal dari *Indian Liver Patient Dataset* (ILPD) yang memuat berbagai variabel klinis, seperti bilirubin, albumin, dan enzim hati. Tahapan penelitian meliputi pra-pemrosesan data, penerapan algoritma KNN, serta evaluasi kinerja menggunakan metrik klasifikasi. Pengujian dilakukan dengan memvariasikan nilai parameter k dari 1 hingga 20 untuk memperoleh performa terbaik. Hasil penelitian menunjukkan bahwa nilai $k = 15$ menghasilkan akurasi tertinggi sebesar 76,07%, meningkat dibandingkan pengujian awal yang mencapai 67%. Temuan ini menunjukkan bahwa algoritma KNN mampu mengklasifikasikan data pasien penyakit liver dengan cukup baik, namun kinerjanya sangat dipengaruhi oleh pemilihan nilai k . Oleh karena itu, optimasi parameter dan pengembangan model lebih lanjut masih diperlukan untuk meminimalkan kesalahan klasifikasi.

Kata kunci: penyakit liver, K-Nearest Neighbor, klasifikasi, machine learning, ILPD.

Abstract

This study aims to analyze the performance of the K-Nearest Neighbor (KNN) algorithm in classifying liver disease to support early detection and reduce the risk of complications caused by delayed diagnosis. The data used in this study were obtained from the Indian Liver Patient Dataset (ILPD), which contains various clinical variables such as bilirubin, albumin, and liver enzymes. The research stages include data preprocessing, implementation of the KNN algorithm, and performance evaluation using classification metrics. Experiments were conducted by varying the parameter k from 1 to 20 to obtain the optimal performance. The results show that $k = 15$ achieved the highest accuracy of 76.07%, representing an improvement compared to the initial testing accuracy of 67%. These findings indicate that the KNN algorithm is capable of classifying liver disease patient data effectively; however, its performance is highly dependent on the selection of the k value. Therefore, further parameter optimization and model development are required to minimize classification errors.

Keywords: liver disease, K-Nearest Neighbor, classification, machine learning, ILPD.

1. PENDAHULUAN

Penyakit liver merupakan salah satu gangguan kesehatan yang masih menjadi perhatian serius karena memiliki angka kejadian dan tingkat kematian yang relatif tinggi di berbagai wilayah. Penurunan fungsi hati dapat disebabkan oleh berbagai faktor, seperti infeksi virus, konsumsi alkohol jangka panjang, gangguan metabolik, serta pola hidup yang tidak sehat. Apabila tidak terdeteksi dan ditangani sejak dini, penyakit liver berpotensi berkembang menjadi komplikasi serius, seperti sirosis dan kanker hati. Namun, proses diagnosis secara klinis umumnya memerlukan pemeriksaan laboratorium yang kompleks, biaya yang tidak sedikit, serta waktu dan ketelitian yang tinggi, sehingga berisiko menyebabkan keterlambatan penanganan [1][2].

Perkembangan teknologi informasi dan pengolahan data membuka peluang baru dalam membantu proses diagnosis penyakit melalui penerapan teknik machine learning. Data klinis pasien yang tersimpan dalam rekam medis dapat dianalisis menggunakan algoritma klasifikasi untuk memprediksi kondisi kesehatan seseorang, termasuk risiko penyakit liver. Salah satu algoritma klasifikasi yang banyak digunakan adalah K-Nearest Neighbor (KNN). Algoritma ini menentukan kelas suatu data berdasarkan tingkat kemiripan atau jarak terdekat dengan data lain, sehingga relatif mudah diterapkan dan efektif pada berbagai permasalahan klasifikasi [3][4].

Sejumlah penelitian sebelumnya telah menerapkan algoritma KNN untuk klasifikasi penyakit liver dengan tingkat akurasi yang beragam. Namun, sebagian penelitian masih menunjukkan keterbatasan, khususnya dalam pemilihan nilai parameter k yang optimal, yang sangat berpengaruh terhadap stabilitas dan akurasi hasil klasifikasi. Oleh karena itu, diperlukan kajian lebih lanjut untuk menganalisis pengaruh variasi nilai k terhadap kinerja algoritma KNN agar diperoleh hasil klasifikasi yang lebih optimal dan andal.

Penelitian ini menggunakan data pasien penyakit liver sebanyak 583 sampel yang bersumber dari Indian Liver Patient Dataset (ILPD), dengan berbagai variabel klinis seperti usia, jenis kelamin, kadar bilirubin, enzim hati, total protein, albumin, serta rasio albumin dan globulin. Setiap data telah dilengkapi dengan label kelas yang menunjukkan status pasien sebagai penderita atau bukan penderita penyakit liver, sehingga sesuai digunakan untuk pengujian kinerja algoritma klasifikasi.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengkaji performa algoritma K-Nearest Neighbor dalam mengklasifikasikan penyakit liver berdasarkan data klinis pasien. Evaluasi kinerja dilakukan menggunakan beberapa metrik, yaitu akurasi, presisi, recall, dan confusion matrix. Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai efektivitas algoritma KNN sebagai alat bantu deteksi dini penyakit liver serta mendukung pengembangan sistem pendukung keputusan di bidang medis [5].

2. METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif dengan desain eksperimen untuk mengevaluasi performa algoritma K-Nearest Neighbor (KNN) dalam klasifikasi penyakit liver. Data yang digunakan bersumber dari *Indian Liver Patient Dataset (ILPD)* yang memuat berbagai variabel klinis, meliputi usia, jenis kelamin, kadar bilirubin total dan langsung, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, total protein, albumin, serta rasio albumin terhadap globulin.[6]

Tabel 1. Data Indian Liver Patient Dataset

No	1	2	3	4	5
Age	65	62	62	58	72
Gender	Female	Male	Male	Male	Male
Total_Bilirubin	0.7	10.9	7.3	1.0	3.9
Direct_Bilirubin	0.1	5.5	4.1	0.4	2.0
Alkaline_Phosphotase	187	699	490	182	195
Alanine_Aminotransferase	16	64	60	14	27
Aspartate_Aminotransferase	18	100	68	20	59
Total_Proteins	6.8	7.5	7.0	6.8	7.3
Albumin	3.3	3.2	3.3	3.4	2.4
Albumin_and_Globulin_Ratio	0.90	0.74	0.89	1.00	0.40
Dataset	1	1	1	1	1

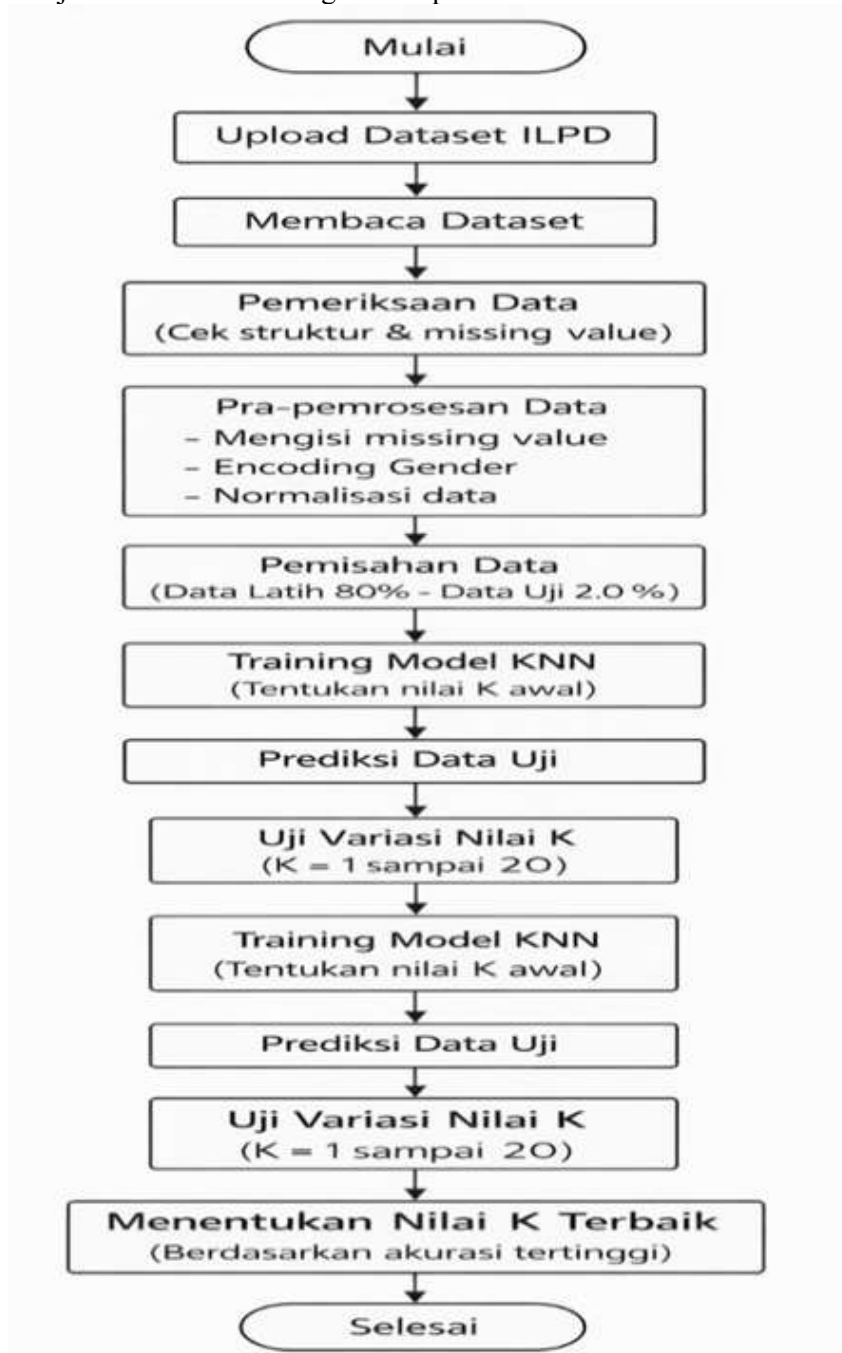
Dataset tersebut telah dilengkapi dengan label kelas yang menunjukkan kondisi setiap pasien, yaitu penderita dan bukan penderita penyakit liver. Tabel 1 menyajikan contoh lima data awal yang menggambarkan variasi karakteristik pasien berdasarkan parameter klinis. Seluruh contoh data tersebut dikategorikan sebagai penderita penyakit liver, yang menunjukkan perbedaan nilai laboratorium antar pasien meskipun berada dalam kelas yang sama.[7]

Tahapan penelitian diawali dengan proses pra-pemrosesan data. Pada tahap ini dilakukan pemeriksaan dan penanganan data yang tidak lengkap, konversi atribut kategorikal seperti jenis kelamin ke dalam bentuk numerik, serta normalisasi data numerik agar setiap variabel memiliki skala yang seimbang. Proses normalisasi diperlukan untuk mencegah dominasi atribut tertentu dalam perhitungan jarak pada algoritma KNN. Setelah pra-pemrosesan selesai, dataset dibagi menjadi data pelatihan dan data pengujian dengan proporsi 70% sebagai data latih dan 30% sebagai data uji. Pembagian data ini bertujuan untuk melatih model sekaligus menguji kemampuannya dalam mengklasifikasikan data yang belum pernah digunakan sebelumnya.[8]

Selanjutnya, algoritma KNN diterapkan pada data pelatihan dengan menentukan nilai parameter k sebagai jumlah tetangga terdekat yang digunakan dalam proses klasifikasi. Perhitungan jarak antar data dilakukan

menggunakan jarak Euclidean karena sesuai untuk data numerik yang telah dinormalisasi. Data pengujian kemudian diklasifikasikan ke dalam kelas penderita atau bukan penderita penyakit liver berdasarkan kelas mayoritas dari k tetangga terdekat.[9]

Tahap akhir penelitian difokuskan pada evaluasi kinerja model KNN. Evaluasi dilakukan menggunakan beberapa metrik performa, yaitu akurasi, presisi, recall, dan *confusion matrix*. [10] Hasil evaluasi ini digunakan untuk menilai kemampuan algoritma KNN dalam mengklasifikasikan penyakit liver secara tepat berdasarkan data klinis pasien. Seluruh proses pengolahan data, pelatihan model, dan evaluasi dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka *pandas*, *NumPy*, dan *scikit-learn*. Alur metodologi penelitian secara keseluruhan disajikan dalam bentuk diagram alir pada Gambar 1.



Gambar 1. Flowchart Metodologi Penelitian

3. HASIL DAN PEMBAHASAN

Penelitian dimulai dengan penetapan tujuan utama, yaitu mengkaji performa algoritma K-Nearest Neighbor (KNN) dalam mengelompokkan data pasien penyakit liver dengan memanfaatkan Indian Liver Patient Dataset (ILPD). Pada tahap awal ini juga ditentukan batasan penelitian, variabel yang dianalisis, serta teknik evaluasi yang digunakan untuk menilai hasil klasifikasi.

3.1. Pemeriksaan Dataset

Tahap pemeriksaan dataset dilakukan untuk mengetahui kondisi awal data sebelum digunakan dalam proses pembentukan model. Kegiatan ini mencakup pengecekan keberadaan nilai kosong (*missing value*) pada setiap atribut yang terdapat dalam dataset pasien penyakit liver.

Tabel 2. Hasil Pemeriksaan Missing Value pada Dataset

No	Nama Atribut	Jumlah Missing Value
1	Age	0
2	Gender	0
3	Total_Bilirubin	0
4	Direct_Bilirubin	0
5	Alkaline_Phosphotase	0
6	Alanine_Aminotransferase	0
7	Aspartate_Aminotransferase	0
8	Total_Proteins	0
9	Albumin	0
10	Albumin_and_Globulin_Ratio	4
11	Dataset	0

Tahap pemeriksaan dataset dilakukan untuk mengetahui kondisi awal data sebelum digunakan dalam proses pembentukan model. Kegiatan ini mencakup pengecekan keberadaan nilai kosong (*missing value*) pada setiap atribut yang terdapat dalam dataset pasien penyakit liver.

Berdasarkan hasil pemeriksaan, diketahui bahwa hampir seluruh atribut memiliki data yang lengkap. Dari sebelas atribut yang dianalisis, sepuluh atribut tidak mengandung nilai kosong, yaitu *Age*, *Gender*, *Total_Bilirubin*, *Direct_Bilirubin*, *Alkaline_Phosphotase*, *Alanine_Aminotransferase*, *Aspartate_Aminotransferase*, *Total_Proteins*, *Albumin*, serta atribut kelas *Dataset*. Kondisi ini menunjukkan bahwa sebagian besar data memiliki kualitas yang baik dan dapat langsung dimanfaatkan dalam proses pemodelan.

Atribut *Albumin_and_Globulin_Ratio* menjadi satu-satunya variabel yang mengandung data tidak lengkap dengan jumlah empat nilai kosong. Jumlah tersebut relatif kecil jika dibandingkan dengan keseluruhan jumlah data, sehingga tidak memengaruhi distribusi data secara signifikan. Meskipun demikian, keberadaan nilai kosong tetap berpotensi mengganggu proses perhitungan jarak pada algoritma KNN apabila tidak ditangani dengan benar.

Secara umum, hasil pemeriksaan menunjukkan bahwa dataset memiliki tingkat kelengkapan yang tinggi dan hanya membutuhkan perlakuan minimal pada tahap pra-pemrosesan. Dengan melakukan pengisian nilai kosong pada atribut *Albumin_and_Globulin_Ratio* menggunakan metode imputasi yang sesuai, seperti nilai rata-rata atau median, dataset dapat digunakan secara optimal dalam proses pelatihan dan pengujian model klasifikasi.

3.2 Pra-pemrosesan Data

Tahap pra-pemrosesan merupakan bagian penting karena kualitas data sangat menentukan keberhasilan algoritma KNN. Pada tahap ini dilakukan serangkaian langkah untuk menyiapkan data agar siap digunakan dalam proses pelatihan model. Langkah pertama adalah menangani data yang tidak lengkap. Berdasarkan hasil pemeriksaan, hanya atribut *Albumin_and_Globulin_Ratio* yang memiliki beberapa nilai kosong. Nilai tersebut kemudian dilengkapi melalui teknik imputasi dengan mengganti nilai yang hilang menggunakan nilai statistik tertentu, seperti rata-rata atau median.

Selanjutnya dilakukan perubahan data kategorikal, khususnya pada atribut *Gender*, ke dalam bentuk numerik. Proses ini bertujuan agar atribut tersebut dapat diproses oleh algoritma KNN. Pengkodean dilakukan dengan memberikan nilai numerik pada masing-masing kategori, misalnya 1

untuk laki-laki dan 0 untuk perempuan.

Tahap berikutnya adalah melakukan normalisasi terhadap seluruh atribut numerik. Normalisasi diperlukan karena setiap variabel memiliki rentang nilai yang berbeda, sehingga berpotensi menimbulkan dominasi atribut tertentu dalam perhitungan jarak. Oleh karena itu, digunakan metode normalisasi seperti *Min-Max Scaling* atau *Standardization* agar seluruh variabel berada pada skala yang seimbang.

Melalui tahapan pra-pemrosesan ini, data menjadi lebih bersih, seragam, dan terstandarisasi. Dengan kondisi data yang lebih baik, proses pelatihan dan pengujian model KNN diharapkan dapat menghasilkan performa klasifikasi yang lebih stabil dan akurat.

3.3. Pemisahan Data Latih dan Data Uji

Setelah data dipersiapkan, langkah selanjutnya adalah membagi data set menjadi dua kelompok, yaitu data pelatihan dan data pengujian. Pemisahan ini bertujuan agar model dapat dibangun menggunakan sebagian data sekaligus diuji kemampuannya dalam melakukan prediksi pada data yang belum pernah digunakan sebelumnya.

Dalam penelitian ini, dataset dibagi dengan proporsi 80% sebagai data latih dan 20% sebagai data uji. Data latih digunakan untuk membentuk model KNN dengan mempelajari pola hubungan antara variabel klinis dan status penyakit liver, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam melakukan generalisasi. Pembagian data dilakukan secara acak agar distribusi kedua kelompok tetap mewakili karakteristik dataset secara keseluruhan.

Tahap pemisahan data ini sangat penting karena berpengaruh langsung terhadap hasil evaluasi. Tanpa pemisahan, model berisiko mengalami *overfitting*, yaitu hanya mampu mengenali data pelatihan tetapi gagal memberikan prediksi yang baik pada data baru. Dengan adanya data uji yang terpisah, kinerja model dapat dinilai secara lebih objektif.

3.4. Hasil Klasifikasi Awal Menggunakan KNN

Setelah pembagian data selesai, proses klasifikasi awal dilakukan menggunakan algoritma KNN dengan nilai parameter k awal yang telah ditentukan. Model dilatih menggunakan data latih untuk mempelajari pola hubungan antara atribut klinis dan kelas penyakit liver, kemudian digunakan untuk memprediksi kelas pada data uji.

Tabel 3. Hasil Classification Report Model KNN

Kelas	Precision	Recall	F1-Score	Support
Non-Penderita (0)	0.38	0.50	0.43	30
Penderita (1)	0.81	0.72	0.76	87
Accuracy	–	–	0.67	117
Macro Average	0.60	0.61	0.60	117
Weighted Average	0.70	0.67	0.68	117

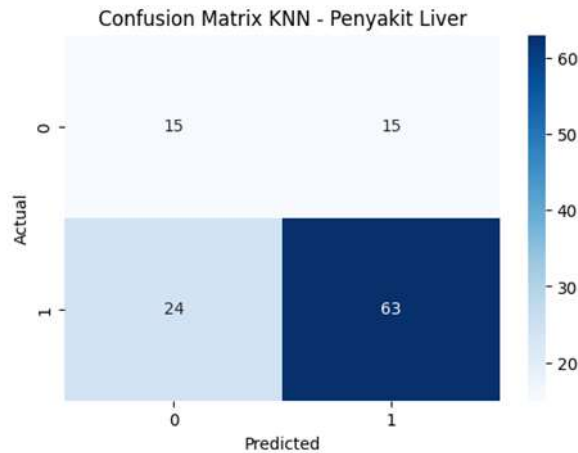
Hasil evaluasi menunjukkan bahwa performa model berbeda antara kedua kelas. Pada kelas non-penderita, nilai presisi, recall, dan f1-score masih tergolong rendah, yang menunjukkan bahwa model belum mampu membedakan pasien sehat secara optimal. Sebaliknya, pada kelas penderita, nilai presisi dan recall lebih tinggi, yang menandakan bahwa model cukup baik dalam mengenali pasien yang benar-benar menderita penyakit liver.

Nilai akurasi keseluruhan sebesar 0,67 menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar. Namun, ketidak seimbangan jumlah data antar kelas menyebabkan model cenderung lebih mempelajari pola dari kelas mayoritas. Nilai *macro average* dan *weighted average* menunjukkan bahwa performa model masih berada pada tingkat sedang dan masih memerlukan peningkatan, terutama pada kelas non-penderita.

3.5. Analisis Confusion Matrix

Confusion matrix digunakan untuk mengevaluasi kesalahan dan keberhasilan prediksi pada masing-masing kelas. Berdasarkan hasil analisis, pada kelas non-penderita terdapat 15 data yang berhasil diklasifikasikan dengan benar dan 15 data yang salah diklasifikasikan sebagai penderita. Hal ini menunjukkan

tingkat kesalahan yang cukup tinggi pada kelas tersebut.



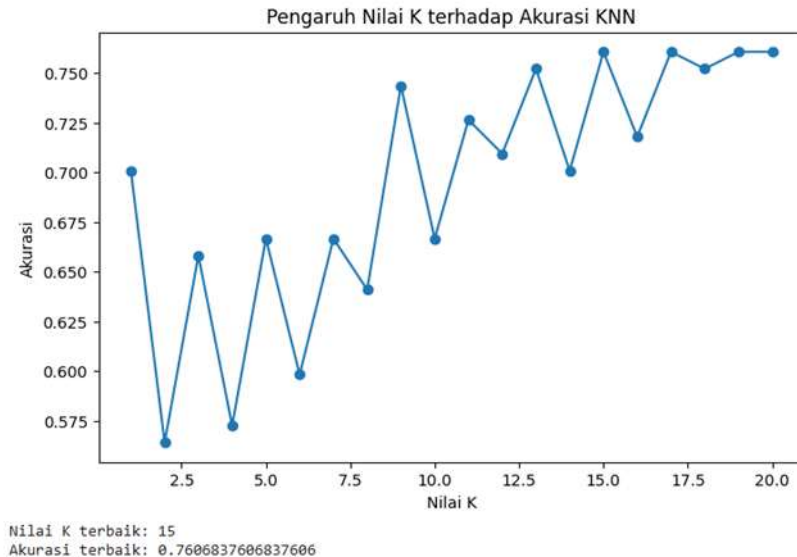
Gambar 2. Confusion Matrix Hasil Klasifikasi KNN

Pada kelas penderita, sebanyak 63 data berhasil dikenali dengan benar, sedangkan 24 data penderita salah diklasifikasikan sebagai non-penderita. Kesalahan jenis ini sangat penting diperhatikan karena berpotensi menyebabkan pasien sakit tidak terdeteksi oleh sistem.

Secara keseluruhan, hasil confusion matrix menunjukkan bahwa model lebih baik dalam mengenali pasien penderita dibandingkan pasien non-penderita. Meskipun demikian, tingginya nilai *false positive* dan *false negative* menunjukkan bahwa performa model masih belum optimal dan memerlukan perbaikan lebih lanjut.

3.6. Pengujian Variasi Nilai Parameter K

Pengujian variasi nilai k dilakukan untuk mengetahui pengaruh jumlah tetangga terdekat terhadap kinerja model KNN. Nilai k diuji pada rentang 1 hingga 20 untuk memperoleh parameter yang paling optimal.



Gambar 3. Grafik Pengaruh Nilai K terhadap Akurasi KNN

Berdasarkan grafik pengaruh nilai parameter k terhadap akurasi algoritma K-Nearest Neighbor (KNN), terlihat bahwa nilai k memiliki pengaruh yang signifikan terhadap performa klasifikasi penyakit liver. Pada nilai k yang kecil, khususnya pada $k = 1$ hingga $k = 4$, akurasi model menunjukkan fluktuasi yang cukup besar dan cenderung rendah. Nilai akurasi terendah terjadi pada $k = 2$ dengan nilai sekitar 0,56. Hal ini menunjukkan bahwa penggunaan jumlah tetangga yang terlalu sedikit menyebabkan model sangat sensitif terhadap noise dan variasi data, sehingga menghasilkan performa yang tidak stabil.

Seiring dengan meningkatnya nilai k, akurasi model cenderung mengalami peningkatan dan

menjadi lebih stabil. Pada rentang $k = 5$ hingga $k = 10$, akurasi mulai menunjukkan tren naik meskipun masih mengalami beberapa penurunan. Peningkatan performa yang lebih konsisten terlihat pada kisaran $k = 11$ hingga $k = 16$, di mana akurasi berada di atas 0,70. Akurasi tertinggi diperoleh pada $k = 15$ dengan nilai sebesar 0,7607 atau sekitar 76%, yang merupakan performa terbaik dari seluruh variasi nilai k yang diuji.

Setelah mencapai nilai $k = 15$, grafik menunjukkan bahwa akurasi cenderung stabil hingga $k = 20$ tanpa peningkatan yang signifikan. Hal ini mengindikasikan bahwa penambahan jumlah tetangga di atas nilai tersebut tidak memberikan kontribusi berarti terhadap peningkatan kinerja model, bahkan berpotensi mengurangi kemampuan model dalam membedakan batas antar kelas karena pengaruh data yang semakin luas.

Secara keseluruhan, hasil analisis ini menunjukkan bahwa pemilihan nilai parameter k sangat menentukan keberhasilan algoritma KNN dalam mengklasifikasikan penyakit liver. Nilai $k = 15$ memberikan keseimbangan terbaik antara bias dan varians sehingga menghasilkan akurasi tertinggi dan performa yang paling stabil. Oleh karena itu, nilai k tersebut dipilih sebagai parameter optimal dalam penelitian ini untuk menghasilkan model klasifikasi yang lebih akurat dan andal.

G Penentuan Nilai K Terbaik

Penentuan nilai parameter k terbaik dilakukan berdasarkan hasil pengujian variasi nilai k terhadap tingkat akurasi algoritma K-NearestNeighbor (KNN). Dari hasil pengujian yang telah dilakukan pada rentang nilai $k = 1$ hingga $k = 20$, diperoleh bahwa nilai akurasi tertinggi dicapai pada $k = 15$ dengan nilai sebesar 0,7607 atau sekitar 76%. Nilai ini merupakan performa terbaik dibandingkan dengan nilai k lainnya, sehingga dipilih sebagai parameter optimal dalam proses klasifikasi penyakit liver.

Pemilihan nilai $k = 15$ menunjukkan bahwa penggunaan jumlah tetangga yang relatif lebih banyak mampu memberikan hasil klasifikasi yang lebih stabil dan akurat. Nilai k yang terlalu kecil terbukti menghasilkan akurasi yang rendah dan tidak stabil karena model sangat sensitif terhadap noise dan variasi data. Sebaliknya, nilai k yang terlalu besar tidak memberikan peningkatan akurasi yang signifikan dan cenderung mengaburkan batas antar kelas. Oleh karena itu, nilai $k = 15$ dianggap mampu memberikan keseimbangan yang optimal antara bias dan varians.

Dengan ditetapkannya nilai k terbaik ini, model KNN kemudian digunakan kembali untuk melakukan klasifikasi akhir terhadap data uji. Penggunaan nilai k optimal diharapkan dapat meningkatkan kemampuan model dalam mengenali pola penyakit liver secara lebih tepat, mengurangi kesalahan klasifikasi, serta menghasilkan performa yang lebih andal sebagai sistem pendukung keputusan dalam membantu deteksi penyakit liver.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Nearest Neighbor (KNN) mampu mengklasifikasikan penyakit liver berdasarkan data klinis pasien dengan kinerja yang cukup baik. Proses pemeriksaan dan pra-pemrosesan data menunjukkan bahwa sebagian besar variabel pada *Indian Liver Patient Dataset* (ILPD) berada dalam kondisi lengkap, sedangkan data yang hilang pada atribut *Albumin and Globulin Ratio* dapat ditangani dengan baik sehingga tidak memengaruhi proses analisis. Hasil pengujian awal model KNN menghasilkan tingkat akurasi sebesar 67%, dengan kemampuan yang lebih baik dalam mengenali pasien penderita penyakit liver dibandingkan pasien non-penderita. Namun, hasil evaluasi juga menunjukkan masih adanya kesalahan klasifikasi, yang mengindikasikan keterbatasan model dalam membedakan kedua kelas secara konsisten. Pengujian variasi nilai parameter k menunjukkan bahwa pemilihan jumlah tetangga terdekat sangat berpengaruh terhadap kinerja model. Nilai $k = 15$ menghasilkan akurasi tertinggi sebesar 76,07% dan memberikan keseimbangan yang optimal antara stabilitas model dan ketelitian prediksi, sehingga dipilih sebagai parameter terbaik dalam penelitian ini. Secara keseluruhan, algoritma KNN memiliki potensi untuk digunakan sebagai alat bantu klasifikasi penyakit liver berbasis data klinis. Meskipun demikian, pengembangan lebih lanjut masih diperlukan, khususnya untuk meningkatkan performa pada kelas non-penderita. Penelitian selanjutnya dapat dilakukan dengan memperbaiki distribusi data, menerapkan teknik seleksi fitur, serta membandingkan algoritma KNN dengan metode klasifikasi lain guna memperoleh sistem dengan tingkat akurasi dan reliabilitas yang lebih tinggi.

DAFTAR PUSTAKA

- [1] E. Pusporani, S. Qomariyah, and I. Irhamah, “Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning,” *Inferensi*, vol. 2, no. 1, p. 25, 2019.
- [2] S. R. Velu, V. Ravi, and K. Tabianan, “Data mining in predicting liver patients using classification model,” *Health Technol. (Berl.)*, vol. 12, no. 6, pp. 1211–1235, 2022.
- [3] K. Widya Kayohana, “Klasifikasi Penyakit Hati Menggunakan Random Forest Dan Knn,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 4, pp. 7924–7929, 2024.
- [4] A. Ganji, D. Usha, and P. S. Rajakumar, “Enhanced Early Diagnosis of Liver Diseases Using Feature Selection and Machine Learning Techniques on the Indian Liver Patient Dataset,” *Scalable Comput. Pract. Exp.*, vol. 26, no. 3, pp. 1104–1115, 2025.
- [5] W. N. Rohman and I. M. A. Agastya, “Evaluation of SMOTE Technique in the Comparison of XGBoost and Random Forest Algorithms for Liver Disease Prediction,” vol. 9, no. 6, 2025.
- [6] S. Aldana and J. S. Wibowo, “Penerapan Data Mining Terhadap Klasifikasi Pasien Penderita Penyakit Liver Menggunakan Metode K-Nearest Neighbor,” *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 124, 2024.
- [7] I. N. Abrar, A. Abdullah, and S. Sucipto, “Liver Disease Classification Using the Elbow Method to Determine Optimal K in the K-Nearest Neighbor (K-NN) Algorithm,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 2, pp. 218–228, 2023.
- [8] T. A. Assegie, “Support Vector Machine And K-Nearest Neighbor Based Liver Disease Classification Model,” *Indones. J. Electron. Electromed. Eng. Med. informatics*, vol. 3, no. 1, pp. 9–14, 2021.
- [9] P. Singh, A. K. Yadav, and S. Gangwar, “Forecasting Liver Disorders with Machine Learning Models,” *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 9, pp. 237–243, 2023.
- [10] S. Sutrisno and Jupron, “Analisa Klasifikasi Penyakit Diabetes Dengan Algoritma Neural Network,” *bit-Tech*, vol. 6, no. 3, pp. 303–310, 2024.