

# OPTIMASI NILAI $K$ PADA ALGORITMA $K$ -NEAREST NEIGHBOR UNTUK KLASIFIKASI KESEHATAN JANIN

Anindya Khrisna Wardhani<sup>1</sup>, Rano Indradi Sudra<sup>2</sup>, Ega Nugraha<sup>3</sup>, Astrid Novita Putri<sup>4</sup>

<sup>1,2,3</sup> Rekam Medis dan Informasi Kesehatan, Politeknik Rukun Abdi Luhur, Kudus

<sup>4</sup> Teknik Informatika, Universitas Semarang, Semarang

<sup>1</sup>akhrisnawardhani@gmail.com, <sup>2</sup>rano.indradi@gmail.com, <sup>3</sup>eganugrahamkm@gmail.com

, <sup>4</sup>astrid@usm.ac.id

## Abstrak

Klasifikasi kesehatan janin merupakan langkah penting dalam mendukung diagnosis dini kondisi janin dan mencegah risiko komplikasi kehamilan yang dapat mengancam jiwa ibu maupun janin. Algoritma *K-Nearest Neighbor* ( $K$ -NN) menjadi salah satu metode yang umum digunakan untuk klasifikasi ini, karena kemampuannya dalam mengenali pola dari data pelatihan. Namun, performa  $K$ -NN sangat dipengaruhi oleh pemilihan nilai  $K$ , yaitu jumlah tetangga terdekat yang dipertimbangkan dalam menentukan kelas suatu data baru. Penelitian ini bertujuan untuk mengoptimalkan nilai  $K$  pada algoritma  $K$ -NN guna memperoleh akurasi tertinggi dalam klasifikasi kesehatan janin. Data yang digunakan adalah hasil pemeriksaan kardiotokografi (CTG) yang diklasifikasikan ke dalam tiga kelas yaitu *Normal*, *Suspect*, dan *Pathological*. Nilai  $K$  diuji dari 1 hingga 20 untuk menentukan nilai  $K$  optimal berdasarkan akurasi klasifikasi. Hasil penelitian menunjukkan bahwa nilai  $K$  optimal berada pada rentang  $K = 1$  hingga  $K = 3$ , dengan akurasi tertinggi dicapai pada  $K = 1$  dan  $K = 2$  sebesar 99,91%. Sementara itu, nilai  $K$  yang lebih besar dari 3 menunjukkan penurunan akurasi secara signifikan. Berdasarkan hasil ini,  $K = 3$  dipilih sebagai nilai optimal untuk menjaga keseimbangan antara akurasi tinggi dan kemampuan generalisasi model. Kesimpulannya, optimasi nilai  $K$  pada algoritma  $K$ -NN mampu meningkatkan akurasi klasifikasi kesehatan janin, yang berpotensi mendukung keputusan medis lebih akurat dan tepat waktu. Temuan ini dapat dimanfaatkan dalam pengembangan sistem prediksi kesehatan janin yang lebih andal, sehingga berkontribusi pada pengurangan risiko komplikasi kesehatan ibu dan janin.

**Kata Kunci:** *K-Nearest Neighbor*, optimasi, klasifikasi, kesehatan janin

## Abstract

*Fetal health classification is an essential step in supporting early diagnosis of fetal conditions and preventing pregnancy complications that can threaten the lives of both mother and fetus. The K-Nearest Neighbor (K-NN) algorithm is one of the commonly used methods for this classification due to its ability to recognize patterns from training data. However, the performance of K-NN is greatly influenced by the choice of the K value, which is the number of nearest neighbors considered in determining the class of new data. This study aims to optimize the K value in the K-NN algorithm to achieve the highest accuracy in fetal health classification. The data used consists of cardiotocography (CTG) examination results classified into three categories is Normal, Suspect, and Pathological. K values from 1 to 20 were tested to determine the optimal K value based on classification accuracy. The results indicate that the optimal K range is between 1 and 3, with the highest accuracy achieved at  $K = 1$  and  $K = 2$ , both at 99.91%. In contrast, K values greater than 3 show a significant decrease in accuracy. Based on these findings,  $K = 3$  is chosen as the optimal value to balance high accuracy with the model's generalization capability. In conclusion, optimizing the K value in the K-NN algorithm can improve the accuracy of fetal health classification, potentially supporting more accurate and timely medical decision-making. These findings can be applied in developing a more reliable fetal health prediction system, thereby contributing to reducing the risk of maternal and fetal health complications*

**Keywords:** *K-Nearest Neighbor, optimization, classification, fetal health*

## 1. Pendahuluan

Kesehatan janin merupakan salah satu aspek penting yang menentukan keberhasilan kehamilan dan keselamatan ibu serta bayi.

Tingginya angka kematian bayi dan ibu, khususnya di negara-negara dengan sumber daya terbatas, menuntut adanya sistem dan metode yang dapat mendeteksi secara dini kondisi kesehatan janin. Menurut data dari

Organisasi Kesehatan Dunia (WHO), sebagian besar kematian bayi baru lahir dan ibu disebabkan oleh komplikasi yang sebenarnya dapat dicegah jika terdeteksi sejak dini. Oleh karena itu, klasifikasi kesehatan janin dengan memanfaatkan data medis yang tersedia menjadi sangat penting untuk mendukung intervensi medis yang tepat waktu.

Salah satu metode yang sering digunakan dalam klasifikasi kesehatan janin adalah algoritma *K-Nearest Neighbor* (K-NN). *K-Nearest Neighbor* (K-NN) adalah algoritma *machine learning* yang digunakan dalam klasifikasi dan regresi, bekerja dengan cara mengukur kedekatan antara data baru dengan data pelatihan yang telah diklasifikasikan sebelumnya. K-NN mengukur jarak, biasanya dengan metode *Euclidean distance*, lalu memilih  $K$  tetangga terdekat untuk menentukan kelas berdasarkan mayoritas tetangga tersebut (Zhang, 2022). K-NN memiliki keunggulan kesederhanaan, fleksibilitas, dan sifatnya yang non-parametrik, namun memerlukan penyesuaian nilai  $K$  dan normalisasi data, serta bisa lambat jika jumlah data besar. Algoritma ini bekerja dengan mengidentifikasi pola dari data sebelumnya berdasarkan kedekatan jarak antara titik data yang ingin diklasifikasikan dengan data-data yang telah diberi label (Ali et al., 2020). Dalam kasus klasifikasi kesehatan janin, algoritma K-NN dapat membantu mengidentifikasi apakah kondisi janin tergolong normal, mencurigakan (*suspect*), atau patologis berdasarkan hasil pemeriksaan Kardiotokografi (CTG). Dengan demikian, algoritma ini berpotensi memberikan panduan awal bagi tenaga kesehatan dalam mengambil langkah-langkah pencegahan yang diperlukan.

Namun, salah satu tantangan dalam penggunaan algoritma K-NN adalah menentukan nilai  $K$  yang optimal. Nilai  $K$  merupakan jumlah tetangga terdekat yang dipertimbangkan dalam menentukan kelas dari data yang akan diklasifikasikan. Pemilihan nilai  $K$  yang tidak tepat dapat menyebabkan hasil klasifikasi yang tidak akurat, baik karena terlalu sensitif terhadap variasi data (*overfitting*) atau tidak cukup responsif terhadap pola yang ada (*underfitting*) (Khrisna Wardhani et al., 2020). Oleh karena itu, penting untuk melakukan optimasi nilai  $K$  agar algoritma K-NN dapat bekerja dengan akurasi maksimal dalam klasifikasi kesehatan janin. Penelitian ini

bertujuan untuk mengoptimalkan nilai  $K$  pada algoritma K-NN dalam klasifikasi kesehatan janin. Dengan optimasi ini, diharapkan algoritma K-NN dapat memberikan hasil klasifikasi yang lebih akurat, sehingga membantu meningkatkan kualitas deteksi dini kondisi janin yang berpotensi berisiko. Hal ini pada akhirnya diharapkan dapat berkontribusi dalam menurunkan angka kematian bayi dan ibu, serta mendukung upaya pencapaian tujuan pembangunan berkelanjutan dalam bidang kesehatan ibu dan anak.

Penelitian tentang algoritma *K-Nearest Neighbor* (K-NN) telah berkembang pesat, khususnya dalam bidang klasifikasi data medis seperti kesehatan janin. Algoritma K-NN dikenal sebagai metode sederhana yang efektif untuk klasifikasi data berbasis jarak. Algoritma ini bekerja dengan cara menghitung jarak antara data baru dan data pelatihan, dan mengklasifikasikan data baru tersebut berdasarkan mayoritas kelas dari  $K$  tetangga terdekatnya. Dalam konteks klasifikasi kesehatan janin, pemanfaatan algoritma K-NN telah dipelajari untuk mendeteksi kondisi kesehatan janin, yaitu apakah dalam keadaan normal, dicurigai (*suspect*), atau patologis.

Beberapa penelitian menunjukkan bahwa performa K-NN sangat dipengaruhi oleh pemilihan nilai  $K$ . Argina (2020) menerapkan algoritma K-NN untuk klasifikasi data penderita diabetes dengan fokus pada evaluasi performa melalui penghitungan akurasi, presisi, *recall*, dan F-measure untuk nilai  $K$  tertentu ( $K=3, 4, 5$ ). Hasil penelitian menunjukkan bahwa nilai  $K=3$  memberikan performa terbaik dengan akurasi 39%, presisi 65%, *recall* 36%, dan F-measure 46%, yang menekankan pentingnya pemilihan nilai  $K$  yang optimal dalam meningkatkan akurasi klasifikasi (Argina, 2020).

Cahyanti et al. (2020) juga menggunakan K-NN untuk klasifikasi dataset pasien kanker payudara dengan penerapan *cross-validation* menggunakan metode *k-fold* untuk nilai  $K=3$  hingga  $K=5$ . Studi ini menemukan bahwa K-NN pada rasio data latih dan data uji 80:20 mampu menghasilkan akurasi hingga 93% dan F-measure 94% pada nilai  $K$  tertentu, yang menunjukkan potensi tinggi metode K-NN dalam mengklasifikasi dataset medis yang kompleks dengan pengaturan data yang tepat. Sementara itu, Putri (2021) menerapkan K-NN pada dataset penyakit kardiovaskular dengan berbagai rasio dataset dan *cross-validation*.

Melalui simulasi rasio dataset yang berbeda (20:80, 50:50, 80:20) serta variasi nilai K (3 hingga 13), penelitian ini menunjukkan hasil terbaik pada K=3 dengan akurasi 87% dan F-measure 92%, menunjukkan bahwa K-NN, jika dikombinasikan dengan teknik *cross-validation*, mampu meningkatkan keakuratan klasifikasi dan relevansi prediksi dalam aplikasi kesehatan (Cahyanti et al., 2020).

Laksono et al. (2020) juga mengoptimalkan nilai K dalam algoritma K-NN untuk klasifikasi email spam dan non-spam melalui pendekatan *frequency distribution clustering* dan *k-means clustering*. Hasil penelitian menunjukkan bahwa nilai K=1 memberikan akurasi tertinggi sebesar 91.4%, sedangkan optimasi menggunakan *frequency distribution clustering* dan *k-means clustering* menghasilkan akurasi hingga 100% dan 99%, yang menggarisbawahi pentingnya optimasi parameter K untuk meningkatkan performa K-NN, terutama dalam kasus data yang tidak seimbang seperti klasifikasi spam (Laksono et al., 2020).

Penelitian mengenai algoritma K-Nearest Neighbor (K-NN) telah banyak diterapkan dalam berbagai bidang, termasuk analisis sentimen, diagnosis kesalahan mesin, dan klasifikasi berbasis pembelajaran mesin. Studi yang diterapkan dalam analisis sentimen menunjukkan bahwa algoritma K-NN efektif untuk klasifikasi opini publik terhadap kebijakan, seperti dalam penelitian yang menganalisis respons publik terhadap pembelajaran daring menggunakan data tweet. Penelitian ini memanfaatkan pembobotan kata TF-IDF dan menghasilkan akurasi tertinggi saat nilai K diatur pada 10 dengan akurasi mencapai 84,65% (Isnain et al., 2021). Di sisi lain, penelitian lain menunjukkan bahwa algoritma K-NN dapat ditingkatkan melalui metode seperti Enhanced K-NN (EK-NN) yang menggabungkan ekstraksi fitur otomatis dan klasifikasi berbasis kasus. EK-NN mengoptimalkan pencarian tetangga terdekat dengan ekstraksi fitur sparsa, meningkatkan ketepatan dalam diagnosis kondisi kesehatan komponen mesin (Lu et al., 2021). Review literatur tentang K-NN juga menunjukkan penggunaannya dalam berbagai aplikasi pembelajaran yang diawasi, dari klasifikasi teks hingga deteksi penyakit, yang mendukung perannya sebagai algoritma sederhana namun

efektif dalam tugas klasifikasi berbasis pembelajaran mesin (Suyal & Goyal, 2022)

Algoritma K-NN adalah salah satu metode pembelajaran mesin paling sederhana dan sering digunakan dalam klasifikasi, terutama di bidang medis dan keamanan jaringan. Kelebihan K-NN terletak pada kemampuannya untuk melakukan klasifikasi berdasarkan kedekatan data latih terhadap data uji menggunakan metrik jarak seperti Euclidean dan Mahalanobis. Meskipun mudah digunakan, algoritma ini memiliki kekurangan seperti ketergantungan pada parameter K yang optimal dan pemrosesan yang lambat pada dataset besar karena sifatnya yang "lazy learning" (tidak melakukan pelatihan sebelumnya) (Liu et al., 2022; Uddin et al., 2022).

Berbagai varian K-NN telah dikembangkan untuk mengatasi kekurangan ini, termasuk K-NN adaptif, fuzzy K-NN, K-NN berbasis clustering, dan K-NN berbobot. Varian ini bertujuan untuk meningkatkan akurasi dan efisiensi dengan mengoptimalkan parameter K, memperbaiki metrik jarak, serta mengurangi kompleksitas komputasi. Sebagai contoh, varian Hassanat K-NN yang menggunakan metrik jarak Hassanat terbukti lebih efektif untuk dataset medis dengan akurasi yang tinggi. Pendekatan berbasis ensemble juga menunjukkan peningkatan kinerja dengan menghilangkan ketergantungan pada parameter K tetap (Uddin et al., 2022)

Di sisi lain, dalam konteks deteksi intrusi pada jaringan sensor nirkabel (WSN), algoritma K-NN sering dikombinasikan dengan metode optimasi, seperti Arithmetic Optimization Algorithm (AOA) dan pendekatan edge computing, untuk mengatasi keterbatasan daya komputasi dan ruang penyimpanan. Model deteksi intrusi berbasis K-NN yang diperkuat oleh AOA berhasil meningkatkan akurasi dan kecepatan deteksi serangan DoS (Denial of Service) di WSN. Penggunaan algoritma AOA yang diparalelkan dengan strategi Lévy Flight menambah diversitas populasi dalam proses optimasi, sehingga menghindari perangkap solusi lokal dan meningkatkan performa K-NN secara signifikan dalam skenario ini (Liu et al., 2022)

Algoritma K-Nearest Neighbors (K-NN) telah lama digunakan sebagai salah satu teknik dalam pembelajaran mesin untuk klasifikasi data. K-NN bekerja dengan membandingkan

titik data baru dengan sejumlah titik data terdekat yang ada dalam data pelatihan, di mana nilai KKK merupakan jumlah tetangga terdekat yang digunakan untuk menentukan kelas titik data baru. Tantangan utama dalam penerapan K-NN adalah menentukan nilai KKK yang optimal serta pemilihan metrik jarak yang paling sesuai untuk mencapai akurasi terbaik dalam prediksi data baru (Pietrzak & Wolkiewicz, 2021). Selain itu, penggunaan K-NN dalam dataset yang tidak seimbang memerlukan pendekatan khusus, seperti teknik pembelajaran cost-sensitive yang memberikan bobot lebih tinggi pada kelas minoritas untuk meningkatkan akurasi klasifikasi, terutama dalam aplikasi diagnosis penyakit kanker yang menggunakan biomarker protein darah (Song & Li, 2022).

Dalam bidang diagnosis medis, K-NN telah diterapkan secara luas, misalnya dalam deteksi penyakit jantung dan kanker. Dalam kasus deteksi penyakit jantung, studi menunjukkan bahwa algoritma K-NN cenderung memiliki akurasi yang lebih rendah dibandingkan dengan Support Vector Machine (SVM), terutama tanpa normalisasi data. Meskipun demikian, K-NN masih sering digunakan karena kesederhanaan implementasi dan efisiensinya dalam pengelompokan data medis berdasarkan kesamaan atribut (Anggoro, 2020). Di sisi lain, dalam diagnosis penyakit kanker, modifikasi K-NN berbasis entropi dan pembelajaran cost-sensitive telah diusulkan untuk menangani data yang sangat tidak seimbang, seperti pada prediksi kanker hati dan ovarium berdasarkan marker protein (Song & Li, 2022).

Dalam konteks industri, seperti deteksi kerusakan motor listrik, K-NN digunakan dalam analisis komponen simetris arus stator untuk mendeteksi dan mengklasifikasikan kesalahan pada gulungan stator motor sinkron magnet permanen (PMSM). Efektivitas K-NN dalam aplikasi ini didukung oleh penggunaan analisis spektral yang memfasilitasi deteksi kerusakan sejak tahap awal, sehingga mengurangi potensi kerugian akibat kerusakan sistem motor secara keseluruhan (electronics-10-01786).

Literatur mengenai klasifikasi diabetes menggunakan algoritma *k-nearest neighbors* (K-NN) berfokus pada penggunaan metode pembelajaran mesin untuk mengklasifikasikan dan mendiagnosis penyakit berdasarkan data

medis. Studi ini menunjukkan K-NN sebagai algoritma populer yang memiliki beragam jenis, seperti Fine K-NN, Weighted K-NN, Medium K-NN, dan Cubic K-NN, yang masing-masing memiliki akurasi berbeda dalam mengklasifikasikan data diabetes (Ali et al., 2020).

Beberapa penelitian juga menyoroti upaya untuk mengatasi tantangan dalam pengaturan nilai KKK dan pencarian tetangga terdekat secara optimal, karena langkah ini bisa menjadi hambatan dalam efisiensi K-NN. Zhang dan Li mengusulkan pendekatan satu langkah dengan pemanfaatan matriks hubungan, yang bertujuan untuk meningkatkan efisiensi proses K-NN dengan menyatukan pengaturan nilai KKK dan pencarian tetangga terdekat dalam satu komputasi (Zhang & Li, 2023).

Berdasarkan berbagai penelitian, algoritma *K-Nearest Neighbor* (K-NN) terbukti bergantung pada pemilihan nilai K yang optimal untuk mencapai performa terbaik dalam klasifikasi, terutama pada dataset medis dan aplikasi lainnya. Argina (2020) dan Cahyanti et al. (2020) menunjukkan bahwa variasi nilai K memengaruhi akurasi, presisi, *recall*, dan F-measure dalam klasifikasi penyakit seperti diabetes dan kanker, di mana  $K=3$  menghasilkan performa optimal dalam kedua kasus ini. Putri (2021) menambahkan bahwa teknik *cross-validation* mampu meningkatkan keakuratan prediksi K-NN pada klasifikasi penyakit kardiovaskular, menekankan peran penting dari metode validasi dalam meningkatkan performa K-NN. Laksono et al. (2020) melakukan optimasi pada nilai K dalam klasifikasi email spam dengan pendekatan *frequency distribution clustering* dan *k-means clustering*, yang menunjukkan bahwa optimasi K dapat meningkatkan akurasi hingga 100%.

Penelitian ini mengindikasikan bahwa K-NN adalah algoritma yang efektif namun memiliki ketergantungan tinggi pada nilai K yang dipilih serta sensitivitas terhadap jenis dataset yang digunakan. Upaya optimasi K untuk berbagai aplikasi, termasuk kesehatan dan klasifikasi email, telah dilakukan dengan metode seperti *cross-validation*, pembobotan berbasis frekuensi, dan *clustering*. Novelty untuk penelitian "Optimasi Nilai K pada Algoritma K-Nearest Neighbor untuk Klasifikasi Kesehatan Janin" dapat difokuskan pada pengembangan metode optimasi K yang dirancang khusus untuk mengatasi tantangan

dalam klasifikasi kesehatan janin, yang mungkin memerlukan pendekatan baru atau modifikasi teknik validasi untuk menghadapi ketidakseimbangan data dan kompleksitas atribut dalam data janin. Dengan demikian, penelitian ini berpotensi memberikan solusi yang lebih efektif dan presisi dalam klasifikasi kesehatan janin menggunakan K-NN.

## 2. Metode Penelitian

Penelitian ini bertujuan untuk mengoptimalkan nilai  $K$  pada algoritma  $K$ -Nearest Neighbor ( $K$ -NN) dalam klasifikasi kesehatan janin dengan menggunakan data pemeriksaan Kardiotokografi (CTG). Langkah-langkah metodologi dalam penelitian ini meliputi tahap pengumpulan data, praproses data, implementasi algoritma  $K$ -NN, optimasi nilai  $K$ , dan evaluasi kinerja model.

Data yang digunakan dalam penelitian ini adalah data pemeriksaan Kardiotokografi (CTG) yang didapatkan dari *Kaggle*. Data mencakup sejumlah fitur seperti detak jantung janin, frekuensi gerakan janin, dan kontraksi uterus. Dataset penelitian ini mengandung 2.126 catatan yang telah diklasifikasikan oleh dokter spesialis obstetri menjadi tiga kelas, yaitu *Normal*, *Suspect*, dan *Pathological*. Data ini akan dibagi menjadi dua bagian yaitu data pelatihan (80%) dan data pengujian (20%) menggunakan metode *stratified sampling* agar komposisi kelas terjaga di setiap set data (Khrisna Wardhani et al., 2020).

Tahap praproses data mencakup beberapa langkah yang diperlukan untuk memastikan kualitas data yang baik sebelum implementasi algoritma  $K$ -NN. Langkah-langkah ini meliputi:

- Normalisasi Data:** Karena  $K$ -NN sensitif terhadap skala fitur, seluruh fitur akan dinormalisasi menggunakan metode *min-max normalization*, yang akan mengubah nilai fitur ke rentang  $[0, 1]$ .
- Pemeriksaan Outlier:** Outlier yang signifikan akan diidentifikasi dan ditangani untuk menghindari distorsi dalam perhitungan jarak.
- Pembagian Data:** Data yang telah dinormalisasi akan dibagi ke dalam data pelatihan dan data pengujian (Ardilla et al., 2021).

Algoritma  $K$ -NN akan diterapkan pada data pelatihan untuk membangun model

klasifikasi kesehatan janin. Beberapa nilai awal  $K$  (misalnya,  $K = 1, 2, 3, 4, 5$ , dst.) akan diuji untuk mengetahui pengaruhnya terhadap akurasi klasifikasi. Jarak antara data baru dan data pelatihan akan dihitung menggunakan *Euclidean distance*, yang merupakan metode umum untuk mengukur jarak dalam algoritma  $K$ -NN dengan rumus sebagai berikut :

Jarak Euclidean =

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + \dots + (z_n - z_n)^2} \quad (1)$$

di mana:

$x_1, y_1, z_1, \dots$  adalah koordinat titik pertama

$x_2, y_2, z_2, \dots$  adalah koordinat titik kedua

$n$  adalah jumlah dimensi

Optimasi nilai  $K$  dilakukan untuk menemukan nilai  $K$  yang menghasilkan akurasi terbaik pada data pengujian. Pendekatan yang digunakan untuk optimasi adalah metode *grid search* dan *cross-validation* dengan menggunakan 5-fold *cross-validation* pada data pelatihan. Pada tahap ini, berbagai nilai  $K$  akan dicoba, dan hasil klasifikasi dari setiap nilai  $K$  akan dibandingkan untuk menemukan nilai yang memberikan hasil paling akurat. Metode ini dipilih karena *grid search* memungkinkan eksplorasi komprehensif terhadap parameter  $K$ , sementara *cross-validation* membantu menghindari *overfitting* pada model.

Model yang telah dioptimalkan akan dievaluasi menggunakan data pengujian berdasarkan akurasi yaitu persentase prediksi yang benar dari keseluruhan prediksi. Metode evaluasi ini bertujuan untuk memastikan bahwa nilai  $K$  yang terpilih mampu menghasilkan model yang robust dan akurat dalam mengklasifikasikan kesehatan janin. Setelah optimasi dan evaluasi model selesai, hasil klasifikasi akan dianalisis untuk melihat apakah nilai  $K$  berhasil meningkatkan akurasi dan kinerja klasifikasi kesehatan janin secara keseluruhan. Hasil penelitian ini akan dibandingkan dengan hasil dari literatur sebelumnya untuk mengukur sejauh mana kontribusi dari optimasi nilai  $K$  dalam meningkatkan akurasi model  $K$ -NN pada klasifikasi kesehatan janin.

## 3. Hasil Penelitian dan Pembahasan

Penelitian ini berfokus pada optimasi nilai  $K$  dalam algoritma  $K$ -Nearest Neighbor ( $K$ -NN) untuk klasifikasi kesehatan janin, dengan tujuan menentukan nilai  $K$  yang optimal guna memaksimalkan akurasi klasifikasi. Hasil penelitian ini menunjukkan bahwa akurasi

algoritma K-NN sangat dipengaruhi oleh nilai  $K$ , dengan hasil akurasi sebagai berikut:

Tabel 1 Skor Akurasi

Nilai $K$	Persentase (%)
1	99,91
2	99,91
3	98,64
4	95,91
5	92,57
6	93,56
7	90,92
8	91,77
9	90,12
10	90,83
11	89,42
12	90,12
13	88,57
14	88,85
15	87,63
16	88,1
17	87,21
18	87,25
19	86,41
20	86,69

Berdasarkan hasil akurasi di atas, dapat dianalisis jika nilai  $K = 1$  dan  $K = 2$  menghasilkan akurasi tertinggi, yaitu 99,91%. Hal ini menunjukkan bahwa K-NN memberikan performa terbaik ketika memperhitungkan sedikit tetangga terdekat untuk klasifikasi kesehatan janin. Dengan mempertimbangkan hanya satu atau dua tetangga terdekat, model dapat menangkap pola lokal dari data yang mungkin relevan dalam menentukan kondisi kesehatan janin. Namun, nilai  $K$  yang terlalu kecil berpotensi menyebabkan *overfitting*, karena model menjadi sangat sensitif terhadap data terdekat dan kurang mampu mengenali pola umum dari seluruh dataset. Ketika  $K$  meningkat dari 3 hingga 20, terjadi tren penurunan akurasi secara bertahap. Pada  $K = 3$ , akurasi masih relatif tinggi di angka 98,64%, tetapi mulai menurun signifikan pada  $K = 4$  (95,91%) dan semakin menurun hingga mencapai 86,41% pada  $K = 19$ . Hal ini disebabkan oleh efek *smoothing* yang terjadi ketika  $K$  lebih besar, di mana model menjadi kurang sensitif terhadap pola yang spesifik pada data dan lebih cenderung mengklasifikasikan data

baru berdasarkan rata-rata tetangga yang lebih luas. Nilai  $K$  yang besar dapat menyebabkan *underfitting*, di mana model kehilangan kemampuan untuk mengenali pola-pola penting dalam data kesehatan janin. Pada nilai  $K$  menengah ( $K = 6$  hingga  $K = 12$ ), terdapat variasi dalam akurasi, dengan nilai yang cenderung stabil di sekitar 90% hingga 93%. Pada  $K = 6$ , akurasi naik sedikit menjadi 93,56% setelah penurunan pada  $K = 5$ , tetapi kemudian akurasi kembali berfluktuasi hingga  $K = 12$ . Fluktuasi ini mengindikasikan bahwa, meskipun model dengan nilai  $K$  menengah dapat memberikan performa yang cukup baik, pemilihan nilai  $K$  dalam rentang ini tidak memberikan hasil akurasi yang optimal. Pada nilai  $K$  yang lebih besar dari 12, akurasi model semakin menurun dan stabil pada rentang 86% hingga 88%. Pada  $K = 13$  hingga  $K = 20$ , akurasi tertinggi hanya mencapai 88,85% pada  $K = 14$ , dengan akurasi terendah pada  $K = 19$  sebesar 86,41%. Nilai  $K$  yang terlalu besar menyebabkan model terlalu generalis dalam memprediksi kelas data baru, sehingga kurang peka terhadap variasi pada data kesehatan janin. Model dengan nilai  $K$  besar cenderung mengabaikan perbedaan antara kelas dan lebih mengarah pada hasil klasifikasi yang rata-rata. Berdasarkan hasil ini, nilai  $K$  optimal berada pada rentang  $K = 1$  hingga  $K = 3$ , dengan akurasi tertinggi dicapai pada  $K = 1$  dan  $K = 2$  sebesar 99,91%. Di sisi lain,  $K = 3$  juga memberikan akurasi tinggi sebesar 98,64% dan memiliki risiko *overfitting* yang lebih rendah dibandingkan  $K = 1$  atau  $K = 2$ . Oleh karena itu,  $K = 3$  dapat dipertimbangkan sebagai nilai  $K$  optimal yang menghasilkan keseimbangan antara akurasi tinggi dan kemampuan generalisasi model terhadap data yang baru. Nilai  $K$  yang lebih besar dari 3 menghasilkan penurunan akurasi yang signifikan, menunjukkan bahwa model mulai kehilangan sensitivitas terhadap pola yang unik dalam data. Meskipun  $K$  yang besar lebih stabil terhadap data bervariasi, dalam kasus ini stabilitas tersebut mengorbankan akurasi model, yang sangat penting dalam klasifikasi kesehatan janin.

Pemilihan nilai  $K$  yang optimal penting dalam memastikan akurasi tinggi dalam klasifikasi kesehatan janin, yang dapat mendukung diagnosis yang lebih cepat dan lebih akurat untuk kesehatan janin. Model dengan akurasi tinggi ini dapat membantu mendeteksi kondisi *Normal*, *Suspect*, atau *Pathological* pada janin dengan lebih akurat, sehingga intervensi medis yang diperlukan dapat diberikan lebih awal.

#### 4. Kesimpulan

Hasil penelitian menunjukkan bahwa nilai  $K$  optimal untuk klasifikasi kesehatan janin dengan algoritma K-NN adalah antara 1 hingga 3, dengan akurasi tertinggi pada  $K = 1$  dan  $K = 2$  sebesar 99,91%, serta  $K = 3$  sebesar 98,64% yang menawarkan keseimbangan lebih baik. Nilai  $K$  yang lebih besar dari 3 menunjukkan penurunan akurasi yang signifikan, yang mengindikasikan bahwa nilai  $K$  yang lebih tinggi membuat model kurang efektif dalam mengklasifikasikan kondisi kesehatan janin secara akurat. Optimasi nilai  $K$  ini diharapkan dapat meningkatkan keandalan sistem pendeteksian dini kesehatan janin, sehingga mendukung pengambilan keputusan medis yang lebih baik dan berpotensi mengurangi angka komplikasi kesehatan janin. Penelitian ini memiliki beberapa keterbatasan yaitu model K-NN sangat bergantung pada distribusi data dalam set pelatihan, sehingga jika ada perubahan pada distribusi data, performa model mungkin tidak seoptimal pada data pelatihan yang ada. Pemilihan nilai  $K$  dalam penelitian ini hanya didasarkan pada akurasi. Penelitian lanjutan dapat mempertimbangkan metrik tambahan seperti *precision*, *recall*, dan *F1-score* untuk mengukur performa model lebih komprehensif pada tiap kelas kesehatan janin.

#### 5. Daftar Pustaka

- Ali, A., Alrubei, M., Hassan, L. F. M., Al-Ja'afari, M., & Abdulwahed, S. (2020). Diabetes classification based on KNN. *IJUM Engineering Journal*, 21(1), 175–181. <https://doi.org/10.31436/iiumej.v21i1.1206>
- Anggoro, D. A. (2020). Comparison of Accuracy Level of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) Algorithms in Predicting Heart Disease. *International Journal of Emerging Trends in Engineering Research*, 8(5), 1689–1694. <https://doi.org/10.30534/ijeter/2020/32852020>
- Ardilla, Y., Wardhani, A. K., Mustika, Manuhutu, A., Ahmad, N., Hasbi, I., Guntoro, Manuhutu, M. A., Ridwan, M., Hozairi, Alim, S., Romli, I., Religia, Y., Octafian, T., Sufandi, U. U., & Ernawati, I. (2021). *DATA MINING DAN APLIKASINYA* (Rismawati, Ed.; 1st ed.). Widina Bhakti Persada Bandung. <https://repository.penerbitwidina.com/media/publications/351768-data-mining-dan-aplikasinya-7b2a8129.pdf>
- Argina, A. M. (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, 1(2), 29–33.
- Cahyanti, D., Rahmayani, A., & Ainy Husniar, S. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. *Indonesian Journal of Data and Science*, 1(2), 39–43.
- Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(2), 121. <https://doi.org/10.22146/ijccs.65176>
- Khrisna Wardhani, A., Novita Putri, A., Fathi Salim Ashour, S., Medis dan Informasi Kesehatan, R., Informatika, T., Science, C., & Rukun Abdi Luhur, P. (2020). *Telematika An Improved K-NN Algorithm and Bagging for Liver Disease Classification*. 15(2), 100–107. <https://doi.org/10.35671/telematika.v15i2.1247>
- Laksono, E. P., Basuki, A., & Abdurrachman Bachtiar, F. (2020). Optimasi Nilai K pada Algoritma KNN untuk Klasifikasi Spam dan Ham Email. *JURNAL RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(2), 377–383.
- Liu, G., Zhao, H., Fan, F., Liu, G., Xu, Q., & Nazir, S. (2022). An Enhanced Intrusion Detection Model Based on Improved kNN in WSNs. *Sensors*, 22(4). <https://doi.org/10.3390/s22041407>
- Lu, J., Qian, W., Li, S., & Cui, R. (2021). Enhanced k-nearest neighbor for intelligent fault diagnosis of rotating machinery. *Applied Sciences (Switzerland)*, 11(3), 1–15. <https://doi.org/10.3390/app11030919>
- Pietrzak, P., & Wolkiewicz, M. (2021). On-line detection and classification of pmsm stator winding faults based on stator current symmetrical components analysis and the knn algorithm. *Electronics (Switzerland)*, 10(15). <https://doi.org/10.3390/electronics10151786>
- Song, C., & Li, X. (2022). Cost-Sensitive KNN Algorithm for Cancer Prediction Based on Entropy Analysis. *Entropy*, 24(2). <https://doi.org/10.3390/e24020253>
- Suyal, M., & Goyal, P. (2022). A Review on Analysis of K-Nearest Neighbor Classification Machine Learning Algorithms based on Supervised Learning. In *International Journal of Engineering Trends and Technology* (Vol. 70, Issue 7, pp. 43–48). Seventh Sense Research Group. <https://doi.org/10.14445/22315381/IJETT-V70I7P205>

- Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-10358-x>
- Zhang, S. (2022). Challenges in KNN Classification. *IEEE Transactions on Knowledge and Data Engineering*, 34(10), 4663–4675. <https://doi.org/10.1109/TKDE.2021.3049250>
- Zhang, S., & Li, J. (2023). KNN Classification With One-Step Computation. *IEEE Transactions on Knowledge and Data Engineering*, 35(3), 2711–2723. <https://doi.org/10.1109/TKDE.2021.3119140>