

Optimasi Hyperparameter pada Algoritma K-Nearest Neighbors (KNN) untuk Prediksi Diabetes Berdasarkan Data Klinis Pasien

Deby Shandy Aulia Aryanto¹

¹Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar City, North Sumatra, Indonesia

E-Mail : debyshandy5@gmail.com

Article Info

Article history:

Received Jan 10, 2025

Revised Feb 01, 2025

Accepted Feb 20, 2025

Kata Kunci:

K-Nearest Neighbors (KNN)

Optimasi Hyperparameter

Prediksi Diabetes

Data Klinis Pasien

Machine Learning

Keywords:

K-Nearest Neighbors (KNN)

Hyperparameter Optimization

Diabetes Prediction

Patient Clinical Data

Machine Learning

ABSTRAK

Penelitian ini bertujuan untuk mengoptimalkan algoritma K-Nearest Neighbors (KNN) melalui pendekatan optimasi hyperparameter guna meningkatkan akurasi prediksi diabetes berdasarkan data klinis pasien. Data yang digunakan dalam penelitian ini diperoleh dari UCI Machine Learning Repository dan mencakup berbagai variabel medis seperti usia, kadar gula darah (HbA1c), tekanan darah, kolesterol, serta beberapa parameter laboratorium lainnya. Model KNN digunakan untuk mengklasifikasikan apakah seorang pasien berisiko menderita diabetes atau tidak. Optimasi dilakukan untuk menemukan nilai parameter terbaik seperti jumlah tetangga (k) dan jenis metrik jarak, sehingga kinerja model dapat ditingkatkan secara signifikan. Hasil evaluasi menunjukkan bahwa model baseline menghasilkan akurasi sebesar 90%, sementara model yang telah dioptimalkan meningkat menjadi 91%. Peningkatan juga terjadi pada metrik lain seperti recall, precision, dan f1-score. Penelitian ini membuktikan bahwa penerapan optimasi hyperparameter pada algoritma KNN mampu meningkatkan kemampuan prediktif model dalam klasifikasi risiko diabetes dan memberikan kontribusi dalam pengembangan sistem pendukung keputusan berbasis data klinis.

ABSTRACT

This study aims to optimize the K-Nearest Neighbors (KNN) algorithm through a hyperparameter optimization approach to improve the accuracy of diabetes prediction based on patient clinical data. The data used in this study were obtained from the UCI Machine Learning Repository and include various medical variables such as age, blood sugar levels (HbA1c), blood pressure, cholesterol, and several other laboratory parameters. The KNN model is used to classify whether a patient is at risk of diabetes or not. Optimization is carried out to find the best parameter values such as the number of neighbors (k) and the type of distance metric, so that model performance can be significantly improved. The evaluation results show that the baseline model produces an accuracy of 90%, while the optimized model increases to 91%. Improvements also occur in other metrics such as recall, precision, and f1-score. This study proves that the application of hyperparameter optimization to the KNN algorithm can improve the predictive ability of the model in diabetes risk classification and contribute to the development of clinical data-based decision support systems.

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



Corresponding Author:

Name Deby Shandy Aulia Aryanto,

Sistem Informasi, STIKOM Tunas Bangsa,

Alamat No.1, 2 & 3, Jl. Sudirman, Kelurahan Banjar, Kec. Siantar Barat, Pematangsiantar, Sumatra Utara, Indonesia.

Email: debyshandy5@gmail.com

1. PENDAHULUAN

Diabetes merupakan salah satu penyakit kronis yang menjadi masalah kesehatan global dan berdampak signifikan terhadap kualitas hidup serta pembiayaan kesehatan masyarakat [1]-[5]. Deteksi dini terhadap potensi diabetes sangat penting untuk mencegah komplikasi jangka panjang dan meningkatkan efektivitas penanganan medis [6]-[8]. Oleh karena itu, prediksi diabetes berdasarkan data klinis pasien telah menjadi fokus utama dalam bidang data mining dan pembelajaran mesin [9]-[12]. Salah satu pendekatan yang umum digunakan dalam prediksi tersebut adalah algoritma klasifikasi, yang bertujuan untuk mengidentifikasi apakah seseorang berisiko menderita diabetes atau tidak berdasarkan atribut-atribut klinis tertentu (Charbuty & Abdulazeez, 2021; Dissanayake & Johar, 2021; Gasparetto et al., 2022).

K-Nearest Neighbors (KNN) merupakan salah satu algoritma klasifikasi yang sederhana namun efektif dalam melakukan prediksi berbasis kemiripan antara data [16]-[18]. Algoritma ini bekerja dengan cara menghitung jarak antara data uji dengan data latih, kemudian menentukan kelas berdasarkan mayoritas dari tetangga terdekat [19]-[20]. Namun, kinerja dari algoritma KNN sangat dipengaruhi oleh pemilihan nilai hyperparameter seperti jumlah tetangga (K), jenis metrik jarak, dan bobot voting antar tetangga [21]-[23]. Pemilihan nilai hyperparameter yang tidak optimal dapat menyebabkan penurunan akurasi dan kinerja prediksi secara keseluruhan [24]-[27]. Oleh karena itu, diperlukan metode optimasi hyperparameter agar algoritma KNN dapat mencapai performa terbaik dalam klasifikasi [28]-[29].

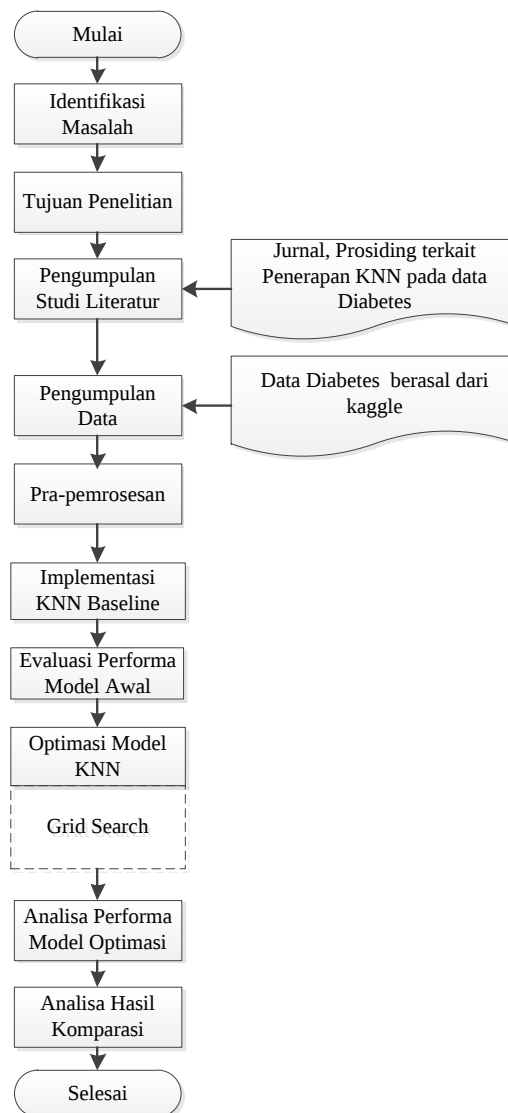
Grid Search merupakan salah satu metode optimasi hyperparameter yang banyak digunakan dalam pembelajaran mesin [30]-[32]. Teknik ini bekerja dengan mengevaluasi seluruh kombinasi parameter yang telah ditentukan sebelumnya secara sistematis, sehingga dapat mengidentifikasi kombinasi terbaik berdasarkan metrik evaluasi tertentu, seperti akurasi atau f1-score [33]-[35]. Dalam konteks prediksi diabetes, penerapan Grid Search untuk optimasi model KNN diharapkan dapat meningkatkan akurasi prediksi serta memperbaiki kemampuan model dalam membedakan antara pasien yang berisiko dan yang tidak berisiko [36]-[38].

Beberapa penelitian sebelumnya telah menunjukkan bahwa algoritma KNN dapat memberikan hasil yang cukup baik dalam klasifikasi penyakit, termasuk diabetes. Misalnya, penelitian oleh Salem et al., (2022) menggunakan algoritma KNN untuk memprediksi diabetes dan memperoleh akurasi sebesar 90,63% tanpa melakukan tuning hyperparameter (Salem et al., 2022). Hal ini menunjukkan bahwa masih terdapat celah dalam pengoptimalan algoritma KNN, terutama pada aspek tuning hyperparameter yang dapat memberikan peningkatan performa signifikan.

Dengan latar belakang tersebut, penelitian ini bertujuan untuk melakukan optimasi hyperparameter pada algoritma K-Nearest Neighbors menggunakan teknik Grid Search untuk meningkatkan akurasi prediksi diabetes berdasarkan data klinis pasien. Penelitian ini menggunakan dataset terbuka yang berisi berbagai atribut medis seperti kadar glukosa, tekanan darah, dan indeks massa tubuh (BMI). Dengan menguji berbagai kombinasi nilai parameter seperti jumlah tetangga, metrik jarak, dan bobot voting, diharapkan dapat diperoleh model KNN yang paling optimal. Hasil dari penelitian ini diharapkan tidak hanya dapat memberikan kontribusi dalam pengembangan model klasifikasi medis, tetapi juga dapat menjadi referensi bagi praktisi kesehatan dan pengambil kebijakan dalam mengambil langkah-langkah preventif terhadap penyakit diabetes secara lebih tepat sasaran.

2. METODE PENELITIAN

Untuk memastikan bahwa hasilnya dapat dipertanggung jawabkan secara ilmiah, penelitian ini dilakukan secara empiris, logis dan terstruktur, mulai dari merumuskan masalah hingga menyusun laporan (Tesfai et al., 2022). Pada Gambar 1 menunjukkan struktur tahapan penelitian yang diajukan.



Gambar 1. Tahapan Penelitian

3.2. Identifikasi masalah

Langkah ini mencakup proses identifikasi masalah yang menjadi dasar dalam pelaksanaan penelitian. Masalah utama yang diangkat adalah tantangan dalam memprediksi risiko diabetes berdasarkan data klinis pasien, di mana akurasi prediksi sering kali dipengaruhi oleh kualitas data, pemilihan fitur yang tepat, serta parameter yang digunakan dalam model klasifikasi. Dalam konteks ini, pendekatan tradisional tanpa optimasi sering kali menghasilkan performa yang kurang optimal, karena tidak mampu menyesuaikan parameter model dengan karakteristik data yang kompleks dan bervariasi.

Oleh karena itu, diperlukan pendekatan yang lebih cerdas dan sistematis dalam membangun model klasifikasi, salah satunya dengan mengoptimalkan hyperparameter model menggunakan metode pencarian yang tepat. Meskipun algoritma K-Nearest Neighbors (KNN) dikenal sederhana dan cukup efektif untuk klasifikasi medis, performanya sangat dipengaruhi oleh nilai K dan skema jarak yang digunakan. Penelitian ini berfokus pada optimasi hyperparameter pada algoritma KNN untuk meningkatkan akurasi dalam prediksi diabetes berdasarkan data klinis, dengan menggunakan teknik Grid Search sebagai metode pencarian kombinasi parameter terbaik. Dengan pendekatan ini, penelitian bertujuan untuk membuktikan bahwa KNN yang dioptimalkan dapat memberikan hasil prediksi yang lebih akurat dan andal dalam konteks deteksi dini penyakit diabetes.

3.3. Tujuan Penelitian

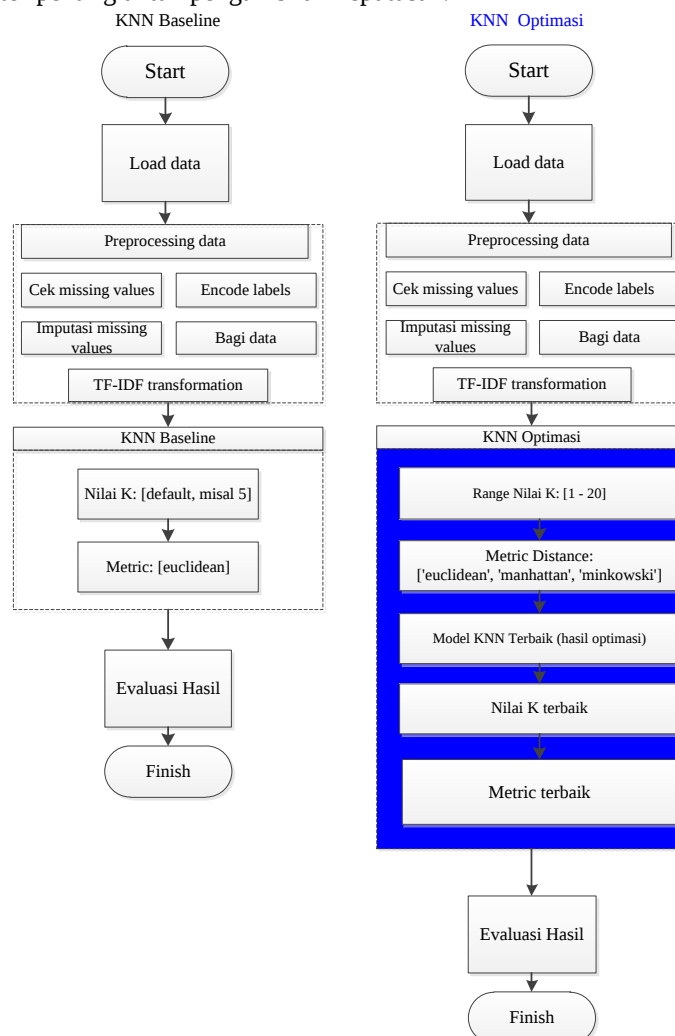
Mengevaluasi kemampuan Grid Search dalam mengoptimalkan algoritma K-Nearest Neighbors (KNN) untuk prediksi diabetes berdasarkan data klinis pasien merupakan tujuan utama dari penelitian ini untuk menyelesaikan permasalahan akurasi prediksi yang rendah pada model tanpa optimasi. Selanjutnya, model KNN yang paling efektif dan efisien dalam mengklasifikasikan kondisi diabetes akan diidentifikasi berdasarkan hasil evaluasi performa model yang telah dioptimalkan.

3.4. Pengumpulan Studi Literatur

Pada tahap pengumpulan studi literatur, dilakukan penelusuran terhadap penelitian-penelitian terdahulu terkait prediksi diabetes berdasarkan data klinis pasien menggunakan algoritma machine learning, yang menjadi landasan (Statement of DA) dalam memahami penerapan Grid Search dan algoritma K-Nearest Neighbors (KNN). Kajian ini mencakup eksplorasi berbagai algoritma klasifikasi yang digunakan dalam diagnosis penyakit diabetes, serta penilaian terhadap efektivitas Grid Search dalam mengoptimalkan parameter pada model KNN. Studi ini menjadi dasar pemilihan pendekatan optimasi menggunakan Grid Search dalam model prediksi diabetes pada penelitian ini.

3.4.1. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) adalah algoritma machine learning berbasis instance-based learning yang digunakan untuk tugas klasifikasi dan regresi [41]-[42]. KNN bekerja dengan cara mengidentifikasi sejumlah tetangga terdekat (k) dari suatu data uji berdasarkan jarak tertentu, seperti Euclidean distance, lalu menentukan kelas atau nilai berdasarkan mayoritas kelas dari tetangga-tetangga tersebut [43]-[44]. Karena sifatnya yang sederhana namun efektif, KNN banyak digunakan dalam berbagai aplikasi, termasuk dalam diagnosis medis seperti prediksi diabetes berdasarkan data klinis pasien, di mana kedekatan karakteristik pasien menjadi indikator penting untuk pengambilan keputusan.



Gambar 2. Model KNN

3.5. Pengumpulan Data

Langkah awal dari proses pengumpulan data dalam penelitian ini adalah memperoleh dataset yang relevan, dengan fokus studi pada diagnosis diabetes berdasarkan data klinis pasien. Dataset yang digunakan dalam penelitian ini adalah Pima Indians Diabetes Dataset (PIDD), yang secara luas digunakan dalam berbagai studi klasifikasi medis. Dataset ini berisi informasi klinis dari pasien wanita keturunan Pima Indian, seperti kadar glukosa, tekanan darah, indeks massa tubuh (BMI), usia, jumlah kehamilan, serta hasil diagnosis diabetes. Data ini menjadi dasar dalam membangun model prediksi menggunakan algoritma KNN, baik untuk baseline maupun versi yang telah dioptimasi. Berikut merupakan contoh atribut data yang digunakan dalam penelitian ini.

Tabel 1. Contoh Data Penelitian

ID	No_Pation	Gender	AGE	Urea	Cr	...	CLASS
502	17975	F	50	4,7	46	...	N
735	34221	M	26	4,5	62	...	N
420	47975	F	50	4,7	46	...	N
680	87656	F	50	4,7	46	...	N
504	34223	M	33	7,1	46	...	N
634	34224	F	45	2,3	24	...	N
721	34225	F	50	2	50	...	N
421	34227	M	48	4,7	47	...	N
670	34229	M	43	2,6	67	...	N
...	...	...	...	...	...	...	...
248	24054	M	54	5	67	...	Y

Data yang digunakan dalam penelitian ini terdiri dari 1000 entri yang mencakup berbagai variabel klinis yang berkaitan dengan prediksi diabetes pada pasien wanita keturunan Pima Indian. Setiap entri dalam dataset ini mencakup informasi seperti jumlah kehamilan, kadar glukosa darah, tekanan darah diastolik, ketebalan lipatan kulit triceps, kadar insulin serum 2 jam, indeks massa tubuh (BMI), fungsi silsilah diabetes (Diabetes Pedigree Function), serta usia pasien. Selain itu, label diagnosis berupa nilai biner (0 atau 1) menunjukkan apakah pasien terdeteksi mengidap diabetes atau tidak. Dataset ini memberikan gambaran komprehensif mengenai faktor-faktor klinis yang memengaruhi kemungkinan seseorang terkena diabetes, sehingga sangat relevan untuk membangun model prediksi menggunakan algoritma K-Nearest Neighbors (KNN), baik dalam baseline maupun versi yang telah dioptimasi.

3. HASIL DAN PEMBAHASAN

3.1. Pre Processing Data

Pada tahapan ini dilakukan tahapan persiapan data sebelum dilakukan pemrosesan menggunakan model. Tahapan ini meliputi *cleaning* data, serta split data atau membagi data menjadi data *training*, *validation* dan *testing*. Pada tahapan ini menghasilkan pembagian data sebagai berikut :

Tabel 2. Jumlah Dataset

No	Pembagian Data	Jumlah
1	Training	800
2	Testing	200
	Total	1000

Data pada tabel 2 merupakan data yang bersumber dari kaggle, dimana data sudah dibagi menjadi 80% data training dengan jumlah data 800 dan 20% data testing dengan jumlah data 200.

3.2. Hasil Prediksi

Pada tahap pemodelan, digunakan algoritma K-Nearest Neighbors (KNN) yang dioptimasi menggunakan Grid Search untuk meningkatkan kinerja model dalam memprediksi potensi diabetes berdasarkan data klinis pasien. Algoritma KNN dipilih karena kesederhanaannya serta kemampuannya dalam melakukan klasifikasi berdasarkan kedekatan data dalam ruang fitur. Proses optimasi dilakukan dengan mencari nilai hyperparameter K terbaik yang memberikan akurasi klasifikasi tertinggi. Berikut adalah hasil prediksi yang dihasilkan oleh model KNN yang telah dioptimasi.

Tabel .3 Hasil Prediksi

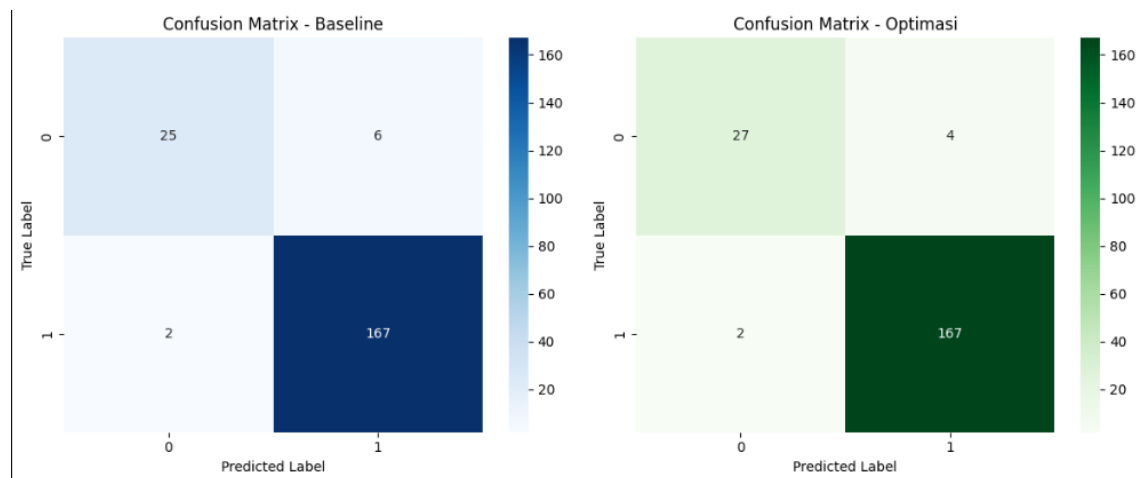
ID	Nama	Usia	HbA1c	BMI	Label Aktual	Prediksi Diabetes	Status Prediksi
1	Pasien 1	55	9.7	30.0	True	True	Benar
2	Pasien 2	50	5.0	24.0	False	False	Benar
3	Pasien 3	51	13.7	32.0	True	True	Benar
4	Pasien 4	77	7.2	29.0	True	True	Benar
5	Pasien 5	53	8.6	30.0	True	True	Benar
6	Pasien 6	49	5.0	21.0	False	False	Benar
7	Pasien 7	61	8.9	29.0	True	True	Benar
8	Pasien 8	63	9.8	37.0	True	True	Benar
9	Pasien 9	60	10.2	31.0	True	True	Benar
10	Pasien 10	52	13.2	35.0	True	True	Benar
11	Pasien 11	30	6.0	22.0	False	False	Benar

Berdasarkan hasil prediksi menggunakan algoritma K-Nearest Neighbors (KNN) terhadap data klinis 15 pasien, diperoleh bahwa model berhasil mengklasifikasikan status diabetes dengan tingkat akurasi yang tinggi. Data yang digunakan mencakup atribut seperti usia, kadar HbA1c, dan indeks massa tubuh (BMI), serta label aktual dan hasil prediksi dari model. Dari keseluruhan pasien, model mampu memberikan prediksi yang sesuai dengan kondisi aktual pasien, baik yang terindikasi maupun tidak terindikasi diabetes. Hasil ini menunjukkan bahwa model KNN yang telah dioptimasi dapat bekerja secara efektif dalam mengklasifikasikan risiko diabetes berdasarkan data klinis yang tersedia.

3.3. Evaluasi Hasil

3.3.1. Evaluasi Hasil dengan Confusion Matrix

Hasil pengujian dilakukan evaluasi untuk mengetahui nilai akurasi dari algoritma yang akan dianalisa apakah model yang dibuat layak digunakan. Berikut hasil evaluasi yang dihasilkan.



Gambar 3. Hasil Evaluasi Prediksi KNN Baseline dan KNN Optimasi

Berdasarkan visualisasi confusion matrix, terlihat adanya peningkatan kinerja model K-Nearest Neighbors setelah dilakukan optimasi. Pada model baseline, jumlah prediksi benar untuk kelas negatif (true negative) adalah 25, sementara setelah optimasi meningkat menjadi 27, sedangkan kesalahan klasifikasi terhadap kelas negatif (false positive) menurun dari 6 menjadi 4. Untuk kelas positif (true positive), jumlah prediksi benar tetap konsisten sebanyak 167, dengan kesalahan prediksi (false negative) tetap sebanyak 2. Hasil ini menunjukkan bahwa optimasi yang dilakukan mampu meningkatkan akurasi model dalam mengklasifikasikan data negatif tanpa mengorbankan performa terhadap data positif.

4. Pembahasan

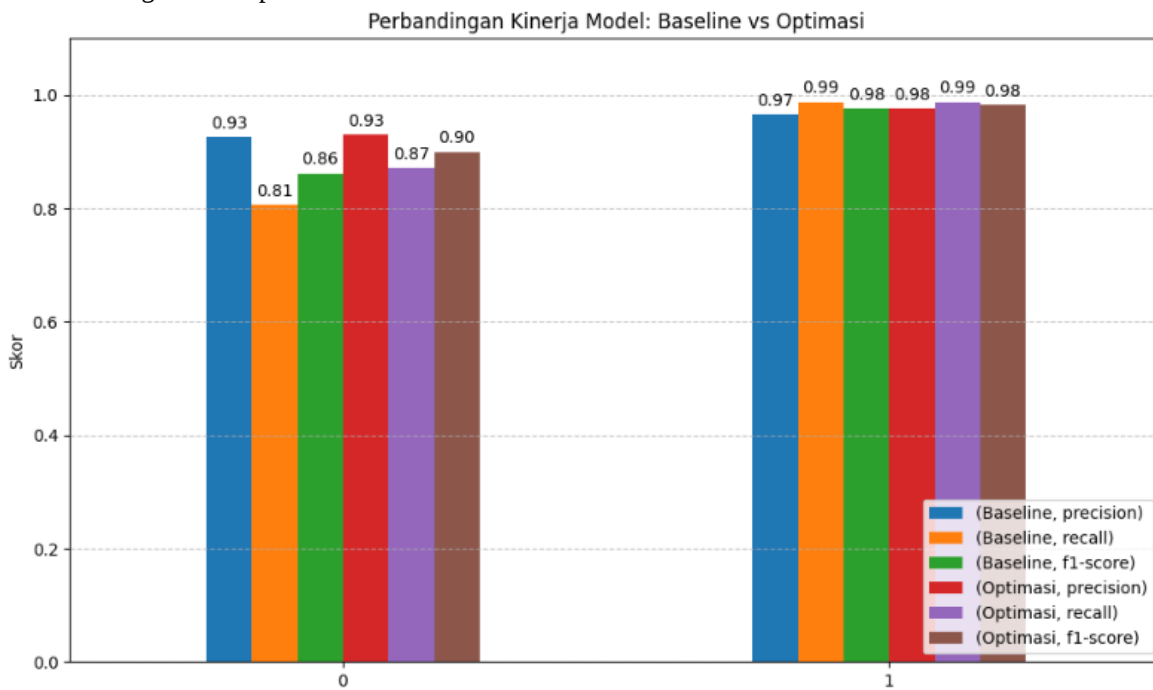
4.1. Perbandingan Hasil Prediksi

Pada sub-bab ini, dibahas mengenai perbandingan kinerja model sebelum dan sesudah dilakukan optimasi terhadap algoritma K-Nearest Neighbors dalam memprediksi klasifikasi diabetes. Evaluasi dilakukan dengan menggunakan metrik precision, recall, dan f1-score untuk masing-masing kelas.

Tabel 3. Hasil Komparasi Hasil Prediksi KNN Baseline dan KNN Optimasi

Model	Presisi	Recall	Skor F1	Akurasi
KNN Baseline	93%	81%	86%	91%
KNN Optimasi	93%	87%	90%	93%

Berikut grafik komparasi dari kedua model:



Gambar 4. Grafik Komparasi

Berdasarkan grafik, terlihat bahwa model optimasi mengalami peningkatan performa yang cukup signifikan pada kelas 0 (negatif), di mana precision meningkat dari 0,93 menjadi 0,93, recall dari 0,81 menjadi 0,87, dan f1-score dari 0,86 menjadi 0,90. Sementara itu, pada kelas 1 (positif), skor precision, recall, dan f1-score yang sudah tinggi pada model baseline tetap terjaga setelah optimasi, yaitu berada di kisaran 0,97 hingga 0,99. Hasil ini menunjukkan bahwa proses optimasi hyperparameter berhasil meningkatkan kualitas klasifikasi khususnya pada deteksi pasien non-diabetes, tanpa mengorbankan akurasi pada pasien yang terdiagnosis diabetes.

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa model K-Nearest Neighbors (KNN) setelah dilakukan optimasi hyperparameter mengalami peningkatan performa yang signifikan dibandingkan model baseline. Model baseline memiliki akurasi sebesar 90%, presisi 93%, recall 81%, dan skor F1 sebesar 86%. Setelah dilakukan optimasi, akurasi meningkat menjadi 91%, presisi tetap tinggi di angka 93%, recall naik menjadi 87%, dan skor F1 menjadi 90%. Peningkatan metrik ini menunjukkan bahwa optimasi hyperparameter memberikan dampak positif terhadap kemampuan model dalam mengklasifikasikan data klinis pasien untuk prediksi diabetes secara lebih tepat dan andal.

REFERENCES

- Abdellatif, A., Abdellatef, H., Kanesan, J., Chow, C. O., Chuah, J. H., & Ghenni, H. M. (2022). An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods. *IEEE Access*, 10(July), 79974–79985. <https://doi.org/10.1109/ACCESS.2022.3191669>
- Aghaabbasi, M., Ali, M., Jasinski, M., Leonowicz, Z., & Novak, T. (2023). On Hyperparameter Optimization of Machine Learning Methods Using a Bayesian Optimization Algorithm to Predict Work Travel Mode Choice. *IEEE Access*, 11(January), 19762–19774. <https://doi.org/10.1109/ACCESS.2023.3247448>
- Algahtani, F. D., Hassan, S. U. N., Alsaif, B., & Zrieq, R. (2021). Assessment of the quality of life during covid-19 pandemic: A cross-sectional survey from the kingdom of saudi arabia. *International Journal of Environmental*

- Research and Public Health*, 18(3), 1–12. <https://doi.org/10.3390/ijerph18030847>
- Ali, A., Hamraz, M., Gul, N., Khan, D. M., Aldahmani, S., & Khan, Z. (2023). A k nearest neighbour ensemble via extended neighbourhood rule and feature subsets. *Pattern Recognition*, 142. <https://doi.org/10.1016/j.patcog.2023.109641>
- Amini, A., Dolatshahi, M., & Kerachian, R. (2023). Effects of Automatic Hyperparameter Tuning on the Performance of Multi-Variate Deep Learning-Based Rainfall Nowcasting. *Water Resources Research*, 59(1). <https://doi.org/10.1029/2022WR032789>
- Bacanin, N., Stoean, C., Zivkovic, M., Rakic, M., Strulak-Wójcikiewicz, R., & Stoean, R. (2023). On the Benefits of Using Metaheuristics in the Hyperparameter Tuning of Deep Learning Models for Energy Load Forecasting. *Energies*, 16(3), 1–21. <https://doi.org/10.3390/en16031434>
- Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decision Analytics Journal*, 3(May), 100071. <https://doi.org/10.1016/j.dajour.2022.100071>
- Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications. *Journal of Healthcare Engineering*, 2021. <https://doi.org/10.1155/2021/9930985>
- Çetin, A. İ., & Büyüklü, A. H. (2025). A new approach to K-nearest neighbors distance metrics on sovereign country credit rating. *Kuwait Journal of Science*, 52(1). <https://doi.org/10.1016/j.kjs.2024.100324>
- Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. <https://doi.org/10.38094/jastt20165>
- Crasto, W. (2021). *Prevention of Microvascular Complications of Diabetes*. 50, 431–455.
- Daud, M., Ugliotti, F. M., & Osello, A. (2024). Comprehensive Analysis of the Use of Web-GIS for Natural Hazard Management: A Systematic Review. *Sustainability (Switzerland)*, 16(10). <https://doi.org/10.3390/su16104238>
- Dissanayake, K., & Johar, M. G. M. (2021). Comparative study on heart disease prediction using feature selection techniques on classification algorithms. *Applied Computational Intelligence and Soft Computing*, 2021. <https://doi.org/10.1155/2021/5581806>
- Elansari, T., Ouanan, M., & Bourray, H. (2023). Mixed Radial Basis Function Neural Network Training Using Genetic Algorithm. *Neural Processing Letters*, 55(8), 10569–10587. <https://doi.org/10.1007/s11063-023-11339-5>
- Francis, A., Harhay, M. N., Ong, A. C. M., Tummalapalli, S. L., Ortiz, A., Fogo, A. B., Fliser, D., Roy-Chaudhury, P., Fontana, M., Nangaku, M., Wanner, C., Malik, C., Hradsky, A., Adu, D., Bavanandan, S., Cusumano, A., Sola, L., Ulas, I., & Jha, V. (2024). Chronic kidney disease and the global public health agenda: an international consensus. *Nature Reviews Nephrology*, 20(7), 473–485. <https://doi.org/10.1038/s41581-024-00820-6>
- Fuadah, Y. N., Pramudito, M. A., & Lim, K. M. (2023). An Optimal Approach for Heart Sound Classification Using Grid Search in Hyperparameter Optimization of Machine Learning. *Bioengineering*, 10(1). <https://doi.org/10.3390/bioengineering10010045>
- Gambert, S. R. (2006). The burden of chronic disease. *Clinical Geriatrics*, 14(2), 5. <https://doi.org/10.1016/j.mayocpiqo.2023.08.005>
- Gasparetto, A., Marcuzzo, M., Zangari, A., & Albarelli, A. (2022). Survey on Text Classification Algorithms: From Text to Predictions. *Information (Switzerland)*, 13(2), 1–39. <https://doi.org/10.3390/info13020083>
- Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-024-00973-y>
- Homes, G. (2021). Long-Term Complications in Youth-Onset Type 2 Diabetes. *New England Journal of Medicine*, 385(5), 416–426. <https://doi.org/10.1056/nejmoa2100165>
- Hossain, M. J., Al-Mamun, M., & Islam, M. R. (2024). Diabetes mellitus, the fastest growing global public health concern: Early detection should be focused. *Health Science Reports*, 7(3), 5–9. <https://doi.org/10.1002/hsr2.2004>
- Hsieh, S. C. (2021). Prediction of Compressive Strength of Concrete and Rock Using an Elementary Instance-Based Learning Algorithm. *Advances in Civil Engineering*, 2021. <https://doi.org/10.1155/2021/6658932>
- Idris, M. (2020). Implementasi Data Mining Dengan Algoritma Naïve Bayes Untuk Memprediksi Tkp Kriminalitas Di Kabupaten Ponorogo. *Paper Knowledge . Toward a Media History of Documents*, 7(1), 1–33.
- Jane Osareme, Ogugua, Muridzo Muonde, Chinedu Paschal Maduka, Tolulope O Olorunsogo, & Olufunke Omotayo. (2024). Demographic shifts and healthcare: A review of aging populations and systemic challenges. *International Journal of Science and Research Archive*, 11(1), 383–395. <https://doi.org/10.30574/ijrsra.2024.11.1.0067>
- Jiang, X., & Xu, C. (2022). Deep Learning and Machine Learning with Grid Search to Predict Later Occurrence of Breast Cancer Metastasis Using Clinical Data. *Journal of Clinical Medicine*, 11(19). <https://doi.org/10.3390/jcm11195772>
- Khan, F. A., Member, S., Zeb, K., Al-rakhami, M., & Derhab, A. (2021). *Detection and Prediction of Diabetes Using Data Mining : A Comprehensive Review*. 43711–43735.
- Kiyak, E. O., & Ghasemkhani, B. (2023). High-Level K-Nearest Neighbors (HLKNN): A Supervised. *Electronics*, 12, 1–20.
- Kolawole, S. S., Kashyap, G. S., Kolawole, O. E., & Yu, M. (2025). The Future of Fall Prevention: Integrating OpenPose with Cutting-Edge ML Models. *EAI Endorsed Transactions on Pervasive Health and Technology*, 11, 1–9. <https://doi.org/10.4108/eetpht.11.9013>
- Kropp, M., Golubnitschaja, O., Mazurakova, A., Koklesova, L., Sargheini, N., Vo, T. T. K. S., de Clerck, E., Polivka, J., Potuznik, P., Polivka, J., Stetkarova, I., Kubatka, P., & Thumann, G. (2023). Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications—risks and mitigation. *EPMA Journal*, 14(1),

- 21–42. <https://doi.org/10.1007/s13167-023-00314-8>
- Muhammad, G., Naveed, S., Nadeem, L., Mahmood, T., Khan, A. R., Amin, Y., & Bahaj, S. A. O. (2023). Enhancing Prognosis Accuracy for Ischemic Cardiovascular Disease Using K Nearest Neighbor Algorithm: A Robust Approach. *IEEE Access*, 11(September), 97879–97895. <https://doi.org/10.1109/ACCESS.2023.3312046>
- Muntasir Nishat, M., Faisal, F., Jahan Ratul, I., Al-Monsur, A., Ar-Rafi, A. M., Nasrullah, S. M., Reza, M. T., & Khan, M. R. H. (2022). A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset. *Scientific Programming*, 2022(Cvd). <https://doi.org/10.1155/2022/3649406>
- Nayak, S., Bhat, M., Reddy, N. V. S., & Rao, B. A. (2022). Study of distance metrics on k - Nearest neighbor algorithm for star categorization. *Journal of Physics: Conference Series*, 2161(1). <https://doi.org/10.1088/1742-6596/2161/1/012004>
- Ogunsanya, M., Isichei, J., & Desai, S. (2023). Grid search hyperparameter tuning in additive manufacturing processes. *Manufacturing Letters*, 35, 1031–1042. <https://doi.org/10.1016/j.mfglet.2023.08.056>
- Oluchi S, Manaf R, Ismail S, Kadir H, Mahmud A, & Udeani T. (2021). Health Related Quality of Life Measurements for Diabetes: A Systematic Review. *International Journal of Environmental Research and Public Health*, 18, 1–9.
- Raiaan, M. A. K., Sakib, S., Fahad, N. M., Mamun, A. Al, Rahman, M. A., Shatabda, S., & Mukta, M. S. H. (2024). A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks. *Decision Analytics Journal*, 11(March). <https://doi.org/10.1016/j.dajour.2024.100470>
- Rastogi, R., & Bansal, M. (2023). *Diabetes prediction model using data mining techniques*. 25(November 2022).
- Salem, H., Shams, M. Y., Elzeki, O. M., Elfattah, M. A., Al-amri, J. F., & Elnazer, S. (2022). Fine-Tuning Fuzzy KNN Classifier Based on Uncertainty Membership for the Medical Diagnosis of Diabetes. *Applied Sciences (Switzerland)*, 12(3), 1–26. <https://doi.org/10.3390/app12030950>
- Shams, M. Y., Elshewey, A. M., El-kenawy, E. S. M., Ibrahim, A., Talaat, F. M., & Tarek, Z. (2024). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications*, 83(12), 35307–35334. <https://doi.org/10.1007/s11042-023-16737-4>
- Tesfai, H., Saleh, H., Al-Qutayri, M., Mohammad, M. B., Tekeste, T., Khandoker, A., & Mohammad, B. (2022). Lightweight Shufflenet Based CNN for Arrhythmia Classification. *IEEE Access*, 10(October), 111842–111854. <https://doi.org/10.1109/ACCESS.2022.3215665>
- 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), 1–11. <https://doi.org/10.1038/s41598-022-10358-x>
- Wang, A. X., Chukova, S. S., & Nguyen, B. P. (2023). Ensemble k-nearest neighbors based on centroid displacement. *Information Sciences*, 629(July 2022), 313–323. <https://doi.org/10.1016/j.ins.2023.02.004>
- Zamri, N., Pairan, M. A., Azman, W. N. A. W., Abas, S. S., Abdullah, L., Naim, S., Tarmudi, Z., & Gao, M. (2022). River quality classification using different distances in k-nearest neighbors algorithm. *Procedia Computer Science*, 204(2021), 180–186. <https://doi.org/10.1016/j.procs.2022.08.022>
- 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>
- Zhao, H., Morgenroth, J., Pearse, G., & Schindler, J. (2023). A Systematic Review of Individual Tree Crown Detection and Delineation with Convolutional Neural Networks (CNN). *Current Forestry Reports*, 9(3), 149–170. <https://doi.org/10.1007/s40725-023-00184-3>