JURNAL KOMPREHENSIF
E ISSN: 3031-0970 | Vol. 4 No. 2 Juli-Desember 2026 | pp. 217-226
Available online at: https://ejournal.edutechjaya.com/index.php/komprehensif

Perbandingan Algoritma *K-Nearest* Neighbor dan *Decision Tree* Untuk Prediksi Diabetes

Elsa Elvira Awal[1], Elfina Novalia[2], Nurhayati[3]

[1,2,3]Universitas Buana Perjuangan Karawang, Jawa Barat, Indonesia

Email: elsaelvira@ubpkarawang.ac.id[1], elfinanovalia@ubpkarawang.ac.id[2], nurhayati@ubpkarawang.ac.id[3]

ABSTRAK

Perkembangan teknologi informasi telah meningkatkan ketersediaan data dalam jumlah besar sehingga diperlukan metode klasifikasi yang mampu menghasilkan prediksi secara akurat. Salah satu penerapan klasifikasi yang penting adalah prediksi penyakit diabetes untuk membantu proses pengambilan keputusan di bidang kesehatan. Penelitian ini bertujuan membandingkan kinerja algoritma K-Nearest Neighbor (KNN) dan *Decision Tree* dalam memprediksi diabetes serta menentukan algoritma dengan performa terbaik. Dataset yang digunakan adalah Diabetes Prediction Dataset yang diperoleh dari Kaggle, terdiri atas 100.000 data, kemudian dilakukan proses preprocessing berupa penghapusan data duplikat, one-hot encoding pada atribut kategorikal, normalisasi data menggunakan StandardScaler, serta pembagian data latih dan data uji dengan rasio 80:20. Algoritma KNN dioptimalkan melalui pengujian beberapa nilai K, sedangkan *Decision Tree* dioptimalkan menggunakan GridSearchCV. Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, F1-score, dan confusion matrix. Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* memberikan performa yang lebih baik dibandingkan KNN dengan nilai accuracy sebesar 97%, sedangkan KNN memperoleh accuracy sebesar 96%. Selain itu, *Decision Tree* menghasilkan nilai recall dan F1-score yang lebih tinggi pada kelas positif diabetes sehingga lebih efektif dalam mengidentifikasi pasien yang berpotensi menderita diabetes. Berdasarkan hasil tersebut, *Decision Tree* direkomendasikan sebagai algoritma yang lebih sesuai untuk prediksi diabetes pada dataset yang digunakan.
Kata Kunci: K-Nearest Neighbor, *Decision Tree*, Prediksi Diabetes, Klasifikasi, Machine Learning.

ABSTRACT

The rapid development of information technology has led to the increasing availability of large-scale data, creating the need for classification methods capable of producing accurate predictions. One important application of classification is diabetes prediction, which can support decision-making in the healthcare sector. This study aims to compare the performance of the K-Nearest Neighbor (KNN) and Decision Tree algorithms in predicting diabetes and to determine the algorithm with the best performance. The study utilized the Diabetes Prediction Dataset obtained from Kaggle, consisting of 100,000 records. Data preprocessing included duplicate removal, one-hot encoding for categorical attributes, data normalization using StandardScaler, and an 80:20 split of training and testing data. The KNN algorithm was optimized by evaluating several K values, while the Decision Tree algorithm was optimized using GridSearchCV. Model performance was evaluated using accuracy, precision, recall, F1-score, and the confusion matrix. The experimental results show that the Decision Tree algorithm outperformed KNN, achieving an accuracy of 97%, while KNN achieved an accuracy of 96%. Furthermore, Decision Tree obtained higher recall and F1-score values for the positive diabetes class, indicating better capability in identifying patients with diabetes. Based on these findings, the Decision Tree algorithm is recommended as a more suitable classification method for diabetes prediction using the selected dataset.
Keywords: K-Nearest Neighbor, Decision Tree, Diabetes Prediction, Classification, Machine Learning.

PENDAHULUAN

Diabetes merupakan salah satu penyakit kronis yang menjadi perhatian utama dalam bidang kesehatan karena prevalensinya terus meningkat di berbagai negara (Silva, et al., 2020). Penyakit ini ditandai dengan tingginya kadar glukosa dalam darah akibat gangguan produksi atau kerja hormon insulin. Apabila tidak ditangani secara tepat, diabetes dapat menyebabkan berbagai komplikasi serius, seperti penyakit jantung, stroke, gagal ginjal, gangguan penglihatan, hingga kerusakan saraf. Oleh karena itu, deteksi dini terhadap risiko diabetes menjadi langkah penting dalam membantu tenaga medis melakukan penanganan lebih awal sehingga dapat menurunkan risiko komplikasi dan meningkatkan kualitas hidup pasien (Sadiq, et al., 2025).

Perkembangan teknologi informasi dan digitalisasi layanan kesehatan telah menghasilkan data medis dalam jumlah yang sangat besar. Kondisi tersebut membuka peluang penerapan machine learning sebagai salah satu pendekatan untuk membantu proses prediksi penyakit berdasarkan pola yang terdapat pada data (Sadiq, et al., 2025). Berbagai algoritma klasifikasi telah digunakan dalam prediksi diabetes, seperti K-Nearest Neighbor (KNN), *Decision Tree*, Support Vector Machine, Random Forest, hingga Artificial Neural Network (Hameed, *et al.*, 2025). Masing-masing algoritma memiliki karakteristik, kelebihan, dan kelemahan yang berbeda sehingga menghasilkan tingkat akurasi yang bervariasi bergantung pada karakteristik dataset yang digunakan (Alzboon, *et al.*, 2025).

Penelitian mengenai prediksi diabetes menggunakan algoritma machine learning telah banyak dilakukan dalam beberapa tahun terakhir. Yunianto & Subhiyakto (2025) membandingkan algoritma KNN dan *Decision Tree* menggunakan Pima Indians Diabetes Database yang telah diseimbangkan menggunakan teknik Synthetic Minority Oversampling Technique (SMOTE). Hasil penelitian menunjukkan bahwa kualitas preprocessing dan penanganan ketidakseimbangan data berpengaruh terhadap performa model klasifikasi. Penelitian lain oleh Fatah & Anam (2026) membandingkan KNN dan *Decision Tree* menggunakan perangkat lunak RapidMiner pada dataset Pima Indians Diabetes serta mengevaluasi model menggunakan K-Fold Cross Validation. Hasil penelitian tersebut menunjukkan bahwa kedua algoritma mampu memberikan performa yang baik, namun hasil evaluasi dipengaruhi oleh metode validasi dan karakteristik data yang digunakan. Sementara itu, Sadiq, *et al.*, (2025) membandingkan *Decision Tree* dengan Artificial Neural Network dan memperoleh hasil bahwa *Decision Tree* memberikan akurasi, precision, recall, dan F1-score yang lebih tinggi pada dataset yang digunakan.

Meskipun berbagai penelitian sebelumnya telah membahas prediksi diabetes menggunakan algoritma klasifikasi, sebagian besar penelitian masih menggunakan Pima Indians Diabetes Dataset yang berukuran relatif kecil, yaitu sekitar 768 data, serta lebih banyak berfokus pada perbandingan beberapa algoritma atau penggunaan teknik penyeimbangan data tertentu (Yunianto & Subhiyakto, 2025). Selain itu, hasil penelitian yang diperoleh masih menunjukkan variasi performa karena dipengaruhi

oleh karakteristik dataset, tahapan preprocessing, serta metode optimasi yang diterapkan (Fatah & Anam, 2026). Kondisi tersebut menunjukkan bahwa belum terdapat satu algoritma yang secara konsisten memberikan performa terbaik pada seluruh dataset sehingga diperlukan evaluasi lebih lanjut menggunakan dataset yang lebih besar dan lebih representatif (Sadiq, et al., 2025).

Berdasarkan kondisi tersebut, penelitian ini melakukan perbandingan kinerja algoritma KNN dan *Decision Tree* menggunakan Diabetes Prediction Dataset yang diperoleh dari Kaggle dengan jumlah data yang jauh lebih besar dibandingkan dataset yang umum digunakan pada penelitian sebelumnya. Penelitian ini menerapkan tahapan preprocessing yang meliputi penghapusan data duplikat, one-hot encoding pada atribut kategorikal, normalisasi menggunakan StandardScaler untuk algoritma KNN, serta optimasi parameter menggunakan pencarian nilai K pada KNN dan GridSearchCV pada *Decision Tree*. Evaluasi model dilakukan menggunakan accuracy, precision, recall, F1-score, dan confusion matrix sehingga mampu memberikan gambaran performa model secara lebih komprehensif.

Kebaruan penelitian ini terletak pada penggunaan Diabetes Prediction Dataset berukuran besar yang memuat karakteristik pasien lebih beragam dibandingkan dataset Pima Indians yang banyak digunakan pada penelitian sebelumnya, serta penerapan proses optimasi parameter pada kedua algoritma sebelum dilakukan evaluasi. Pendekatan tersebut diharapkan menghasilkan perbandingan performa yang lebih objektif sehingga dapat memberikan rekomendasi algoritma klasifikasi yang paling sesuai untuk prediksi diabetes pada dataset modern.

Berdasarkan uraian tersebut, penelitian ini bertujuan membandingkan performa algoritma KNN dan *Decision Tree* dalam melakukan prediksi diabetes serta menentukan algoritma yang memberikan hasil terbaik berdasarkan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian diharapkan dapat menjadi referensi dalam pemilihan algoritma machine learning yang tepat untuk mendukung sistem prediksi diabetes dan penelitian lanjutan di bidang informatika kesehatan.

METODE PENELITIAN

Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen (experimental research) yang bertujuan membandingkan kinerja algoritma K-Nearest Neighbor (KNN) dan *Decision Tree* dalam melakukan prediksi diabetes. Pendekatan eksperimen digunakan untuk mengevaluasi performa kedua algoritma pada dataset yang sama sehingga hasil yang diperoleh dapat dibandingkan secara objektif berdasarkan metrik evaluasi yang telah ditentukan.

Penelitian dilaksanakan pada bulan Juni hingga Juli 2026. Seluruh proses pengolahan data, pembangunan model, dan evaluasi dilakukan secara daring menggunakan platform Google Colaboratory (Google Colab) menggunakan bahasa pemrograman Python.

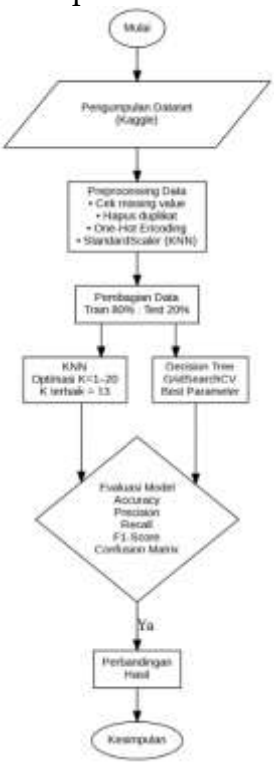
Data Penelitian

Penelitian ini menggunakan data sekunder berupa Diabetes Prediction Dataset yang diperoleh dari platform Kaggle. Dataset awal terdiri atas 100.000 data dengan sembilan atribut, yaitu gender, age, hypertension, heart_disease, smoking_history, bmi, HbA1c_level, blood_glucose_level, dan diabetes sebagai variabel target.

Sebelum dilakukan proses pembangunan model, dataset melalui tahap pembersihan data sehingga diperoleh sebanyak 96.146 data setelah penghapusan data duplikat. Dataset yang telah melalui proses preprocessing selanjutnya digunakan sebagai data penelitian.

Prosedur Penelitian

Penelitian dilakukan melalui beberapa tahapan yang disusun secara sistematis mulai dari pengumpulan dataset hingga penarikan kesimpulan. Setiap tahapan dirancang untuk memastikan bahwa proses pengolahan data dan pembangunan model dilakukan secara terstruktur sehingga menghasilkan model klasifikasi yang optimal. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1 Tahapan Penelitian

Setelah dataset diperoleh dari Kaggle, dilakukan tahap preprocessing yang meliputi pemeriksaan struktur data, identifikasi missing value, penghapusan data duplikat, transformasi atribut kategorikal menggunakan One-Hot Encoding, serta normalisasi menggunakan StandardScaler (Saxena, *et al.*, 2022). Tahap normalisasi hanya diterapkan pada algoritma KNN karena algoritma tersebut menggunakan perhitungan jarak antar data, sedangkan *Decision Tree* tidak dipengaruhi oleh perbedaan skala atribut.

Dataset yang telah diproses kemudian dibagi menjadi data latih dan data uji menggunakan metode train-test split dengan rasio 80% : 20%. Pembagian data dilakukan menggunakan stratified sampling untuk mempertahankan proporsi masing-masing kelas sehingga distribusi data latih dan data uji tetap seimbang.

Tahap berikutnya adalah pembangunan model menggunakan algoritma KNN dan *Decision Tree*. Pada algoritma KNN dilakukan proses optimasi dengan menguji nilai K dari 1 hingga 20 sehingga diperoleh nilai K = 13 sebagai parameter terbaik. Sementara itu, algoritma *Decision Tree* dioptimalkan menggunakan GridSearchCV dengan beberapa kombinasi parameter criterion, max_depth, dan min_samples_split. Hasil optimasi menunjukkan bahwa parameter terbaik adalah criterion = gini, max_depth = 3, dan min_samples_split = 2.

Teknik Analisis Data

Evaluasi performa model dilakukan menggunakan data uji yang tidak digunakan pada proses pelatihan model (Aparicio, *et al.*, 2021). Kinerja masing-masing algoritma dianalisis menggunakan beberapa metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score yang diperoleh melalui classification report (Silva, et al., 2020). Selain itu, digunakan confusion matrix untuk memberikan gambaran mengenai jumlah prediksi yang benar maupun salah pada masing-masing kelas.

Nilai accuracy digunakan untuk mengukur tingkat ketepatan model secara keseluruhan. Precision menunjukkan kemampuan model dalam memberikan prediksi positif yang benar, sedangkan recall mengukur kemampuan model dalam mendeteksi seluruh data yang benar-benar termasuk kelas positif. F1-score digunakan sebagai ukuran keseimbangan antara nilai precision dan recall. Seluruh hasil evaluasi dari algoritma KNN dan *Decision Tree* kemudian dibandingkan untuk menentukan algoritma yang memiliki performa terbaik dalam melakukan prediksi diabetes.

HASIL DAN PEMBAHASAN

Hasil Preprocessing Data

Tahap awal penelitian adalah melakukan preprocessing terhadap Diabetes Prediction Dataset. Dataset awal terdiri atas 100.000 data dengan sembilan atribut, yaitu gender, age, hypertension, heart_disease, smoking_history, bmi, HbA1c_level, blood_glucose_level, dan diabetes sebagai variabel target. Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki missing value, sehingga proses pembersihan data difokuskan pada penghapusan data duplikat.

Setelah dilakukan proses penghapusan data duplikat, jumlah data berkurang menjadi 96.146 data. Selanjutnya dilakukan transformasi atribut kategorikal menggunakan One-Hot Encoding. Khusus untuk algoritma KNN dilakukan normalisasi menggunakan StandardScaler agar seluruh atribut numerik memiliki skala yang sebanding sebelum proses klasifikasi.

Tabel 1 Hasil Preprocessing Dataset

Tahapan	Jumlah Data
Dataset awal	100.000
Missing value	0
Data setelah menghapus duplikat	96.146
Jumlah atribut	9

Berdasarkan Tabel 1, dapat diketahui bahwa kualitas dataset tergolong baik karena tidak ditemukan data yang hilang (missing value). Penghapusan data duplikat menghasilkan dataset yang lebih representatif sehingga dapat meningkatkan kualitas proses pelatihan model.

Pembagian Data

Dataset yang telah melalui proses preprocessing selanjutnya dibagi menjadi data latih dan data uji menggunakan rasio 80:20. Pembagian dilakukan menggunakan stratified sampling sehingga proporsi masing-masing kelas tetap terjaga.

Tabel 2 Pembagian Dataset

Dataset	Jumlah Data
Data Latih	76.916
Data Uji	19.230
Total	96.146

Pembagian data tersebut dilakukan agar model dapat mempelajari pola pada data latih, kemudian diuji menggunakan data yang belum pernah dikenali sebelumnya sehingga hasil evaluasi menjadi lebih objektif.

Hasil Pengujian Algoritma K-Nearest Neighbor

Pada penelitian ini dilakukan pengujian nilai K mulai dari 1 hingga 20. Tujuannya adalah memperoleh nilai K yang menghasilkan performa terbaik.

Tabel 3 Hasil Pengujian Nilai K

Nilai K	Accuracy
1	…
2	…
…	…
13	0,9608
…	…
20	…

Berdasarkan hasil pengujian, nilai K = 13 memberikan akurasi tertinggi sebesar 96,08%, sehingga digunakan sebagai parameter terbaik pada proses evaluasi akhir. Selanjutnya dilakukan evaluasi model menggunakan classification report.

Tabel 4 Hasil Evaluasi KNN

Metrik	Kelas 0	Kelas 1
Precision	0,96	0,95
Recall	1,00	0,58
F1-score	0,98	0,72

Berdasarkan hasil evaluasi, algoritma K-Nearest Neighbor (KNN) memperoleh nilai akurasi sebesar 96% dalam mengklasifikasikan data diabetes. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam melakukan klasifikasi secara keseluruhan. Meskipun demikian, berdasarkan nilai classification report terlihat bahwa recall pada kelas positif diabetes hanya mencapai 58%, yang menunjukkan bahwa masih terdapat sebagian pasien diabetes yang belum berhasil diidentifikasi dengan benar oleh model. Kondisi ini menyebabkan nilai F1-score pada kelas positif hanya sebesar 72%, sehingga kemampuan KNN dalam mendeteksi kasus diabetes masih perlu ditingkatkan.

Hasil Pengujian Algoritma *Decision Tree*

Optimasi parameter *Decision Tree* dilakukan menggunakan GridSearchCV.

Tabel 5 Parameter Terbaik *Decision Tree*

Parameter	Nilai Terbaik
Criterion	Gini
Max Depth	3
Min Samples Split	2
Cross Validation Accuracy	97,06%

Selanjutnya dilakukan evaluasi menggunakan data uji.

Tabel 6 Hasil Evaluasi *Decision Tree*

Metrik	Kelas 0	Kelas 1
Precision	0,97	1,00
Recall	1,00	0,68
F1-score	0,98	0,81

Berdasarkan hasil evaluasi, algoritma *Decision Tree* memperoleh nilai akurasi sebesar 97%, lebih tinggi dibandingkan algoritma KNN yang memperoleh akurasi sebesar 96%. Selain memiliki tingkat akurasi yang lebih baik, *Decision Tree* juga menghasilkan recall sebesar 68% dan F1-score sebesar 81% pada kelas positif diabetes. Hasil tersebut menunjukkan bahwa algoritma *Decision Tree* lebih efektif dalam mengidentifikasi pasien yang benar-benar menderita diabetes, sehingga memberikan performa klasifikasi yang lebih baik pada dataset yang digunakan.

Perbandingan Kinerja Algoritma

Tabel 7 Perbandingan Performa Algoritma

Metrik	KNN	*Decision Tree*
Accuracy	96%	97%

Precision	95%	100%
Recall	58%	68%
F1-score	72%	81%

Tabel 7 menunjukkan bahwa seluruh metrik evaluasi utama lebih unggul pada algoritma *Decision Tree*. Perbedaan paling terlihat terdapat pada nilai recall, yang menunjukkan kemampuan model dalam mendeteksi pasien diabetes secara benar.

Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* memiliki performa yang lebih baik dibandingkan algoritma KNN pada Diabetes Prediction Dataset (Hameed*et al.,* 2025). Hal tersebut ditunjukkan oleh nilai accuracy sebesar 97%, sedangkan KNN memperoleh 96%. Meskipun selisih akurasi hanya sekitar satu persen, *Decision Tree* menghasilkan peningkatan yang lebih signifikan pada nilai recall kelas positif, yaitu 68%, dibandingkan KNN yang hanya mencapai 58%. Kondisi ini menunjukkan bahwa *Decision Tree* lebih efektif dalam mengidentifikasi pasien yang benar-benar menderita diabetes.

Perbedaan performa tersebut dipengaruhi oleh karakteristik kedua algoritma. KNN merupakan algoritma berbasis jarak (distance-based algorithm) sehingga sangat dipengaruhi oleh distribusi data serta kedekatan antar sampel. Sebaliknya, *Decision Tree* membangun struktur pohon keputusan berdasarkan atribut yang paling informatif sehingga lebih mampu memisahkan data yang memiliki karakteristik berbeda tanpa dipengaruhi oleh skala atribut. Oleh karena itu, pada dataset yang memiliki kombinasi atribut numerik dan kategorikal seperti penelitian ini, *Decision Tree* mampu menghasilkan model klasifikasi yang lebih stabil.

Hasil penelitian ini sejalan dengan penelitian Fatah & Anam (2026) yang melaporkan bahwa *Decision Tree* menghasilkan performa klasifikasi yang kompetitif dibandingkan KNN pada prediksi diabetes. Temuan ini juga mendukung penelitian Sadiq, *et al.,* (2025) yang menunjukkan bahwa *Decision Tree* mampu memberikan nilai precision, recall, dan F1-score yang lebih tinggi dibandingkan algoritma pembanding pada dataset diabetes. Dengan demikian, hasil penelitian ini memperkuat temuan sebelumnya bahwa *Decision Tree* merupakan salah satu algoritma yang efektif untuk klasifikasi diabetes.

Kontribusi penelitian ini terletak pada penggunaan Diabetes Prediction Dataset berukuran besar yang terdiri atas lebih dari 96 ribu data setelah proses pembersihan, sehingga mampu memberikan gambaran performa algoritma pada dataset yang lebih representatif dibandingkan penelitian sebelumnya yang umumnya masih menggunakan Pima Indians Diabetes Dataset. Selain itu, optimasi parameter pada kedua algoritma menghasilkan proses perbandingan yang lebih objektif sehingga rekomendasi algoritma didasarkan pada konfigurasi terbaik masing-masing model.

Penelitian ini masih memiliki beberapa keterbatasan. Algoritma yang dibandingkan hanya terdiri atas KNN dan *Decision Tree*, sehingga belum dapat menggambarkan performa algoritma lain seperti Random Forest, Support Vector

Machine, maupun Extreme Gradient Boosting. Selain itu, penelitian ini menggunakan satu dataset sehingga hasil yang diperoleh belum tentu dapat digeneralisasikan pada seluruh data kesehatan. Penelitian selanjutnya dapat mengembangkan perbandingan dengan algoritma lain, menerapkan teknik penyeimbangan data, melakukan feature selection, maupun menggunakan beberapa dataset agar diperoleh model prediksi diabetes yang lebih robust.

KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma *Decision Tree* memiliki performa yang lebih baik dibandingkan K-Nearest Neighbor (KNN) dalam melakukan prediksi diabetes pada Diabetes Prediction Dataset. Hal ini ditunjukkan oleh nilai accuracy sebesar 97%, yang lebih tinggi dibandingkan KNN sebesar 96%, serta nilai recall dan F1-score yang lebih baik pada kelas positif diabetes sehingga lebih efektif dalam mengidentifikasi pasien yang berpotensi menderita diabetes. Dengan demikian, tujuan penelitian untuk membandingkan performa kedua algoritma dan menentukan algoritma yang memberikan hasil terbaik telah tercapai, yaitu *Decision Tree*. Temuan ini menunjukkan bahwa *Decision Tree* dapat dipertimbangkan sebagai algoritma yang lebih sesuai untuk pengembangan sistem prediksi diabetes berbasis machine learning. Penelitian selanjutnya disarankan untuk membandingkan algoritma ini dengan metode klasifikasi lainnya, seperti Random Forest, Support Vector Machine, atau Extreme Gradient Boosting, serta menerapkan teknik feature selection atau penyeimbangan data agar diperoleh model prediksi yang lebih akurat dan memiliki kemampuan generalisasi yang lebih baik.

DAFTAR PUSTAKA

ALZBOON, M. S., AL-BATAH, M., ALQARALEH, M., ABUASHOUR, A., & BADER, A. F. (2025). A COMPARATIVE STUDY OF MACHINE LEARNING TECHNIQUES FOR EARLY PREDICTION OF DIABETES. *ARXIV*.

Aparicio, L. F., Noguez, J., Montesinos, L., & García, J. A. (2021). Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. *Diabetology & Metabolic Syndrome*.

Fatah, Z., & Anam, B. (2026). Application of KNN & *Decision Tree* Algorithms in Predicting Diabetes Using RapidMiner. *Jurnal Ilmiah Sistem Informasi*, 229–241.

Hameed, E. M., Joshi, H., & Kadhim, Q. K. (2025). Advancements in Artificial Intelligence Techniques for Diabetes Prediction: A Comprehensive Literature Review. *Journal of Robotics and Control*.

Sadiq, I. Z., Katsayal, B. S., Ibrahim, B., Ibrahim, M., Hassan, H. A., Ghali, U. M., . . . Abba, S. I. (2025). Data-driven diabetes mellitus prediction and management: a comparative evaluation of *Decision Tree* classifier and artificial neural network models along with statistical analysis. *Scientific Reports, 15*.

Saxena, R., Sharma, S. K., Gupta, M., & Sampada, G. C. (2022). A Comprehensive Review of Various Diabetic Prediction Models: A Literature Survey. *Journal of Healthcare Engineering*.

Silva, K. D., Lee, W. K., Forbes, A., Demmer, R. T., Barton, C., & Enticott, J. (2020). Use and performance of machine learning models for type 2 diabetes prediction in community settings: A systematic review and meta-analysis. *International Journal of Medical Informatics*.

Yunianto, A. H., & Subhiyakto, E. R. (2025). Perbandingan Kinerja Algoritma K-Nearest Neighbors dan *Decision Tree* untuk Klasifikasi Diabetes. *Building of Informatics, Technology and Science*, 2601–2611.