

DETEKSI RISIKO DIABETES BERDASARKAN FAKTOR KESEHATAN MENGUNAKAN LONG SHORT-TERM MEMORY (LSTM) DAN XGBOOST

DIABETES RISK DETECTION BASED ON HEALTH FACTORS USING LONG SHORT-TERM MEMORY (LSTM) AND XGBOOST

Siti Inayatul Mufidah¹, Irma Handayani²

Universitas Teknologi Yogyakarta^{1,2}

inayatulmufidah48@gmail.com¹

ABSTRACT

Diabetes melitus is a chronic disease that has become a global health concern due to its potential to cause severe complications such as kidney failure, heart disease, and death. This study aims to develop a diabetes risk detection system based on a hybrid model combining Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost). The dataset used in this research was obtained from Kaggle and consists of nine main health-related features. The research process includes data preprocessing using SMOTE, LSTM model training for feature extraction, and classification using XGBoost. The experimental results show that the hybrid LSTM-XGBoost model achieved an accuracy of 96,24% and an ROC-AUC of 97,46%. The model was successfully implemented in a web-based system using Flask, allowing users to perform diabetes risk detection quickly, efficiently, and interactively.

Keywords: Diabetes, LSTM, XGBoost, Deep Learning, Detection.

ABSTRAK

Diabetes melitus merupakan penyakit kronis yang menjadi masalah kesehatan global karena dapat menyebabkan komplikasi serius seperti gagal ginjal, penyakit jantung, dan kematian. Penelitian ini bertujuan untuk mengembangkan sistem deteksi risiko diabetes berbasis kombinasi Long Short-Term Memory (LSTM) dan Extreme Gradient Boosting (XGBoost). Dataset yang digunakan berasal dari Kaggle dengan sembilan fitur kesehatan utama. Proses penelitian meliputi preprocessing data menggunakan SMOTE, pelatihan model LSTM untuk ekstraksi fitur, serta klasifikasi menggunakan XGBoost. Hasil pengujian menunjukkan model hybrid LSTM-XGBoost memiliki akurasi sebesar 96,24% dan ROC-AUC 97,46%. Model ini juga berhasil diimplementasikan dalam sistem berbasis web menggunakan Flask yang memudahkan pengguna melakukan deteksi risiko diabetes secara cepat, efisien, dan interaktif.

Kata Kunci: Diabetes, LSTM, XGBoost, Deep Learning, Deteksi.

PENDAHULUAN

Diabetes melitus merupakan salah satu penyakit kronis yang menjadi permasalahan kesehatan global. Menurut World Health Organization (WHO), (2024) jumlah penderita diabetes meningkat dari 200 juta pada tahun 1990 menjadi 830 juta pada tahun 2022. Penyakit ini dapat menyebabkan komplikasi serius seperti kebutaan, gagal ginjal, stroke, dan penyakit jantung, bahkan dapat berujung pada kematian. Oleh karena itu, deteksi dini terhadap risiko diabetes sangat penting agar tindakan pencegahan dapat dilakukan sejak dini untuk mengurangi dampak negatif yang ditimbulkan.

Namun, banyak individu yang tidak menyadari risiko diabetes yang dihadapi hingga kondisi sudah memburuk. Hal ini disebabkan oleh keterbatasan akses terhadap pemeriksaan medis rutin, kurangnya pemahaman terhadap faktor risiko, serta keterbatasan sumber daya kesehatan di beberapa wilayah. Oleh karena itu, diperlukan solusi inovatif yang dapat membantu tenaga medis dalam melakukan deteksi dini risiko diabetes secara cepat, akurat, dan efisien, terutama di lingkungan pelayanan kesehatan dengan fasilitas terbatas.

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*) memberikan peluang baru dalam bidang kesehatan,

khususnya melalui penerapan *machine learning* dan *deep learning*. Salah satu metode *deep learning* yang potensial adalah *Long Short-Term Memory* (LSTM), yaitu bagian dari *Recurrent Neural Network* (RNN) yang dirancang untuk mengenali pola jangka panjang dalam data sekuensial. Beberapa penelitian terdahulu telah menunjukkan keunggulan metode ini. Penelitian oleh (Chowdary & Kumar, 2021) menunjukkan bahwa LSTM merupakan metode yang lebih unggul dalam deteksi diabetes, dengan akurasi 95,6%, lebih tinggi dibandingkan algoritma *machine learning*. Hasil serupa diperoleh oleh (El-Attar et al., 2022), yang menunjukkan akurasi sebesar 95%. Dalam mendeteksi perbedaan antara DNA normal dan bermutasi yang berhubungan dengan risiko diabetes. Selain itu, penelitian oleh (Panigrahy et al., 2024) menunjukkan bahwa LSTM yang dioptimalkan dengan SMOTE dan GridSearchCV mencapai akurasi 97,12%, menjadikannya metode yang sangat efektif untuk diagnosis diabetes berbasis data kesehatan.

Selain itu, algoritma *Extreme Gradient Boosting* (XGBoost) juga terbukti unggul dalam klasifikasi dan deteksi penyakit karena kemampuannya dalam menangani data berskala besar secara efisien. Kombinasi antara LSTM dan XGBoost memungkinkan terciptanya model hibrid yang dapat memanfaatkan keunggulan masing-masing algoritma. LSTM dalam mengenali pola temporal dan XGBoost dalam meningkatkan akurasi klasifikasi.

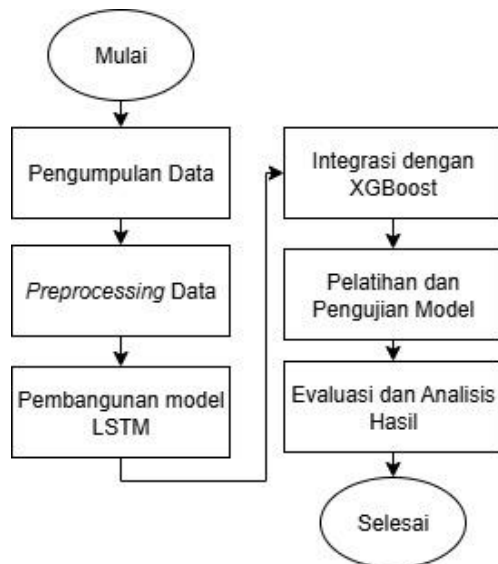
Beberapa penelitian sebelumnya juga mendukung efektivitas penggunaan algoritma XGBoost dalam mendeteksi penyakit berbasis data kesehatan. Menurut (Abdurrosyid & Almais, 2025), metode XGBoost mampu memberikan hasil klasifikasi yang lebih akurat dalam deteksi dini diabetes dibandingkan metode tradisional lainnya. Sementara itu, (Cahyana et al., 2025) menunjukkan kemampuan XGBoost dalam klasifikasi yang sangat unggul dalam deteksi penyakit

kronis seperti diabetes. Dengan akurasi mencapai 96,33% dan nilai ROC-AUC sebesar 0,99. Penelitian lain oleh (Febrianti et al., 2025), algoritma XGBoost mampu mendeteksi risiko diabetes dengan akurasi mencapai 98,7% dan F1-score 99%, menjadikannya salah satu metode paling andal dalam klasifikasi penyakit berbasis data kesehatan.

Berdasarkan hasil-hasil penelitian tersebut, penggunaan kombinasi LSTM dan XGBoost memiliki dasar ilmiah yang kuat untuk pengembangan sistem deteksi risiko diabetes yang lebih akurat dan efisien. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model deteksi risiko diabetes berdasarkan berbagai faktor kesehatan, meliputi usia, jenis kelamin, indeks massa tubuh, hipertensi, penyakit jantung, riwayat merokok, kadar HbA1c, dan kadar glukosa darah. Sistem yang dikembangkan diharapkan dapat membantu tenaga medis dalam melakukan pemeriksaan risiko diabetes secara cepat dan efisien, mendukung proses diagnosis awal, serta meningkatkan efektivitas pelayanan kesehatan di tingkat fasilitas dasar.

METODE

Pada bagian ini dijelaskan tahapan-tahapan yang dilakukan dalam penelitian, mulai dari pengumpulan data, preprocessing, pembangunan model, hingga tahap evaluasi. Penelitian ini menerapkan pendekatan kuantitatif dengan metode eksperimental untuk menguji performa model hibrid yang menggabungkan algoritma *Long Short-Term Memory* (LSTM) dan *Extreme Gradient Boosting* (XGBoost) dalam mendeteksi risiko diabetes. Tahapan penelitian secara umum dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar1, penelitian diawali dengan pengumpulan data dari dataset publik, kemudian dilakukan tahap *preprocessing* untuk pembersihan data, selanjutnya dilakukan Pembangunan model LSTM untuk menghasilkan representasi fitur. Representasi tersebut diintegrasikan dengan algoritma XGBoost yang berfungsi sebagai pengklasifikasi utama. Model hasil integrasi kemudian menjalani tahap pelatihan dan pengujian untuk memperoleh performa terbaik. Tahap terakhir adalah evaluasi dan analisis hasil.

Dataset

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari Kaggle dengan nama *Diabetes Prediction Dataset*. Dataset ini berisi 100.000 data dan beberapa atribut yang menjadi faktor risiko diabetes. Variabel yang digunakan dalam penelitian ini terdiri dari delapan fitur utama, yaitu usia, jenis kelamin, indeks massa tubuh (BMI), hipertensi, penyakit jantung, riwayat merokok, kadar HbA1c, dan kadar glukosa darah, sedangkan label target adalah diabetes dengan nilai biner (0=tidak diabetes, 1=diabetes).

Preprocessing Data

Preprocessing data merupakan tahapan untuk mengolah data mentah agar menjadi format yang lebih mudah dipahami

dan siap digunakan oleh model (Swastika et al., 2023). Tahapan preprocessing meliputi:

1. Pembersihan dan Transformasi Data

Variabel kategorikal seperti gender dan smoking_history diubah ke bentuk numerik menggunakan LabelEncoder agar dapat diproses oleh model. Nilai yang hilang diatasi dengan metode imputasi, dan data duplikat dihapus.

2. Penyeimbangan Data dengan SMOTE

SMOTE merupakan metode yang digunakan untuk menangani ketidakseimbangan data dengan menyeimbangkan jumlah sampel antara kelas mayoritas dan minoritas (Yulianti et al., 2023). Data latih sering kali memiliki distribusi kelas yang tidak seimbang antara pasien diabetes dan non-diabetes, sehingga digunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menambah data sintetis pada kelas minoritas sehingga distribusi data menjadi seimbang dan model tidak bias.

3. Standardisasi Fitur

Semua fitur numerik dinormalisasi menggunakan *StandardScaler* agar setiap atribut memiliki skala nilai yang seragam.

4. Pembagian Data

Dataset dibagi menjadi data latih (80%) dan data uji (20%) untuk memastikan hasil evaluasi yang objektif.

Setelah preprocessing, data diubah ke format tiga dimensi menggunakan fungsi *reshape* agar kompatibel dengan arsitektur LSTM.

Long Short Term-Memory (LSTM)

LSTM merupakan pengembangan dari *Recurrent Neural Network* (RNN) yang mampu menyimpan informasi jangka Panjang dan menghindari masalah vanishing gradient. Dengan kemampuannya dalam mengingat dan memproses informasi dari waktu ke waktu, LSTM efektif dalam mempertahankan data penting dari masa lampau serta mengabaikan data yang tidak relevan, sehingga dapat mengolah pola data yang

bersifat temporal secara lebih akurat (Kwanda et al., 2024).

Model LSTM pada penelitian ini dibangun menggunakan *Sequential API* dari TensorFlow dengan arsitektur pada Tabel 1.

Tabel 1. Arsitektur Model LSTM

Lapisan	Jumlah Neuron	Fungsi Aktivasi
Input Layer	-	-
LSTM Layer	128	<i>tanh</i>
Dropout Layer	-	-
Dense Layer	64	<i>ReLU</i>
Output Layer	1	Sigmoid

Model dilatih menggunakan *optimizer Adam* dengan *loss function binary cross-entropy*. Proses pelatihan berjalan selama maksimal 30 epoch dengan batch size 128, serta menerapkan Teknik *EarlyStopping* dengan *patience* = 3 untuk menghentikan pelatihan lebih awal jika tidak terjadi peningkatan pada validation loss.

XGBoost

XGBoost adalah algoritma *gradient boosting* yang memiliki performa tinggi dan efektif digunakan untuk menyelesaikan permasalahan klasifikasi (Firdaus et al., 2025). Algoritma XGBoost digunakan sebagai model klasifikasi utama setelah proses ekstraksi fitur oleh LSTM. Model XGBoost dilatih menggunakan fitur hasil ekstraksi dari LSTM dengan parameter utama pada Tabel 2.

Tabel 2. Parameter Model XGBoost

Parameter	Nilai	Keterangan
n_estimators	300	Jumlah pohon keputusan yang digunakan.
learning_rate	0.001	Kecepatan pembelajaran yang mengontrol seberapa besar pembaruan bobot dilakukan pada setiap iterasi.
max_depth	5	Kedalaman maksimum setiap pohon untuk mengontrol kompleksitas model dan mencegah <i>overfitting</i> .
subsample	0,8	Proporsi fitur yang digunakan tiap pohon.
colsample_by_tree	0,8	Presentase data yang digunakan tiap iterasi.

eval_metric	logloss	Fungsi evaluasi yang digunakan untuk menghitung kesalahan logaritmik selama pelatihan model.
-------------	---------	--

Hasil dari XGBoost berupa probabilitas risiko diabetes (antara 0-1) yang kemudian dikategorikan menjadi “Terdeteksi Risiko Diabetes” dan “Tidak Terdeteksi Risiko Diabetes”.

Evaluasi Model

Model diuji menggunakan beberapa metrik evaluasi yang meliputi akurasi, presisi, recall, F1-score, ROC-AUC, dan *confusion matrix*.

Akurasi

Akurasi merupakan ukuran performa keseluruhan model prediksi dengan menilai seberapa benar prediksi (Gunawan et al., 2025). Semakin tinggi nilai akurasi, semakin baik model dalam melakukan klasifikasi secara keseluruhan.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Dimana TP adalah jumlah data positif yang terdeteksi benar, TN adalah jumlah data negatif yang terdeteksi benar, FP adalah data negatif yang salah terdeteksi positif, dan FN adalah data positif yang salah terdeteksi negatif.

Presisi

Presisi adalah jumlah prediksi positif yang benar-benar relevan (Aditya et al., 2024). Metrik ini penting untuk memastikan bahwa hasil positif dari sistem benar-benar valid.

$$Precision = \frac{TP}{TP + FP}$$

Recall

Recall adalah rasion prediksi benar positif dibagi dengan semua data yang benar positif (Adji Kusuma et al., 2025).

$$Recall = \frac{TP}{TP + FN}$$

F1-Score

F1-Score merupakan rata-rata harmonis dari recall dan precision (Putri et al., 2025). Nilai F1-Score yang tinggi menunjukkan bahwa model memiliki kinerja yang seimbang antara presisi dan recall.

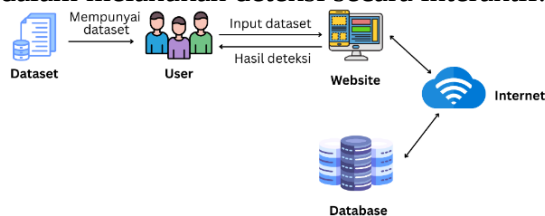
$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

ROC-AUC

ROC-AUC mengukur kemampuan model dalam membedakan kelas positif dan negative secara keseluruhan. Nilai AUC berada pada rentang 0 hingga 1, di mana semakin mendekati 1 berarti model semakin baik dalam membedakan dua kelas tersebut.

Implementasi Sistem

Sistem deteksi risiko diabetes yang dikembangkan pada penelitian ini diimplementasikan dalam bentuk aplikasi berbasis web untuk memudahkan pengguna dalam melakukan deteksi secara interaktif.



Gambar 2. Implementasi Sistem

Pada Gambar 2 menunjukkan alur kerja sistem deteksi risiko diabetes berbasis web. Proses dimulai dari pengguna yang memiliki data kesehatan. Data tersebut diinput melalui antarmuka website yang terhubung dengan server. Website memproses data menggunakan model LSTM dan XGBoost. Selanjutnya, hasil deteksi dikirim kembali kepada pengguna melalui website dalam bentuk informasi. Semua data input dan hasil deteksi disimpan dalam database yang terhubung ke server melalui jaringan internet, sehingga hasil dapat diakses kembali dengan mudah.

HASIL DAN PEMBAHASAN

Pada bagian ini membahas hasil yang diperoleh dari proses pelatihan dan pengujian model deteksi risiko diabetes menggunakan pendekatan hybrid *Long Short-Term Memory* (LSTM) dan *Extreme Gradient Boosting* (XGBoost). Selain itu, pada bagian ini juga dijelaskan analisis performa model berdasarkan metrik evaluasi, implementasi sistem berbasis web, serta pembahasan terhadap hasil penelitian yang telah dilakukan.

Hasil Pengujian Model

Penelitian ini menghasilkan model hybrid yang menggabungkan algoritma *Long Short-Term Memory* (LSTM) dan *Extreme Gradient Boosting* (XGBoost) untuk mendeteksi risiko diabetes berdasarkan data kesehatan pasien. Proses pelatihan dilakukan menggunakan dataset dari Kaggle yang telah melalui tahap pra-pemrosesan dan penyeimbangan data menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE).

Hasil evaluasi terhadap data uji ditunjukkan pada Tabel 3 berikut.

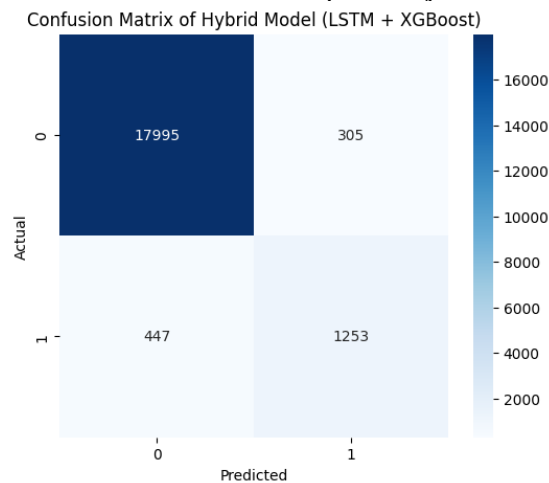
Tabel 3. Hasil Evaluasi Model

Metrik Evaluasi	Nilai
Akurasi	0.9624
Presisi	0.8042
Recall	0.7371
F1-Score	0.7692
ROC-AUC	0.9746

Berdasarkan hasil tersebut, model hybrid LSTM dan XGBoost mencapai akurasi sebesar 96,24% yang menunjukkan tingkat ketepatan model dalam mengklasifikasikan data pasien dengan benar. Nilai ROC-AUC sebesar 0.9746 menunjukkan kemampuan model yang sangat baik dalam membedakan antara pasien berisiko diabetes dan yang tidak berisiko. Nilai presisi sebesar 0.8042 menunjukkan bahwa sekitar 80% dari hasil prediksi positif benar-benar termasuk pasien berisiko diabetes. Sementara, recall sebesar 0.7371 menunjukkan bahwa model

berhasil mendeteksi 73% dari seluruh kasus positif.

Hasil *confolution matrix* pada Gambar 3 memperlihatkan distribusi klasifikasi model terhadap data uji.



Gambar 3. Confusion Matrix

Dari Gambar 3 dapat dilihat bahwa model mengklasifikasikan mayoritas data dengan benar, yaitu 17.995 data negatif dan 1.253 data positif, sementara kesalahan klasifikasi terjadi pada 305 data *false positive* dan 447 data *false negative*. Hal ini menunjukkan bahwa model hybrid LSTM+XGBoost dapat mengenali sebagian besar kasus dengan tepat, baik pada pasien yang berisiko maupun tidak berisiko diabetes.

Analisis Performa Model

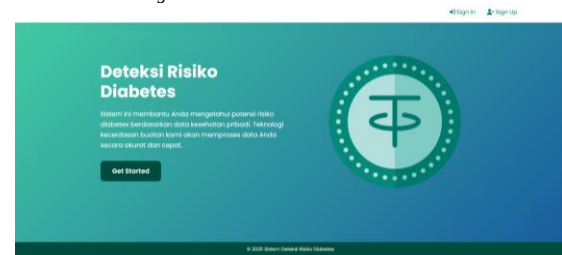
Hasil evaluasi menunjukkan bahwa kombinasi algoritma LSTM dan XGBoost mampu meningkatkan kinerja deteksi dibandingkan penggunaan metode Tunggal. LSTM berperan dalam mengekstraksi fitur, sementara XGBoost memperkuat kemampuan klasifikasi dengan meminimalkan kesalahan deteksi secara iterative.

Nilai akurasi 96,24% dan AUC 0,9746 menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan kelas positif dan negatif. Metrik recall yang lebih tinggi dari presisi mengindikasikan bahwa model cenderung lebih sensitive terhadap kasus positif, yang merupakan karakteristik penting dalam konteks deteksi dini penyakit

seperti diabetes. Nilai f1-Score sebesar 0,7692 menunjukkan keseimbangan yang cukup baik antara presisi dan recall. Secara keseluruhan, model hybrid ini memberikan performa yang kompetitif.

Implementasi Sistem Berbasis Website

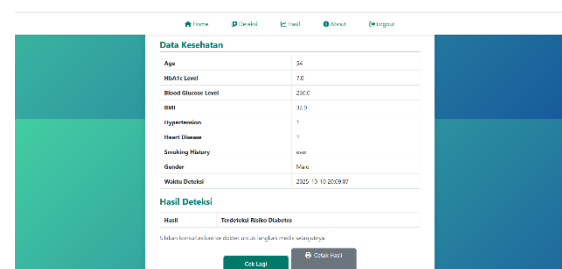
Model hybrid yang telah dilatih diintegrasikan ke dalam sistem berbasis web menggunakan framework Flask. Antarmuka website memungkinkan pengguna memasukkan data kesehatan seperti usia, jenis kelamin, BMI, tekanan darah, kadar HbA1c, dan kadar glukosa darah. Setelah data diinput, sistem akan melakukan proses normalisasi dan mengirimkan data tersebut ke model untuk dilakukan deteksi. Hasil deteksi ditampilkan langsung pada halaman hasil dalam bentuk kategori “Terdeteksi Risiko Diabetes” atau “Tidak Terdeteksi Risiko Diabetes” beserta informasi data kesehatannya.



Gambar 4. Tampilan Halaman Awal



Gambar 5. Tampilan Halaman Input



Gambar 6. Tampilan Halaman Hasil

Selain menampilkan hasil secara langsung, sistem ini juga menyimpan data input dan hasil deteksi ke dalam database MySQL sebagai Riwayat pemeriksaan.

Pembahasan

Dari hasil pengujian dan implementasi, dapat disimpulkan bahwa pendekatan hybrid LSTM dan XGBoost berhasil meningkatkan akurasi klasifikasi serta efisiensi proses deteksi risiko diabetes. Kelebihan utama dari sistem ini adalah kemampuannya untuk diintegrasikan ke dalam aplikasi berbasis web sehingga mudah digunakan. Meskipun performa model sudah sangat baik, nilai recall yang belum mencapai optimal menunjukkan bahwa beberapa kasus positif masih terdeteksi sebagai negatif (*false negative*). Secara keseluruhan, model hybrid yang diusulkan memiliki potensi besar untuk diimplementasikan dalam sistem pendukung keputusan medis, khususnya untuk deteksi dini risiko diabetes yang bersifat preventif.

SIMPULAN

Penelitian ini berhasil mengembangkan sistem deteksi risiko diabetes berbasis kombinasi algoritma *Long Short-Term Memory* (LSTM) dan *Extreme Gradient Boosting* (XGBoost), model hybrid ini dirancang untuk mengoptimalkan kemampuan LSTM dalam mengenali pola data sekuensial serta keunggulan XGBoost dalam klasifikasi yang efisien dan akurat. Berdasarkan hasil pengujian, model hybrid yang diusulkan menunjukkan performa yang sangat baik dengan nilai akurasi sebesar 96,24%, presisi 80,42%, recall 73,70%, F1-score 76,91%, dan ROC-AUC 97,46%. Nilai ROC-AUC yang tinggi menandakan kemampuan model dalam membedakan antara individu berisiko dan tidak berisiko diabetes secara konsisten. Hal ini membuktikan bahwa pendekatan hybrid lebih unggul dibandingkan metode Tunggal karena mampu mencapai keseimbangan antara sensitivitas dan spesifisitas.

Selain itu, model ini juga berhasil diimplementasikan ke dalam sistem berbasis web menggunakan framewrok Flask. Sistem memungkinkan pengguna untuk memasukkan data kesehatannya. Kemudian, data tersebut diproses secara otomatis oleh model untuk menghasilkan deteksi risiko diabetes yang ditampilkan dalam bentuk kategori “Berisiko” atau “Tidak Berisiko”. Dengan demikian, sistem ini dapat dimanfaatkan sebagai alat bantu deteksi dini yang efisien, mudah diakses, dan mendukung tenaga medis dalam pengambilan keputusan awal. Selanjutnya, penelitian ini dapat dikembangkan lebih lanjut dengan melakukan optimasi hyperparameter, penerapan metode feature selection, atau pengujian pada dataset klinis yang lebih beragam agar sistem dapat menghasilkan deteksi risiko diabetes yang semakin akurat dan adaptif terhadap kondisi nyata.

DAFTAR PUSTAKA

- Abdurrosyid, R., & Almais, A. T. W. (2025). Deteksi Dini Diabetes menggunakan Machine Learning dengan Metode PCA dan XGBoost. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 11(1), 51–56.
- Aditya, M. F., Pramuntadi, A., Wijaya, D. P., & Wicaksono, Y. (2024). Implementasi Metode Decision Tree pada Prediksi Penyakit Diabetes Melitus Tipe 2. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(3), 1104–1110.
- Adji Kusuma, D., Ratna Dewi, A., & Rony Wijaya, A. (2025). Perbandingan Random Forest dan Convolutional Neural Network dalam Memprediksi Peralihan Pelanggan. *Jurnal Informatika Sunan Kalijaga*, 10(2), 186–194.
- Cahyana, A., Susanto, E. R., & Parjito. (2025). Penerapan Algoritma XGBoost untuk Prediksi Diabetes: Analisis Confusion Matrix dan ROC

- Curve. *Fountain of Informatics Journal*, 10(1), 40–50.
- Chowdary, P. B. K., & Kumar, Dr. R. U. (2021). An Effective Approach for Detecting Diabetes using Deep Learning Techniques based on Convolutional LSTM Networks. *IJACSA) International Journal of Advanced Computer Science and Applications*, 12(4), 519–525.
- El-Attar, N. E., Moustafa, B. M., & Awad, W. A. (2022). Deep learning model to detect diabetes mellitus based on dna sequence. *Intelligent Automation and Soft Computing*, 31(1), 325–338.
- Febrianti, L. F., Nur, M. I., Haris, M. Al, & Amri, S. (2025). Implementasi Metode XGBoost dan SHAP untuk Klasifikasi dan Analisis Faktor Risiko Penyakit Diabetes Mellitus. *Seminar Nasional Sains Data*, 5(1), 336–346.
- Firdaus, D., Afin, Sumardi, I., & Chazar, C. (2025). Deteksi Serangan Pada Jaringan Internet Of Things Medis Menggunakan Machine Learning Dengan Algoritma XGBoost. *CyberSecurity Dan Forensik Digital*, 8(1), 34–42.
- Gunawan, S., Astuti, R., Prihartono, W., & Hamonangan, R. (2025). PREDIKSI DIABETES MELLITUS TIPE 2 DENGAN ALGORITMA LOGISTIC REGRESSION UNTUK PENDETEKSIAN DINI. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1), 848-854.
- Kwanda, K., Herwindiati, D. E., & Lauro, M. D. (2024). Perbandingan LSTM dan Bidirectional LSTM pada Sistem Prediksi Harga Saham Berbasis Website. *Ranah Research*, 7(1), 26–35.
- Panigrahy, S., Dash, S., & Padhy, S. (2024). SMOTE-based Deep LSTM System with GridSearchCV Optimization for Intelligent Diabetes Diagnosis. *J. Electrical Systems*, 20(7), 804–815.
- Putri, S. R., Arifin, M., & Supriyono. (2025). Sistemasi: Jurnal Sistem Informasi Analisis Sentimen Publik terhadap Nadiem Makarim sebagai Mendikbudristek menggunakan Support Vector Machine (SVM). *Sistemasi: Jurnal Sistem Informasi*, 14(2), 826–834.
- Swastika, R., Mukodimah, S., Susanto, F., Muslihudin, M., & Ipnuwati, S. (2023). *IMPLEMENTASI DATA MINING (Clustering, Association, Prediction, Estimation, Classification)*. Indramayu: Penerbit Adab.
- World Health Organizatio (WHO). (2024, November 14). Who.Int. diakses pada 10 Oktober 2025, pukul 14.00 dari <https://www.who.int/news-room/fact-sheets/detail/diabetes>
- Yulianti, A., Fitri, F., Amalita, N., & Vionanda, D. (2023). The SMOTE Application of CART Methods for Coping Imbalanced Data in Classifying Status Work on Labor Force in the City of Padang. *UNP JOURNAL OF STATISTICS AND DATA SCIENCE*, 1(3), 172–179.