JURNAL DEVICE, VOL. 15 NO 1, Mei 2025, 163-170
ISSN: 0216-9185 | eISSN: 2746-8984

KLASIFIKASI PENYAKIT DIABETES MELITUS MENGGUNAKAN METODE STACKING ENSEMBLE

Noor Herlinawati [1)], Kusrini [2)]

[1,2)] Universitas Amikom Yogyakarta
Email : noor.herlinawati@students.amikom.ac.id [1)], kusrini@amikom.ac.id [2)]

ABSTRAK

Pendeteksian dini terhadap risiko diabetes merupakan tantangan penting dalam dunia medis modern. Tujuan penelitian ini adalah untuk meningkatkan akurasi klasifikasi pasien diabetes menggunakan pendekatan stacking ensemble, yang menggabungkan tiga model pembelajaran mesin: K-Nearest Neighbors (KNN), Random Forest, dan XGBoost. Penelitian menggunakan dataset Pima Indians Diabetes yang terdiri dari 768 data pasien. Setelah dilakukan preprocessing, balancing, dan feature selection, model dibangun dengan Logistic Regression sebagai meta-learner. Model berhasil mencapai akurasi 77,27% dengan skor ROC AUC sebesar 82,91%. Metode ini menunjukkan potensi besar dalam pengembangan sistem diagnosis otomatis yang lebih andal untuk penyakit diabetes.

Kata Kunci : diabetes melitus, stacking ensemble, knn, random forest, XGBoost.

ABSTRACT

Early detection of diabetes risk is a critical challenge in modern medicine. The aim of this study is to improve the accuracy of diabetes patient classification using a stacking ensemble approach, which combines three machine learning models: K-Nearest Neighbors (KNN), Random Forest, and XGBoost. The research utilizes the Pima Indians Diabetes dataset, consisting of 768 patient records. After performing preprocessing, data balancing, and feature selection, the model was built using Logistic Regression as the meta-learner. The resulting model achieved an accuracy of 77.27% and a ROC AUC score of 82.91%. This method demonstrates strong potential for the development of more reliable automated diagnostic systems for diabetes.

Keywords: diabetes melitus, stacking ensemble, knn, random forest, XGBoost

1. PENDAHULUAN

Diabetes merupakan penyakit kronis yang berdampak besar terhadap kesehatan masyarakat global. Kondisi ini ditandai dengan kadar gula darah yang tinggi akibat gangguan dalam produksi atau penggunaan insulin, baik karena defisiensi (diabetes tipe 1) maupun resistensi (diabetes tipe 2). Jika tidak ditangani dengan baik, diabetes dapat memicu berbagai komplikasi serius, seperti kerusakan pada pembuluh darah, ginjal, mata, saraf, bahkan jantung. Data dari WHO mencatat peningkatan angka kematian akibat diabetes sebesar 3% secara global dari tahun 2000 hingga 2019, di mana hampir setengah dari kematian tersebut terjadi sebelum usia 70 tahun, dengan kenaikan signifikan di negara-negara berpenghasilan rendah dan menengah (Carpinteiro et al., 2023; Phongying & Hiriote, 2023).

Selain dampak kesehatan, diabetes juga memberikan beban ekonomi yang besar, terutama akibat biaya perawatan komplikasi yang kian meningkat. Oleh karena itu, deteksi dini dan manajemen penyakit menjadi sangat penting. Dalam beberapa tahun terakhir, pendekatan berbasis teknologi, seperti kecerdasan buatan dan machine learning (ML), mulai banyak diterapkan dalam upaya mendeteksi diabetes secara lebih akurat dan cepat (Nguyen et al., 2023). ML mampu mengenali pola dari data historis dan menggunakannya untuk membuat prediksi terhadap data baru, yang sangat berguna dalam konteks diagnosis medis.

Sebelum proses klasifikasi dilakukan, data perlu dipersiapkan melalui serangkaian tahap preprocessing seperti pembersihan data, penanganan nilai hilang, dan normalisasi. Ini bertujuan agar proses pembelajaran mesin dapat berlangsung secara optimal (Ahmed & Razak, 2024). Dalam klasifikasi, model ML biasanya menentukan hasil berdasarkan mayoritas prediksi dari berbagai algoritma. Untuk meningkatkan akurasi, beberapa pendekatan lanjutan seperti ensemble learning telah dikembangkan. Pendekatan ini menggabungkan beberapa model dasar menjadi satu model yang lebih kuat dan stabil (Kim et al., 2023; Kumari & Upadhaya, 2024).

Metode ensemble terbukti efektif dalam mengatasi kekurangan dari algoritma ML tunggal, terutama saat menghadapi data yang besar, tidak seimbang, atau mengandung noise. Model ensemble seperti stacking atau boosting tidak hanya meningkatkan akurasi tetapi juga memperkuat kemampuan generalisasi prediksi. Misalnya, dalam studi oleh (Nguyen et al., 2023), beberapa algoritma seperti Decision Tree, Logistic Regression, SVM, hingga Random Forest dibandingkan untuk klasifikasi diabetes. Hasilnya menunjukkan bahwa Random Forest menghasilkan akurasi tinggi, namun juga berisiko overfitting karena perbedaan besar antara performa data latih dan uji. Hal ini menunjukkan pentingnya pemilihan model dan teknik evaluasi yang tepat.

Penelitian lain oleh (Liang et al., 2025) menunjukkan efektivitas Bayesian Networks dalam memodelkan hubungan probabilistik antar variabel menggunakan dataset PIMA Indians, dengan Logistic Regression sebagai baseline. Evaluasi menggunakan ROC dan AUC memperlihatkan bahwa pendekatan ini mampu menangkap kompleksitas hubungan antar fitur, meskipun masih memiliki keterbatasan dalam aspek generalisasi model.

Pendekatan stacking ensemble juga diperkuat oleh studi (Reza et al., 2024), yang menggunakan beberapa algoritma dasar dengan logistic regression sebagai meta-learner. Model ini berhasil mencapai akurasi hingga 77,10% pada data pima, menunjukkan bahwa kombinasi model mampu memberikan performa yang lebih tinggi dibanding model tunggal. Hasil ini memperkuat posisi ensemble learning sebagai metode unggulan dalam deteksi dini diabetes.

Oleh karena itu, guna meningkatkan akurasi dan keandalan dalam proses klasifikasi diabetes, penelitian ini mengusulkan penerapan metode stacking ensemble yang menggabungkan kekuatan dari beberapa algoritma pembelajaran mesin, yaitu K-Nearest Neighbors (KNN), Random Forest (RF), dan Extreme Gradient Boosting

(XGBoost) menggunakan Pima Indian Diabetes Database. Pendekatan ini diharapkan mampu serta menghasilkan performa prediksi yang optimal dalam mendeteksi diabetes secara dini.

2. METODE

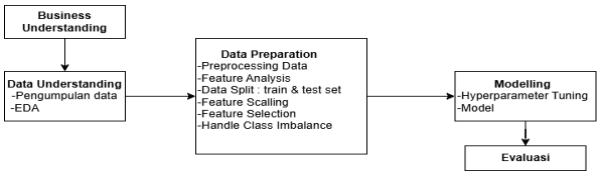
Penelitian ini menggunakan pendekatan kuantitatif untuk mengklasifikasikan diabetes berdasarkan dataset Pima Indians Diabetes Database dari UCI Machine Learning Repositor. Model klasifikasi ini diterapkan untuk mengetahui nilai rata-rata akurasi dalam melakukan klasifikasi penyakit Diabetes melitus menggunakan metode stacking ensemble KNN, Random Forest dan XGBoost.

Pima Indians Diabetes Database digunakan sebagai dataset dalam penelitian ini. Dataset terdiri dari 768 sampel pasien perempuan berusia diatas 21 tahun dari suku PIMA di Amerika Serikat. Setiap sampel memiliki 8 fitur independen, seperti kadar glukosa, tekanan darah, indeks massa tubuh (BMI), dan riwayat diabetes dalam keluarga, serta satu label target yang menunjukkan apakah pasien menderita diabetes atau tidak.

Dataset diperoleh dari UCI Machine Learning Repository yang telah melalui tahap preprocessing data seperti pembersihan data dan normalisasi data untuk meningkatkan akurasi model.

Analisis data dilakukan dengan menerapkan algoritma KNN, XGBoost dan Random Forest sebagai Base Learners pada metode Stacking Ensembles. Evaluasi model menggunakan

Tahapan pada penelitian berdasarkan Gambar 1. Alur Penelitian adalah:



Gambar 1. Alur Penelitian

2.1. Business Understanding

Tahapan ini bertujuan untuk memperoleh pemahaman yang mendalam mengenai objek penelitian melalui kajian literatur yang relevan dengan klasifikasi diabetes melitus, serta merumuskan permasalahan yang akan dipecahkan dan menetapkan tujuan dari penelitian..

2.2. Data Understanding

Tahap selanjutnya adalah mengumpulkan data penyakit diabetes dan dataset PIMA Indians Diabetes Database, kemudian melakukan *Exploratory Data Analysis* dengan tujuan memahami struktur, pola dan potensi masalah pada data.

2.3. Data Preparation

Setelah melakukan data understanding, Langkah selanjutnya menyiapkan data untuk dimodelkan yang terdiri dari tahapan *preprocessing data, feature analysis,data split, feature scalling, feature selection* dan *handle class imbalance.*

2.4. Modelling

Tahapan ini berguna untuk mengoptimalkan parameter model dan membangun ensemble model klasifikasi menggunakan algoritma KNN, Random Forest dan XGBoost.

2.5. Evaluasi

Tahap ini dilakukan untuk menilai kinerja model klasifikasi dengan menggunakan berbagai metrik evaluasi, seperti akurasi, presisi, recall, F1-Score, dan ROC-AUC, guna memastikan seberapa baik model dalam mengklasifikasikan data secara akurat dan seimbang.

3. HASIL DAN PEMBAHASAN

Sistem klasifikasi diabetes melitus menggunakan metode stacking ensemble KNN, Random Forest dan XGBoost sudah dilaksanakan menggunakan dataset PIMA Indians Diabetes Dataset dengan 2 label target yaitu pasien menderita diabetes atau tidak. Data terdiri dari 768 sampel pasien dan terdiri dari 8 fitur yaitu Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function dan Age.

3.1. Diabetes Melitus

Diabetes melitus (DM) adalah kelompok penyakit metabolik ditandai dengan peningkatan kadar gula darah (atau disebut

hiperglikemia) yang diakibatkan oleh kelainan sekresi insulin, kerja insulin, dan atau keduanya. Diabetes melitus terbagi menjadi empat kelompok yaitu Diabetes melitus tipe-1 (DM tipe-1), Diabetes melitus tipe-2 (DM tipe-2), Diabetes melitus gestasional (diabetes pada kehamilan), dan Diabetes tipe spesifik yang berkaitan dengan penyebab lain.

3.2. KNN

Menurut (Parsian, 2015) Algoritma K-Nearest Neighbors (KNN) adalah metode non-parametrik dalam pengenalan pola untuk mengklasifikasikan objek berdasarkan contoh pelatihan terdekat dalam ruang fitur. KNN merupakan jenis instance-based learning atau lazy learning, Dimana fungsi hanya didekati secara local dan seluruh perhitungan ditunda hingga proses klasifikasi. Salah satu algoritma Machine Learning yang paling sederhana adalah algoritma KNN, yang mengklasifikasikan objek berdasarkan suara paling banyak dari tetangganya. Kelas yang paling umum di antara K tetangga terdekat, di mana k adalah bilangan bulat positif, biasanya kecil , dan jika k=1, maka tetangga terdekatnya hanya akan diberi kelas yang sama.

3.3. Random Forest

Menurut (Quinto, 2020)Algoritma Random Forest menggunakan kumpulan keputusan pohon untuk regresi dan klasifikasi. Untuk mengurangi varians sambil mempertahankan bias yang rendah, algoritma ini menggunakan teknik yang dikenal sebagai bagging, atau agregasi bootstrap. Pohon-pohon tertentu dilatih dengan bagasi dari subset data pelatihan. Random Forest menggunakan teknik bagging fitur juga. Feature bagging menggunakan subset fitur, atau kolom, dibandingkan dengan bagging, yang menggunakan subset observasi. Tujuan dari feature bagging adalah untuk mengurangi korelasi antara pohon keputusan. Pohon-pohon individu akan sangat mirip tanpa fitur bagging, terutama dalam kasus di mana satu fitur dominan. Untuk klasifikasi, suara mayoritas modus atau output dari masing-masing pohon menjadi prediksi akhir model.

3.4. XGBoost

XGBoost adalah kerangka kerja pembelajaran mesin yang mendukung bahasa pemrograman Python. XGBoost menjalankan model yang ditingkatkan dengan gradien yang dapat diskalakan dan dapat mempelajari komputasi paralel dan terdistribusi dengan cepat tanpa mengorbankan efisiensi memori. Selain itu, dia adalah pebelajar kelompok XGBoost memiliki kemampuan untuk memecahkan masalah regresi dan klasifikasi. XGBoost menggunakan peningkatan untuk memperbaiki kesalahan pada pohon sebelumnya. Saat model berbasis pohon terlalu sesuai, itu bermanfaat. Untuk menginstal model di lingkungan Python Anda, gunakan pip install xgboost (Nokeri, 2021).

3.5. Stacking Ensemble

Ensemble Learning adalah kombinasi beberapa teknik *Machine Learning* yang dilakukan bersama dengan cepat menjadi standar yang digunakan untuk mendapatkan peningkatan akurasi dalam model *Machine Learning* dalam *Data Science*. Dalam teknik ensemble ini, beberapa model dasar (base learners) terlebih dahulu dilatih secara bersamaan untuk menghasilkan prediksi. Selanjutnya, hasil prediksi dari masing-masing model digunakan sebagai data pelatihan bagi model lain yang disebut *meta-learner*. Teknik ini dapat dianggap sebagai proses penumpukan lapisan-lapisan pembelajaran mesin, di mana satu set model pembelajar ditambahkan di atas model lainnya untuk membentuk suatu *meta-learner*. Penumpukan ini bertujuan untuk menggabungkan kekuatan dari beberapa model guna meningkatkan akurasi prediksi. (Kumar & Jain, 2020).

3.6. Preprocessing Data

Pada tahap ini, dataset *Pima Indian Diabetes* melalui beberapa proses pra-pemrosesan data, yaitu pengecekan duplikasi, nilai hilang (missing values), pembersihan data, deteksi outlier dan winsorisasi, serta identifikasi data yang tidak relevan.

Hasil pengecekan menunjukkan bahwa dataset tidak mengandung duplikasi maupun missing value secara eksplisit. Namun, dalam

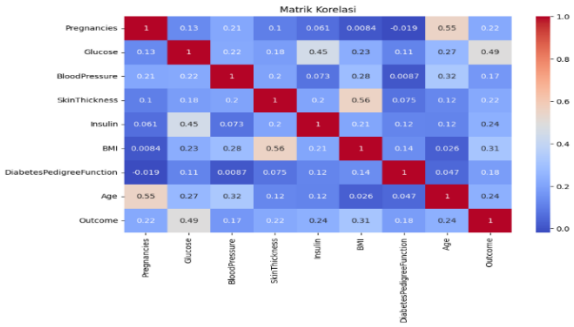
proses pembersihan data, nilai nol pada kolom ‘Glucose’, ‘BloodPressure’, ‘SkinThickness’, ‘Insulin’, dan ‘BMI’ dianggap sebagai nilai kosong, kemudian diganti dengan *NaN* dan diimputasi menggunakan nilai median.

Selanjutnya, dilakukan deteksi outlier menggunakan visualisasi *boxplot* dan metode *Interquartile Range* (IQR). Nilai-nilai yang terdeteksi sebagai outlier kemudian disesuaikan (di-*winsorize*) dengan menggantinya menggunakan nilai batas atas atau bawah.

Terakhir, dilakukan pemeriksaan terhadap fitur yang mungkin tidak relevan. Namun, karena dataset ini bersifat numerik dan berisi data klinis, tidak ditemukan fitur yang secara eksplisit tidak relevan seperti ID pengguna, nama pasien, *timestamp*, atau data berbentuk teks bebas.

3.7. Feature Analysis

Berdasarkan hasil dari analisis fitur didapatkan korelasi matriks antara berbagai fitur dalam *Pima Indian Diabetes Dataset*, termasuk target variabel outcome (1 = diabetes, 0 = tidak diabetes). Korelasi berkisar antara -1 (sangat negative) hinggat +1 (sangat positif).



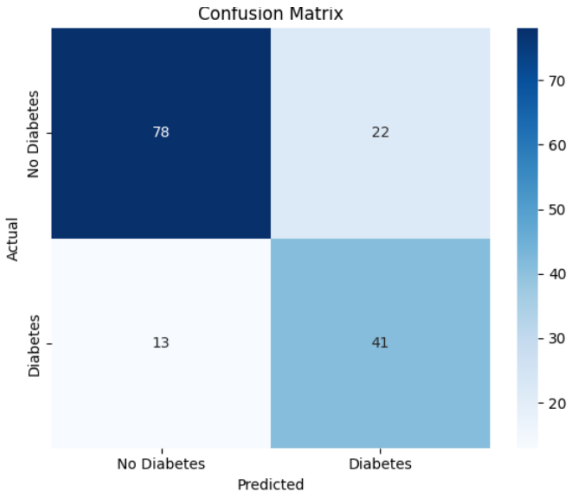
Gambar 2. Matrik Korelasi

Gambar 2. Matrik Korelasi menunjukkan fitur yang memiliki hubungan paling kuat dengan diagnosa diabetes adalah Glucose, BMI, Age, Insulin, Pregnacies, dan SkinThickness. Sedangkan BloodPressure dan DiabetesPedigreFunction merupakan fitur dengan korelasi tidak signifikan.

3.8. Evaluasi

Setelah dilakukan pelatihan menggunakan stacking ensemble yaitu KNN, Random Forest dan XGBoost dan kemudian model dievaluasi didapatkan hasil dengan akurasi keseluruhan sebesar 77.27% dan ROC AUC 82.91%, model stacking ensemble menunjukkan kinerja yang cukup baik.

Evaluasi performa model dilakukan menggunakan confusion matrix dan classification report untuk mengetahui seberapa baik model dapat mengklasifikasikan pasien sebagai diabetes atau non-diabetes.



Gambar 3. Confusion Matrix

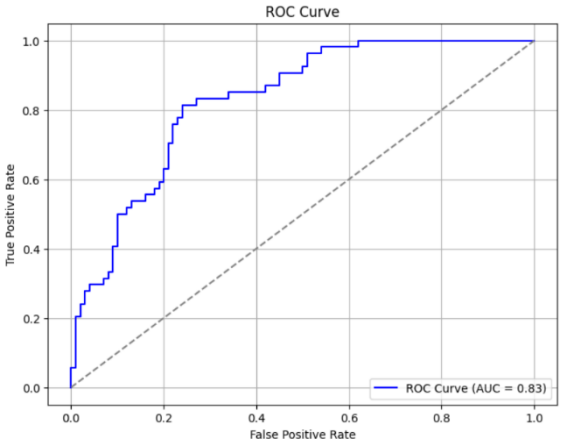
Untuk Gambar 3. Confusion Matrix menunjukkan bahwa dari 100 pasien non-diabetes, 78 diklasifikasikan dengan benar, sedangkan 22 salah diklasifikasikan sebagai penderita diabetes. Dari 54 pasien diabetes, 41 berhasil dideteksi dengan benar, sementara 13 kasus tidak terdeteksi (false negative). Precision kelas diabetes sebesar 0.65 menunjukkan bahwa masih terdapat tingkat false positive yang cukup tinggi (22 kasus), di mana pasien non-diabetes salah diklasifikasikan sebagai diabetes. Recall kelas diabetes sebesar 0.76 mengindikasikan bahwa sebagian besar pasien diabetes berhasil dikenali oleh model. F1-score diabetes sebesar 0.70 menunjukkan keseimbangan yang tepat antara kemampuan model mendeteksi kasus positif secara benar dan menghindari kesalahan klasifikasi. Hasil penelitian ini menunjukkan model ini dapat digunakan sebagai alat bantu skrining awal terhadap pasien berisiko diabetes.

Classification Report:

	precision	recall	f1-score	support
No Diabetes	0.86	0.78	0.82	100
Diabetes	0.65	0.76	0.70	54
accuracy			0.77	154
macro avg	0.75	0.77	0.76	154
weighted avg	0.78	0.77	0.78	154

Gambar 4. Classification Report

Gambar 4. Classification Report menunjukkan performa model dalam mengklasifikasikan dua kelas: "No Diabetes" dan "Diabetes". Model mencapai akurasi sebesar 77%, yang berarti 77 dari 100 prediksi sesuai dengan label sebenarnya. Model lebih akurat dalam mengklasifikasikan "No Diabetes" dengan precision 0.86 dan f1-score 0.82, dibandingkan dengan kelas "Diabetes", yang memiliki precision 0.65 dan f1-score 0.70. Meskipun nilai recall untuk kelas Diabetes cukup tinggi (0.76), rendahnya precision menunjukkan masih adanya false positive yang perlu diminimalkan untuk meningkatkan ketepatan prediksi.



Gambar 5. ROC Curve

Gambar 5. ROC Curve menunjukkan kurva ROC (Receiver Operating Characteristic) dari model klasifikasi, dengan nilai AUC (Area Under Curve) sebesar 0.83. Kurva ROC menggambarkan hubungan antara True Positive Rate (Sensitivity) dan False Positive Rate, pada berbagai ambang batas klasifikasi. Semakin melengkung ke kiri atas, semakin baik kinerja model. Nilai AUC = 0.83 menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan antara kelas "Diabetes" dan "No Diabetes". AUC mendekati 1.0 menandakan performa klasifikasi yang kuat, sementara nilai 0.5 berarti tidak lebih baik dari tebakan acak.

Model stacking ensemble yang digunakan dalam penelitian ini mencapai akurasi sebesar 77,27% dan AUC sebesar 0.83, menandakan performa klasifikasi yang cukup baik. Nilai AUC yang lebih tinggi dari akurasi menunjukkan bahwa meskipun beberapa prediksi masih salah, model tetap memiliki kemampuan yang kuat dalam membedakan antara pasien diabetes dan non-diabetes berdasarkan probabilitas.

Hasil ini konsisten dengan studi sebelumnya. Sebagai contoh, Reza et al. (2024) melaporkan penggunaan stacking ensemble model (menggabungkan Decision Tree, Random Forest dan SVM) pada dataset yang sama dengan hasil akurasi 77.10%. Dengan demikian, performa model dalam penelitian ini sedikit lebih baik secara numerik, yang kemungkinan disebabkan oleh penggunaan algoritma yang lebih kuat (Random Forest dan XGBoost) serta penanganan data imbalance menggunakan SMOTETomek.

Namun demikian, penggunaan tiga algoritma kuat secara bersamaan dalam ensemble yakni KNN, Random Forest, dan XGBoost berpotensi meningkatkan risiko overfitting, terutama jika tidak dibarengi dengan validasi silang dan regularisasi yang memadai. Meskipun hasil pengujian menunjukkan performa yang cukup baik, belum tentu model mempertahankan kinerjanya saat diuji pada data baru di luar dataset PIMA. Oleh karena itu, evaluasi lanjutan seperti cross-validation dan pengujian pada dataset eksternal sangat disarankan agar generalisasi model lebih terjamin.

Secara keseluruhan, pendekatan ini menunjukkan potensi besar dalam membantu deteksi awal diabetes, namun perlu dikaji lebih lanjut sebelum diterapkan dalam skala klinis.

4. PENUTUP

4.1. Kesimpulan

Penelitian ini mengembangkan model klasifikasi diabetes melitus menggunakan

metode stacking ensemble yang dipadukan dengan SMOTETomek untuk menangani ketidakseimbangan data. Model menggabungkan KNN, Random Forest, dan XGBoost sebagai base learners, serta Logistic Regression sebagai meta-learner. Tahapan preprocessing seperti penanganan missing values, outlier, seleksi fitur, dan balancing data turut meningkatkan performa model. Hasil evaluasi menunjukkan akurasi sebesar 77,27% dan skor ROC-AUC 82,91%, menandakan kemampuan klasifikasi yang baik. Pendekatan ini menunjukkan potensi besar sebagai solusi diagnosis diabetes yang andal.

4.2. Saran

Untuk penelitian berikutnya diharapkan bisa menguji model ini dengan menggunakan dataset kesehatan yang lain sehingga model ini dapat dikembangkan lebih lanjut untuk di terapkan ke dalam sistem pendukung keputusan klinis.

5. DAFTAR PUSTAKA

Ahmed, A. Z., & Razak, D. T. A. (2024). Soil Classification using the Stacking Ensemble Learning Technique for Crop Agronomy. In *J. Electrical Systems* (Vol. 20, Issue 7).

Carpinteiro, C., Lopes, J., Abelha, A., & Santos, M. F. (2023). A Comparative Study of Classification Algorithms for Early Detection of Diabetes. *Procedia Computer Science*, *220*, 868–873. https://doi.org/10.1016/j.procs.2023.03.117

Kim, K. B., Park, H. J., & Song, D. H. (2023). Combining Supervised and Unsupervised Fuzzy Learning Algorithms for Robust Diabetes Diagnosis. *Applied Sciences (Switzerland)*, *13*(1). https://doi.org/10.3390/app13010351

Kumar, A., & Jain, M. (2020). Ensemble Learning for AI Developers: Learn Bagging, Stacking, and Boosting Methods with Use Cases. In *Ensemble Learning for AI Developers: Learn Bagging, Stacking, and Boosting Methods with Use Cases*. Apress Media LLC. https://doi.org/10.1007/978-1-4842-5940-5

Kumari, S., & Upadhaya, A. (2024). *Investigating Role of Supervised Machine Learning Approach in Classification of Diabetic Patient*. https://doi.org/https://doi.org/10.52783/jes.2987

Liang, X., Song, W., Yang, W., & Yue, Z. (2025). Enhancing diabetes risk assessment through Bayesian networks: An in-depth study on the Pima Indian population. *Endocrine and Metabolic Science*, *17*. https://doi.org/10.1016/j.endmts.2024.100212

Nguyen, L. P., Tung, D. D., Nguyen, D. T., Le, H. N., Tran, T. Q., Binh, T. Van, & Pham, D. T. N. (2023). The Utilization of Machine Learning Algorithms for Assisting Physicians in the Diagnosis of Diabetes. *Diagnostics*, *13*(12). https://doi.org/10.3390/diagnostics13122087

Nokeri, T. C. (2021). Data Science Solutions with Python: Fast and Scalable Models Using Keras, PySpark MLlib, H2O, XGBoost, and Scikit-Learn. In *Data Science Solutions with Python: Fast and Scalable Models Using Keras, PySpark MLlib, H2O, XGBoost, and Scikit-Learn*. Springer International Publishing. https://doi.org/10.1007/978-1-4842-7762-1

Parsian, Mahmoud. (2015). *Data algorithms : recipes for scaling up with Hadoop and Spark* (A. Spencer & M. Beaugureau, Eds.; Vol. 1). O'Reilly Media.

Phongying, M., & Hiriote, S. (2023). Diabetes Classification Using Machine Learning Techniques. *Computation*, *11*(5). https://doi.org/10.3390/computation11050096

Quinto, B. (2020). Next-generation machine learning with spark: Covers XGBoost, LightGBM, Spark NLP, distributed deep learning with keras, and more. In *Next-Generation Machine Learning with*

Spark: Covers XGBoost, LightGBM, Spark NLP, Distributed Deep Learning with Keras, and More. Apress Media LLC. https://doi.org/10.1007/978-1-4842-5669-5

Reza, M. S., Amin, R., Yasmin, R., Kulsum, W., & Ruhi, S. (2024). Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data. *Heliyon*, *10*(2). https://doi.org/10.1016/j.heliyon.2024.e24536