

ANALISIS KLASIFIKASI DIAGNOSA PENYAKIT DIABETES MELITUS BERDASARKAN KOMPARASI ALGORITMA *SUPERVISED LEARNING*

Sherly Taurin Siridion¹, Bakti Siregar²

Universitas Matana, Banten, Indonesia

Email: sherly.taurin@student.matanauniversity.ac.id

KATA KUNCI

Diabetes Mellitus, Comparison, Machine Learning, Supervised Learning

ABSTRACT

Diabetes mellitus is a serious health problem that threatens various age groups, including children, teenagers, and adults. In 2021, the mortality rate due to diabetes mellitus reached alarming levels, making it a global threat, especially in Indonesia, where the number of patients reached 19.5 million. Efforts to address diabetes mellitus include early prediction of the disease's risk in patients, and machine learning approaches have shown potential in this regard. This study employs a quantitative method by utilizing secondary data from the UC Irvine Machine Learning Repository titled "Early Stages Diabetes Risk Prediction". The data was obtained from questionnaires filled out by diabetes patients at Sylhet Diabetes Hospital and validated by healthcare professionals. A total of 520 data samples with 17 attributes were used for analysis. The tested Supervised Learning algorithms include Logistic Regression, K-Nearest Neighbour, Support Vector Machine, Random Forest, Naïve Bayes Classifier, Artificial Neural Network, Decision Tree C4.5, and Gradient Boosting Classifier. The research findings reveal that the Random Forest algorithm achieved the highest accuracy of 98.71% in diagnosing diabetes mellitus. This study significantly contributes to enhancing the understanding of diabetes mellitus and has the potential for further development in finding the best algorithm for early disease prediction. It is hoped that this research will make a significant contribution to the efforts in preventing and managing diabetes mellitus, ultimately improving the quality of life for patients and reducing its impact on the population.

ABSTRAK

Penyakit diabetes melitus merupakan masalah kesehatan serius yang mengancam berbagai kelompok usia, termasuk anak-anak, remaja, dan dewasa. Pada tahun 2021, angka kematian akibat diabetes melitus mencapai tingkat yang mengkhawatirkan, menjadikan penyakit ini sebagai ancaman global khususnya di Indonesia dengan jumlah penderita mencapai 19,5 juta. Upaya untuk mengatasi diabetes melitus termasuk prediksi dini risiko penyakit pada pasien, dan pendekatan machine learning menunjukkan potensi dalam hal ini. Studi ini menggunakan metode kuantitatif dengan memanfaatkan data sekunder dari UC Irvine Machine Learning Repository berjudul "*Early Stages Diabetes Risk Prediction*". Data tersebut berasal dari kuesioner yang diisi oleh

pasien diabetes di Sylhet Diabetes Hospital dan telah divalidasi oleh tenaga medis. Total 520 data dengan 17 atribut digunakan untuk analisis. *Algoritma Supervised Learning* yang diuji mencakup *Logistic Regression*, *K-Nearest Neighbour*, *Support Vector Machine*, *Random Forest*, *Naïve Bayes Classifier*, *Artificial Neural Network*, *Decision Tree C4.5*, dan *Gradient Boosting Classifier*. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* menghasilkan akurasi tertinggi sebesar 98,71% dalam mendiagnosa diabetes melitus. Penelitian ini memberikan kontribusi penting dalam meningkatkan pemahaman tentang diabetes melitus dan berpotensi untuk pengembangan lebih lanjut guna menemukan algoritma terbaik dalam prediksi dini penyakit ini. Diharapkan bahwa penelitian ini akan memberikan sumbangan yang signifikan dalam upaya pencegahan dan penanganan diabetes melitus, sehingga dapat meningkatkan kualitas hidup pasien dan mengurangi dampaknya pada tingkat populasi.

PENDAHULUAN

Diabetes Melitus adalah penyakit serius dan mematikan yang dapat menyerang individu dari berbagai kelompok usia, termasuk anak-anak, remaja, dan dewasa. Kondisi ini menyebabkan gangguan pada metabolisme tubuh, meningkatkan kadar glukosa dalam darah, serta mengganggu metabolisme karbohidrat, lipid, dan protein akibat kurangnya fungsi insulin (Masenda, 2023). Statistik dari *International Diabetes Federation (IDF)* pada tahun 2021 menunjukkan bahwa terdapat lebih dari 537 juta orang dewasa di seluruh dunia yang menderita diabetes, dengan lebih dari 6,7 juta kematian akibat penyakit ini. Tiongkok menjadi negara dengan jumlah pengidap diabetes terbanyak, mencapai 140,87 juta orang, sedangkan Indonesia berada di peringkat kelima dengan jumlah penderita mencapai 19,5 juta orang (Atlas, 2021). Hal ini menandakan bahwa diabetes melitus merupakan ancaman serius bagi kesehatan masyarakat di Indonesia maupun secara global.

Penting bagi masyarakat untuk memahami ciri-ciri penyakit diabetes melitus guna mendeteksi dan menghadapinya lebih awal (Ismah, 2018). Beberapa gejala umum diabetes meliputi haus berlebihan, sering buang air kecil, penurunan berat badan yang drastis, keluhan kulit, luka yang sulit sembuh, mudah lelah, perubahan penglihatan, serta kesemutan atau mati rasa pada anggota tubuh tertentu (Kemenkes, 2019). Namun, masih banyak masyarakat yang kurang menyadari tanda-tanda ini sebagai gejala diabetes melitus, sehingga diperlukan upaya dalam penanggulangan penyakit ini agar masyarakat dapat lebih memahami dan mengenali gejala-gejala yang timbul.

Di Indonesia, berbagai upaya telah dilakukan untuk mengatasi diabetes melitus, termasuk penyuluhan tentang pencegahan, program Kesehatan Dasar Masyarakat (KDM) yang menyediakan fasilitas pemeriksaan kadar gula darah secara rutin dengan biaya ditanggung negara dimana seseorang akan terdiagnosa terkena penyakit diabetes melitus ketika hasilnya $\geq 200\text{mg/dL}$ bila pemeriksaan gula darah dilakukan sewaktu atau 2 jam sesudah makan dan $\geq 126\text{mg/dL}$ bila pemeriksaan dilakukan setelah puasa, pembuatan roadmap pengendalian kasus diabetes, dan deteksi dini penyakit diabetes melitus (HR, Pratiwi, & Susanti, 2021).

Salah satu metode yang dapat digunakan untuk prediksi dini penyakit diabetes melitus adalah dengan memanfaatkan algoritma machine learning dalam pendekatan klasifikasi (Ridwan, 2020). Penelitian ini bertujuan untuk membandingkan berbagai *algoritma*

supervised learning seperti *Logistic Regression*, *K-Nearest Neighbor*, *Support Vector Machine*, *Random Forest*, *Naïve Bayes Classifier*, *Artificial Neural Network*, *C4.5*, dan *Gradient Boosting Classifier* untuk mendiagnosa diabetes dengan akurasi tertinggi. Bahasa pemrograman *Python* digunakan dalam penelitian ini, dan evaluasi *Confusion Matrix (Accuracy)* dan *ROC Curve* digunakan untuk mengevaluasi kinerja algoritma. Diharapkan penelitian ini dapat memberikan kontribusi penting dalam upaya pencegahan dan penanganan penyakit diabetes melitus dengan mendukung identifikasi dini dan pengembangan algoritma prediksi yang lebih efektif.

Beberapa penelitian sebelumnya telah dilakukan untuk mendiagnosa diabetes melitus menggunakan berbagai algoritma machine learning. (Anggoro & Novitaningrum, 2021) dalam penelitiannya "*Comparison of Accuracy Level of Support Vector Machine (SVM) and Artificial Neural Network (ANN) Algorithms in Predicting Diabetes Mellitus Disease*" membandingkan *Algoritma Support Vector Machine* dan *Artificial Neural Network*, dengan hasil *Artificial Neural Network* memiliki akurasi tertinggi sebesar 85,20%. (Kazerouni et al., 2020) dalam "*Type2 Diabetes Mellitus Prediction Using Data Mining Algorithms Based on The Long-Noncoding Rnas Expression: A Comparison of Four Data Mining Approaches*" mendiagnosa Diabetes Melitus Tipe 2 menggunakan *algoritma K-Nearest Neighbors*, *Support Vector Machine*, *Artificial Neural Network*, dan *Logistic Regression* dengan akurasi tinggi untuk keempat algoritma tersebut. (Deberneh & Kim, 2021) dalam "*Prediction of Type 2 Diabetes Based on Machine Learning Algorithm*" mengenai mendiagnosa diabetes tipe 2 menggunakan *algoritma Logistic Regression*, *Artificial Neural Network*, dan *Decision Tree* dengan akurasi tertinggi diperoleh oleh *Decision Tree* sebesar 77,87%. Selanjutnya, (Putry, 2022) dalam "*Komparasi Algoritma KNN dan Naive Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus*" membandingkan *algoritma K-Nearest Neighbors* dan *Naïve Bayes* dengan hasil algoritma terbaik adalah *Naïve Bayes* dengan akurasi 80%.

METODE PENELITIAN

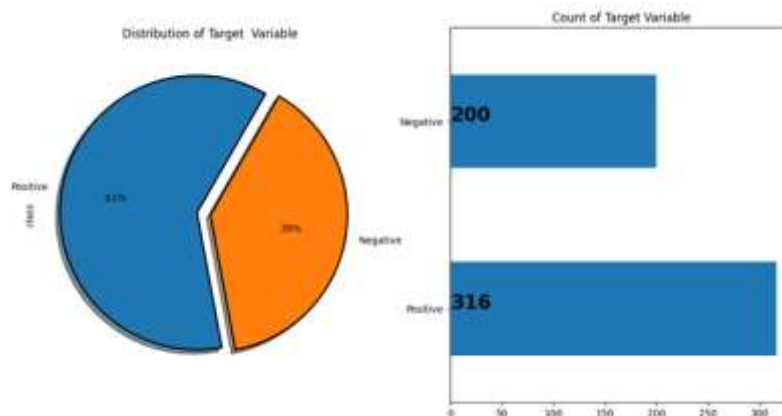
Penelitian ini menggunakan data sekunder yang diperoleh melalui media perantara, yaitu dari *UC Irvine Machine Learning Repository* (Rini, n.d.). Data tersebut berasal dari kuesioner yang diisi oleh pasien diabetes di *Sylhet Diabetes Hospital* di Bangladesh dan telah diverifikasi oleh dokter. Penelitian ini dilakukan karena adanya masalah diabetes, yang dapat mempengaruhi siapa saja dan jika tidak ditangani dengan baik dapat berakibat fatal (Putri, 2020). Masalah tersebut kemudian diidentifikasi, dirumuskan, dan menjadi tujuan penelitian ini. Selanjutnya, dengan menggunakan data sekunder, dilakukan pemilihan data untuk memastikan kesesuaian dengan kebutuhan penelitian.

Tujuan utama klasifikasi adalah mengembangkan model atau aturan yang dapat memprediksi kelas atau label yang tepat untuk setiap objek baru berdasarkan pemahaman dari data latih. Proses klasifikasi data terdiri dari dua tahapan. Tahapan pertama adalah learning, di mana data pelatihan dianalisis menggunakan algoritma klasifikasi. Tahapan kedua adalah classification, di mana data uji digunakan untuk mengestimasi akurasi dari aturan klasifikasi.

HASIL DAN PEMBAHASAN

Tahapan pertama yang dilakukan setelah import data berdasarkan tahapan penelitian adalah pembersihan data. Pada tidak terdapat missing value dan terdapat 4 data outlier yang

diatasi dengan cara dihapus dari data (Ramadhan, Pertiwi, Defanka, & Sari, 2023). Pengujian data duplikat tidak dilakukan karena tidak ada atribut yang dapat digunakan untuk mengecek adanya duplikat pada data secara spesifik (Santoso, 2020). Setelah pembersihan data, jumlah data yang akan digunakan untuk penelitian terdapat 516 data dengan 17 atribut dengan sebaran positif dan negative seperti pada gambar. Setelah melakukan pembersihan data, selanjutnya akan dilakukan pre-processing data untuk mengetahui sebaran datanya yang tertera pada gambar 1.



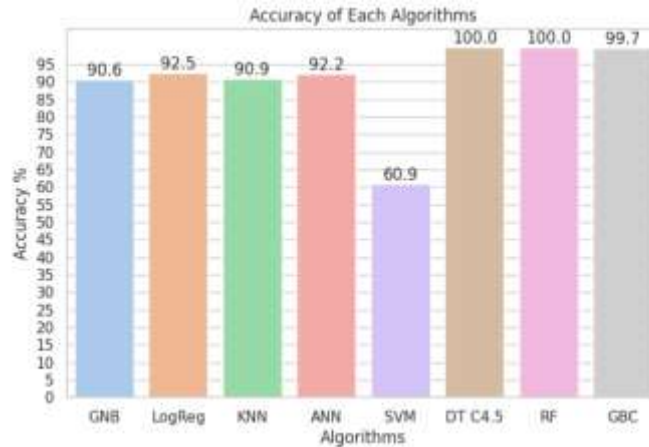
Gambar 1
Visualisasi Sebaran Data Pasien

Berdasarkan visualisasi yang dilakukan di tahap *pre-processing* data pada gambar 1, perbandingan pasien yang positif dan negatif diabetes adalah 61% dan 39% dengan 316 pasien positif dan 200 negatif. Langkah akhir dalam *pre-processing* data adalah mengubah variabel kategorikal ke bentuk numerikal dan membagi data menjadi training dan testing untuk memastikan keakuratan dalam pengujian (Hasibuan, 2022). Pada penelitian ini, data dibagi dengan perbandingan 70:30 mengikuti ukuran data yang dimiliki. Dari hasil pembagian didapatkan 361 data training dan 155 data testing. Dimana keduanya memiliki 61% pasien positif diabetes dan 39% pasien negatif diabetes. Hal ini menunjukkan bahwa pembagian data telah mempertahankan proporsi yang seimbang antara kategori positif dan negatif diabetes dalam kedua subset data.

Pembentukan Model

Setelah *pre-processing* data dilakukan, maka data sudah siap digunakan untuk pembuatan model. Terdapat 8 Model yang dibentuk dan kemudian dari model yang dibentuk dipilih 5 model dengan tingkat akurasi tertinggi untuk dilanjutkan kepengujian model. Hasil yang didapatkan dari pembentukan model terdapat pada gambar 2.

Analisis Klasifikasi Diagnosa Penyakit Diabetes Melitus Berdasarkan Komparasi Algoritma Supervised Learning

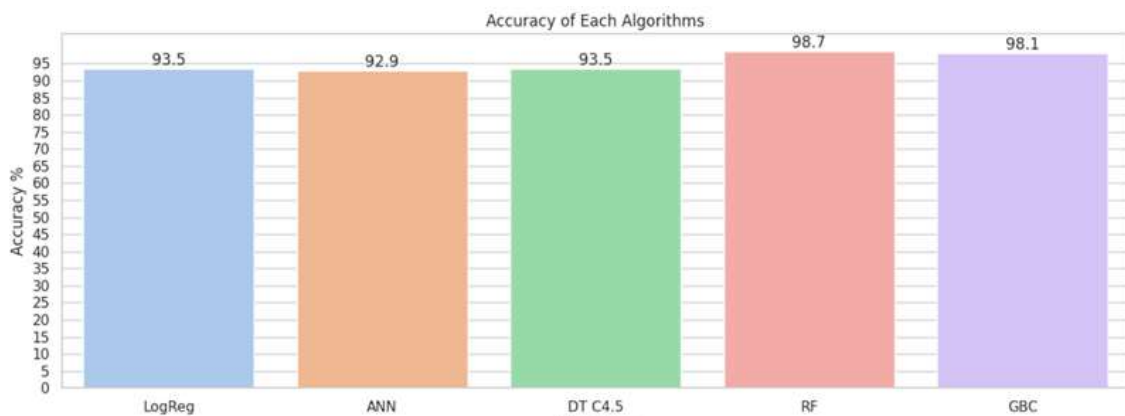


Gambar 2
Perbandingan Tingkat Akurasi Pembentukan Model

Pada visualisasi akurasi model training, didapatkan akurasi untuk *Gaussian Naïve Bayes* adalah 90,6%, *Logistic Regression* 92,5%, *K-Nearest Neighbor* 90,9%, *Artificial Neural Network* 92,2%, *Support Vector Machine* 60,9%, *Algoritma C4.5*, dan *Random Forest* 100%, dan *Gradient Boosting Classifier* 99,7%. Dari hasil tersebut didapatkan lima algoritma dengan akurasi tertinggi sesuai urutannya adalah *Algoritma C4.5*, *Random Forest*, *Gradient Boosting Classifier*, *Logistic Regression*, dan *Artificial Neural Network*. Selanjutnya, ke lima model di uji menggunakan data testing untuk mendapatkan model terbaik dalam mendiagnosa penyakit diabetes (Wiratama & Pradnya, 2022).

Pengujian dan Perbandingan Model

Pengujian terhadap model-model yang sudah dibentuk menggunakan data training dilakukan untuk mendapatkan model terbaik yang dapat digunakan untuk memprediksi penyakit diabetes melitus. Hasil yang didapat dari pengujian model dilihat hasilnya dalam bentuk *confusion matrix* dan kurva ROC yang menunjukkan akurasi hasil prediksi dari setiap model dengan hasil perbandingan seperti gambar 3.



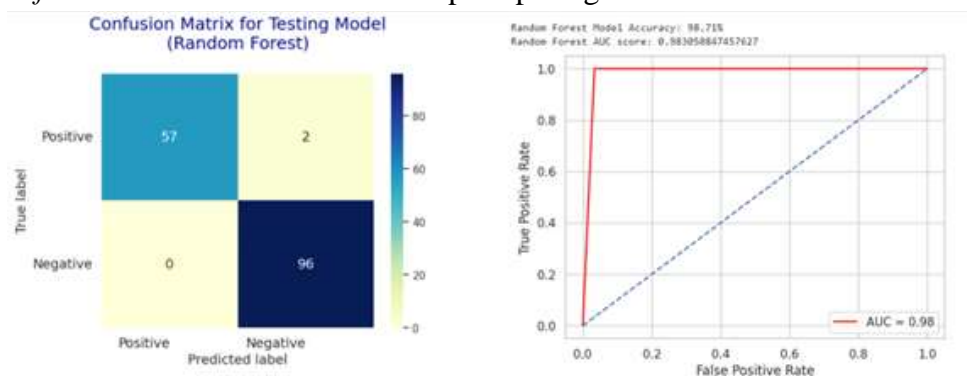
Gambar 3
Perbandingan Hasil Pengujian Model

Diagram batang pada gambar 5 menunjukkan tingkat akurasi untuk *Logistic Regression* adalah 93,5%; *Artificial Neural Network* 94,5%; *Decision Tree C4.5* 93,5%; *Random Forest* 98,7%; dan *Gradient Boosting Classifier* 98,1%. Maka dapat disimpulkan dari hasil pengujian

tersebut, model terbaik dengan tingkat akurasi tertinggi untuk memprediksi penyakit diabetes adalah *Random Forest*, dengan akurasi sebesar 98,71% yang berarti dapat memprediksi dengan sangat akurat dan nilai AUC 0,98305 yang menunjukkan model ini memiliki kemampuan yang sangat baik untuk membedakan kelas positif dan negatif. Namun, model-model lainnya seperti *Gradient Boosting Classifier*, *Decision Tree C4.5*, *Artificial Neural Network*, dan *Logistic Regression* juga memberikan hasil yang baik dengan tingkat akurasi di atas 90%.

Perbandingan Model Terbaik sebelum dan setelah *Feature Selection*

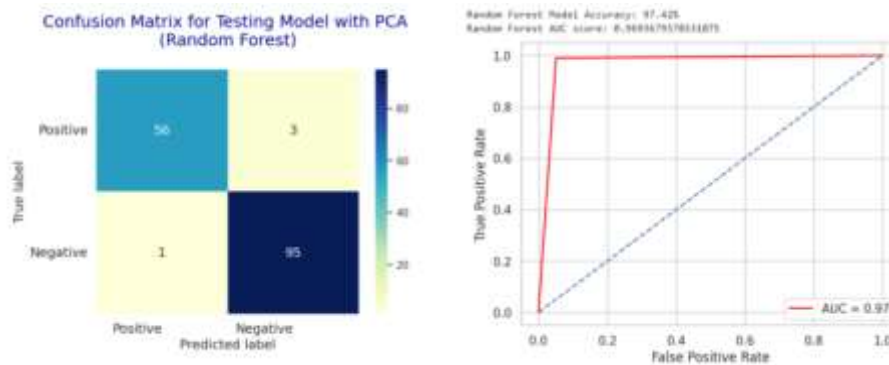
Berdasarkan hasil dari pengujian dan perbandingan model, model terbaik yang dapat digunakan untuk memprediksi penyakit diabetes melitus adalah *Random Forest Classifier* dengan *confusion matrix* dan ROC curve seperti pada gambar 4.



Gambar 4
Hasil Pengujian Model *Random Forest*

Confusion matrix dengan algoritma random forest menunjukkan terdapat 57 pasien positif dan 98 pasien negatif. Dapat disimpulkan terdapat 2 hasil yang error dimana pasien yang seharusnya positif terprediksi negatif. Hasil error ini disebut sebagai *False Negative* (FN) dimana model gagal mengklasifikasikan pasien yang sebenarnya positif dengan benar. Kurva ROC menunjukkan nilai AUC algoritma random forest untuk data ini adalah 0,98305. Nilai AUC yang tinggi ini menunjukkan model ini memiliki kemampuan yang sangat baik untuk membedakan antara kelas positif dan negatif. Selain itu, tingkat akurasi prediksi sebesar 98,71% menunjukkan bahwa model secara keseluruhan memberikan prediksi yang akurat.

Dalam rangka mengurangi ruang penelitian, dilakukan *principal component analysis* untuk mengetahui jumlah variabel yang paling berpengaruh untuk memprediksi diabetes (Satriatama et al., 2023). Dalam kasus ini, didapatkan setidaknya 8 komponen digunakan dengan *variance* sebesar 0,995 yang didapat dari *principal component analysis*, yang berarti komponen-komponen tersebut mampu menjelaskan sekitar 99,5% variasi dalam data. Setelah mendapatkan komponen-komponen utama, dilakukan *feature selection* untuk menentukan variabel-variabel yang paling berpengaruh terhadap hasil prediksi positif dan negatif diabetes (Adinugroho & Sari, 2018). Dalam kasus ini, didapatkan 8 variabel yang paling berpengaruh, yaitu: *Age*, *Gender*, *Polyuria*, *Polydipsia*, *sudden weight loss*, *Irritability*, *Partial Paresis*, dan *Alopecia*. Maka selanjutnya dibentuk *confusion matrix* dan kurva ROC, didapatkan hasil yang ditunjukkan pada gambar 4. Hal ini dilakukan untuk membandingkan hasil prediksi sebelum dan setelah *feature selection*.



Gambar 5
Hasil Pengujian Setelah *Feature Selection*

Dari hasil *confusion matrix* dan kurva ROC setelah menerapkan *feature selection* pada model *random forest*, terlihat bahwa terdapat 4 hasil yang salah klasifikasi. Dari 4 hasil tersebut, 3 hasil seharusnya positif tetapi diprediksi sebagai negatif (*False Negative*), dan 1 hasil seharusnya negatif tetapi diprediksi sebagai positif (*False Positive*). Jumlah pasien yang terprediksi positif adalah 57 dan jumlah pasien yang terprediksi negatif adalah 98. Kurva ROC menunjukkan nilai AUC (*Area Under the Curve*) algoritma *random forest* setelah menerapkan *feature selection* sebesar 0,96937. Nilai AUC yang tinggi ini menunjukkan bahwa model setelah PCA dan *feature selection* masih memiliki kemampuan yang baik dalam membedakan antara kelas positif dan negatif (Irawan & Samopa, 2018). Tingkat akurasi prediksinya sebesar 97,42%, yang menunjukkan bahwa model setelah PCA dan *feature selection* memberikan prediksi yang akurat.

Keakuratan hasil yang didapat setelah menggunakan PCA dan *feature selection* untuk memprediksi penyakit diabetes turun sekitar 1% dari model sebelumnya. Meskipun terdapat sedikit penurunan akurasi dibandingkan dengan model sebelumnya, penggunaan PCA dan *feature selection* dapat membantu dalam mengurangi dimensi data dan mempertahankan informasi penting dari variabel yang paling berpengaruh (Danardono & Fikriyah, 2021). Hal ini dapat membantu dalam meningkatkan efisiensi dan interpretabilitas model, serta mengurangi *overfitting*.

KESIMPULAN

Penelitian ini menggunakan data pasien diabetes melitus yang terdiri dari 520 data awal, tetapi setelah proses pembersihan data, jumlahnya menjadi 516 data yang siap digunakan untuk tahap penelitian. Pada tahap pembentukan model, delapan algoritma supervised learning diaplikasikan, dan lima model dengan akurasi tertinggi yaitu *Decision Tree C4.5*, *Random Forest*, *Gradient Boosting Classifier*, *Logistic Regression*, dan *Artificial Neural Network* dipilih untuk pengujian lebih lanjut menggunakan data testing. Hasil pengujian menunjukkan bahwa semua model memiliki tingkat akurasi yang tinggi, dan model *Random Forest* dan *Gradient Boosting Classifier* mencapai akurasi tertinggi, yaitu 98,71% dan 98,1% secara berurutan. Berdasarkan tingkat akurasi dan nilai AUC, model *Random Forest* dinilai sebagai model terbaik dalam memprediksi diabetes, karena memiliki nilai AUC 0,98305 yang menunjukkan kemampuan yang baik dalam membedakan kelas positif dan negatif. Meskipun setelah menerapkan PCA dan *Feature Selection* pada model *Random Forest*, terdapat sedikit

penurunan akurasi menjadi 97,42%, namun model tersebut masih mampu membedakan kelas dengan baik.

Penelitian ini memberikan kontribusi dalam pengembangan prediksi diabetes melitus berdasarkan gejala yang dialami oleh pasien. Peneliti selanjutnya dapat menggunakan metode yang sama dengan dataset yang berbeda untuk memvalidasi hasil penelitian ini. Selain itu, penelitian dapat dilakukan dengan menggunakan algoritma lain untuk memprediksi diabetes melitus dan membandingkan hasilnya. Hasil dari penelitian ini diharapkan dapat digunakan sebagai dasar untuk membuat aplikasi pendeteksi dini penyakit diabetes melitus berdasarkan gejala yang dialami oleh pasien, sehingga dapat membantu masyarakat dalam memahami risiko dan melakukan tindakan pencegahan yang tepat. Dengan meningkatnya kesadaran tentang pentingnya pencegahan dan deteksi dini penyakit diabetes melitus, aplikasi berbasis *machine learning* seperti yang dihasilkan dari penelitian ini dapat menjadi alat yang berharga dalam mendukung upaya kesehatan masyarakat.

DAFTAR PUSTAKA

- Adinugroho, Sigit, & Sari, Yuita Arum. (2018). *Implementasi data mining menggunakan WEKA*. Universitas Brawijaya Press. [Google Scholar](#)
- Anggoro, Dimas Aryo, & Novitaningrum, Dian. (2021). Comparison of accuracy level of support vector machine (SVM) and artificial neural network (ANN) algorithms in predicting diabetes mellitus disease. *ICIC Express Letters*, 15(1), 9–18. [Google Scholar](#)
- Atlas, I. D. F. Diabetes. (2021). Доступно по: <https://diabetesatlas.org>. Ссылка Активна Ha, 23, 2021. [Google Scholar](#)
- Danardono, S. Si, & Fikriyah, Vidya N. (2021). *Sistem Informasi Geografis dan Aplikasinya di Bidang Geografi*. Muhammadiyah University Press. [Google Scholar](#)
- Deberneh, Henock M., & Kim, Intaek. (2021). Prediction of type 2 diabetes based on machine learning algorithm. *International Journal of Environmental Research and Public Health*, 18(6), 3317. [Google Scholar](#)
- Hasibuan, Feny Iryani. (2022). Algoritma C. 45 Dalam Menganalisis Faktor Penyebab Mahasiswa Tidak Berkontribusi Menyampaikan Pendapat Saat Proses Pembelajaran. *SkripsiKu-2022*, 1(1). [Google Scholar](#)
- HR, Hakim Tobroni, Pratiwi, Tiara Fatma, & Susanti, Novia. (2021). *Cara Jitu Mengatasi Diabetes Mellitus dengan Teknik Komplementer*. Penerbit NEM. [Google Scholar](#)
- Irawan, Ferry, & Samopa, Febriliyan. (2018). A comparative assessment of the random forest and SVM algorithms using combination of principal component analysis and SMOTE for accounts receivable seamless prediction case study company X in surabaya. *Proceedings of the 2nd International Seminar Of Contemporary ONBUSINESS Research & Management. University Warmadewa Bali*. [Google Scholar](#)
- Ismah, Zata. (2018). *Bahan Ajar Dasar Epidemiologi*. [Google Scholar](#)
- Kazerouni, Faranak, Bayani, Azadeh, Asadi, Farkhondeh, Saeidi, Leyla, Parvizi, Nasrin, & Mansoori, Zahra. (2020). Type2 diabetes mellitus prediction using data mining algorithms based on the long-noncoding RNAs expression: a comparison of four data mining approaches. *BMC Bioinformatics*, 21, 1–13. [Google Scholar](#)
- Kemenkes, R. I. (2019). Tanda dan gejala Diabetes Melitus. *P2PTM Kemenkes RI*. [Google Scholar](#)

Analisis Klasifikasi Diagnosa Penyakit Diabetes Melitus Berdasarkan Komparasi Algoritma Supervised Learning

- Masenda, Bobbie Snow. (2023). *A 6-Week Diabetic Self-Management Program on Reducing Hemoglobin A1c Levels, Increasing Patients Knowledge and Attitudes Toward Diabetes*. University of Massachusetts Global.[Google Scholar](#)
- Putri, Rizky Pratama. (2020). *Pengaruh Slow Stroke Back Massage (SSBM) terhadap Depresi pada Penderita Diabetes Melitus Tipe 2 di Wilayah Kerja PUSKESMAS Juanda Kota Samarinda*.[Google Scholar](#)
- Putry, Naisah Marito. (2022). Komparasi algoritma knn dan naïve bayes untuk klasifikasi diagnosis penyakit diabetes mellitus. *Evolusi: Jurnal Sains Dan Manajemen*, 10(1).[Google Scholar](#)
- Ramadhan, Humam Maulana Tsubasanofa, Pertiwi, Aisyah, Defanka, Galan Ahmad, & Sari, Anggraini Puspita. (2023). Clustering RFM (Recency, Frequency, Monetary) Publisher Gim Menggunakan Algoritma K-Means. *Prosiding Seminar Nasional Informatika Bela Negara*, 3, 8–18.[Google Scholar](#)
- Ridwan, Achmad. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *J. SISKOM-KB (Sistem Komput. Dan Kecerdasan Buatan)*, 4(1), 15–21.[Google Scholar](#)
- Rini, Afif Sulfa. (n.d.). *Analisis publikasi artikel perpajakan pada beberapa jurnal internasional di bidang akuntansi tahun 2010-2014*.[Google Scholar](#)
- Santoso, Singgih. (2020). *Panduan Lengkap SPSS 26*. Elex Media Komputindo.[Google Scholar](#)
- Satriatama, Ananda Elang, Wibowo, Ari Prasetyo, Arnold, I. Gusti Ngurah, Pratama, Reyhan Bayu, Masyhuda, Tegar Alwinata, Agusti, Yohannes Alexander, Purwanti, Endah, & Werdiningsih, Indah. (2023). Analisis Klaster Data Pasien Diabetes untuk Identifikasi Pola dan Karakteristik Pasien. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(3), 172–182.[Google Scholar](#)
- Wiratama, Muhammad Abid, & Pradnya, Windha Mega. (2022). Optimasi algoritma data mining menggunakan backward elimination untuk klasifikasi penyakit diabetes. *Jurnal Nasional Pendidikan Teknik Informatika: Janapati*, 11(1), 1–12.[Google Scholar](#)