



Sistem Informasi Model Prediksi Penyakit Diabetes Menggunakan Algoritma Random Forest di Fasyankes

Muhadi^{1*}, Amir Ali²

¹D4 MIK Politeknik Medica Farma Husada Mataram
Tanjung Karang, Mataram, Indonesia
muhadimuzani15@gmail.com

²STIKES Yayasan RS Dr. Soetomo,
Kalijudan, Surabaya, Indonesia
amir.consulting@gmail.com

*Corresponding Author

Abstrak:

Selama ini fasyankes hanya melakukan penyimpanan data transaksi pemeriksaan terhadap pasien diabetes menggunakan aplikasi sistem informasi. Data yang telah terkumpul hanya disimpan dalam database sistem informasi di komputer server. Belum ada analisa mendalam terkait data yang tersimpan dalam server database di aplikasi sistem informasi kesehatan. Tujuan dari penelitian ini yaitu mengembangkan sistem informasi model prediksi penyakit diabetes dengan menggunakan algoritma random forest di fasyankes. Menggunakan metode machine learning dengan algoritma random Forest. Pembagian data pada penelitian ini dilakukan dengan komposisi 80 % data training dan 20 % data testing. Hasil dari pembagian data ini menghasilkan data training sebesar 615 dan data testing sebesar 153 dari total data keseluruhan yang berjumlah 768. menggunakan data dan atribut secara acak sesuai dengan persyaratan. Hasil dari klasifikasi ini menghasilkan model klasifikasi pasien penderita diabetes dengan tingkat akurasi sebesar 76%, dengan nilai mean squares error sebesar 0,24. Penelitian ini berhasil membuat model prediksi untuk penderita penyakit diabetes dengan menginputkan variabel yang menyebabkan penyakit diabetes. Tingkat akurasi yang dihasilkan oleh model ini yaitu sebesar 76% dengan nilai mean square error sebesar 0,24. Dimana model ini menghasilkan nilai yang cukup baik dalam memprediksi pasien masuk dalam kategori diabetes atau tidak. Hasil evaluasi kinerja model yang dihasilkan menghasilkan nilai akurasi sebesar 76%, Presisi sebesar 69%, Recall sebesar 59%, F1-Score sebesar 63%

Kata Kunci:

Algoritma random forest, Machine learning, Akurasi, Prediksi, Mean square error

Abstract:

So far, health facilities only store data on examination transactions for diabetes patients using information system applications. The data that has been collected is only stored in the information system database on the server computer. There has been no in-depth analysis regarding the data stored in the database server in the health information system application. The aim of this research is to develop an information system for diabetes prediction models using the random forest algorithm in health facilities. Using machine learning methods with the random Forest algorithm. Data sharing in this research was carried out with a composition of 80% training data and 20% testing

data. The results of this data sharing produced 615 training data and 153 testing data out of a total of 768 data. Using random data and attributes according to the requirements. The results of this classification produce a classification model for diabetes patients with an accuracy level of 76%, with a mean squares error value of 0.24. This research succeeded in creating a prediction model for diabetes sufferers by inputting variables that cause diabetes. The level of accuracy produced by this model is 76% with a mean square error value of 0.24. Where this model produces quite good values in predicting whether a patient is in the diabetes category or not. The results of the performance evaluation of the resulting model produce an accuracy value of 76%, Precision of 69%, Recall of 59%, F1-Score of 63%

Keywords:

Random forest algorithm, Machine learning, Accuracy, Prediction, Mean square error

1. Pendahuluan

Diabetes mellitus merupakan salah satu penyakit kronis yang menjadi ancaman kesehatan masyarakat secara global. Penyakit ini memiliki dampak serius terhadap kualitas hidup pasien serta menimbulkan beban ekonomi yang signifikan pada sistem kesehatan. Penyakit ini berhubungan dengan peningkatan glukosa darah yang bersirkulasi. Tingkat penyakit ini, kini menyebar dengan cepat dan kejadian global terus meningkat selama beberapa dekade terakhir[1]. Di Indonesia, prevalensi diabetes terus meningkat setiap tahunnya, yang sebagian besar disebabkan oleh gaya hidup yang tidak sehat, pola makan, dan kurangnya aktivitas fisik. Hal ini contohnya bisa ditemukan pada kasus jumlah penderita diabetes mellitus tipe 2 pada anak dan remaja dimana Jumlah diabetes mellitus tipe 2 pada anak-anak dan remaja telah meningkat sebesar 33% dalam 15 tahun terakhir[2]

Prediksi dini terhadap kemungkinan seseorang menderita diabetes dapat membantu dalam pengambilan keputusan klinis, termasuk penanganan dan pencegahan komplikasi. Salah satu metode yang digunakan untuk memprediksi penyakit ini adalah algoritma pembelajaran mesin (machine learning). Random Forest sebagai algoritma pembelajaran mesin berbasis pohon keputusan, telah terbukti efektif dalam memproses data besar dan kompleks serta memiliki akurasi tinggi dalam memprediksi hasil[3]

Penelitian ini bertujuan untuk membangun sistem informasi model prediksi penyakit diabetes dengan menggunakan algoritma Random Forest berdasarkan database pasien diabetes di fasyankes. Sistem informasi ini merupakan pengembangan sistem informasi prediksi dengan menggunakan algoritma random forest. Sistem informasi prediksi yang serupa tetapi menggunakan algoritma lain seperti algoritma Support Vector Machine (SVM) pernah dibuat, seperti penelitian yang dilakukan oleh [4]. Sistem pakar untuk memprediksi penyakit diabetes dengan pendekatan certainty factor seperti yang dilakukan oleh [5]. Penelitian yang dilakukan oleh annisa dkk terkait prediksi penyakit hipertensi menggunakan machine learning dengan algoritma regresi logistik[6]. Penelitian lainnya yang dilakukan oleh Nur tri dkk terkait analisis perbandingan ensemble machine learning dengan teknik smote untuk prediksi diabetes[7]. Pada penelitian ini akan dibuatkan sistem informasi prediksi dari hasil pemodelan dengan pendekatan algoritma random forest dimana diharapkan dapat menjadi alat bantu yang andal bagi tenaga medis dalam proses diagnosis.

1.1. Kajian Masalah

Masalah dalam penelitian ini yaitu belum adanya analisa mendalam terhadap data pasien terutama pasien diabetes yang tersimpan dalam database server di fasyankes. Penggalan informasi terkait big data yang tersimpan dalam database server fasyankes ini sangatlah penting dimana Big Data, dengan kapasitasnya untuk menangani volume data yang besar, kompleksitas tinggi, dan kecepatan pemrosesan yang optimal, telah merevolusi cara data kesehatan dikumpulkan, disimpan, dan dikelola. Pemanfaatan teknologi ini memberikan keunggulan dalam analisis data kesehatan yang mendalam dan menyeluruh. Data besar ini mencakup informasi dari berbagai sumber, seperti catatan medis elektronik, hasil uji laboratorium, hingga data genetik, memberikan pemahaman yang lebih komprehensif tentang kondisi kesehatan individu dan populasi [8]. Tenaga kesehatan dapat memberikan edukasi terhadap pasien terkait penyakit yang diderita pasien terutama bagi pasien diabetes. Pengambilan keputusan klinis yang dihasilkan oleh tenaga kesehatan didasarkan pada data informasi dari analisa data pasien yang tersimpan dalam database komputer server di fasyankes. Pendekatan analisa data ini dapat menggunakan pendekatan teknologi pembelajaran mesin atau yang kita sebut dengan machine learning.

Perkembangan teknologi saat ini khususnya di bidang machine learning sangat bermanfaat dalam dunia kesehatan, dimana dokter dan tim kesehatan lainnya mempunyai kesempatan untuk meningkatkan kualitas kerja tim medis dengan mengimplementasikan model klasifikasi diabetes menggunakan algoritma random forest sebagai salah satu algoritmanya. Algoritma random forest mempunyai model pengembangan yang lebih sesuai sehingga tim medis dapat memanfaatkan untuk mengenali gejala awal penyakit diabetes ini, sehingga kedepannya pasien yang sudah mengalami gejala dapat ditangani lebih cepat dan mencegahnya terkena penyakit lainnya[9]

Pada penelitian ini kami mencoba mengembangkan suatu sistem informasi untuk pemodelan prediksi penyakit diabetes dengan menggunakan algoritma random forest sebagai salah satu algoritma dari machine learning. Dengan adanya sistem informasi ini tentu diharapkan dapat mempermudah tenaga kesehatan dalam mengambil keputusan medis terbaik buat pasien terutama pasien diabetes

1.2. Pembatasan Masalah

Pada penelitian ini kami membatasi permasalahan penelitian dimana data yang digunakan adalah dataset pima indian diabetes yang kami download dari internet sebagai contoh dataset umum yang digunakan untuk keperluan uji coba sistem informasi pemodelan prediksi penyakit diabetes yang telah digunakan. Sistem informasi ini dapat digunakan untuk fasyankes yang membutuhkan. Jika komposisi datasetnya berubah, maka sistem informasi ini dapat diubah inputannya sesuai dengan kebutuhan analisa data oleh fasyankes

2. Metode

Penelitian ini menggunakan metode klasifikasi dengan menggunakan algoritma random forest. Dimana random forest adalah pengembangan dari metode decision tree yang menggunakan beberapa decision tree, dimana setiap decision tree telah dilakukan pelatihan menggunakan sampel individu dan setiap atribut dipecah pada pohon yang dipilih antara atribut subset yang bersifat acak[10]. Random forest mempunyai beberapa keunggulan, yaitu dapat meningkatkan akurasi apabila terdapat data yang hilang serta untuk resisting outliers, dan juga efisien untuk penyimpanan data. Tidak hanya itu, pada Random Forest terdapat proses seleksi fitur dimana mampu mengambil fitur terbaik sehingga meningkatkan performa pada model klasifikasi. Dengan adanya fitur seleksi tentunya Random Forest

mampu bekerja pada data yang besar dengan parameter yang kompleks secara efektif [11]. Random Forest (RF) pertama kali diperkenalkan oleh Leo Breiman (2001). RF merupakan salah satu metode yang dapat meningkatkan hasil akurasi dalam membangkitkan atribut untuk setiap node yang dilakukan secara acak. RF terdiri dari sekumpulan decision tree, dimana kumpulan pohon keputusan ini digunakan untuk mengklasifikasi data ke suatu kelas. Membentuk pohon keputusan pada metode RF sama dengan proses pada Classification and Regression Tree (CART), hanya saja pada RF tidak dilakukan pruning (pemangkasan) [12].

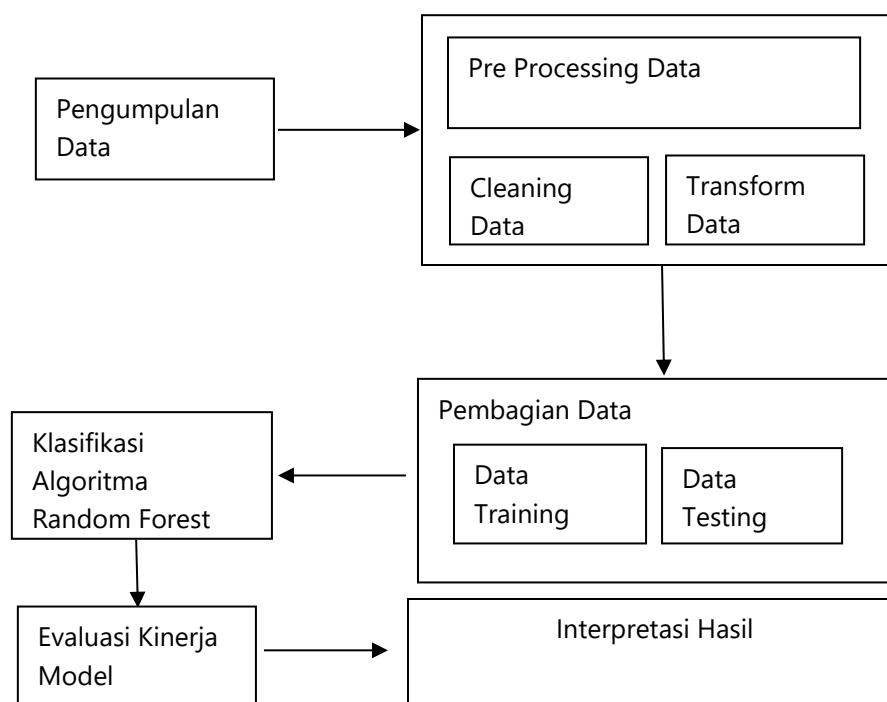
Indeks Gini digunakan untuk memilih fitur di setiap simpul internal dari pohon keputusan. Nilai Indeks Gini dapat dihitung sebagai berikut:

$$Gini(S_i) = 1 - \sum_{i=0}^{c-1} p_i^2 \tag{1}$$

Dengan p_i merupakan frekuensi relative kelas C_i di dalam set C_i merupakan n kelas untuk $i = 1, \dots, c-1$, dan c adalah jumlah kelas yang telah ditentukan. Kualitas split pada fitur k ke dalam subset S_i merupakan jumlah sampel milik kelas C_i , kemudian dihitung sebagai jumlah pertimbangan indikasi Gini dari subset yang dihasilkan. Data dapat dihitung dengan rumus sebagai berikut :

$$Gini_{split} = \sum_{i=0}^{k-1} \left(\frac{n_i}{n}\right) Gini(S_i) \tag{2}$$

Dimana n_i merupakan jumlah sampel dalam subset S_i setelah di split dan n merupakan jumlah sampel di node yang diberikan.



Sumber : Peneliti (2024)

Gambar 1: Kerangka Operasional Penelitian

Pada gambar 1 diatas diawali dari proses pengumpulan data. Pengumpulan data dari penelitian ini menggunakan data publik yang diambil dari kaggle website untuk data Pima Indian Diabetes Database. Data ini sebagai data model yang dapat kita gunakan untuk melakukan simulasi dalam pembuatan sistem informasi model prediksi penyakit diabetes. Tahapan selanjutnya adalah tahapan preprocessing

dimana terdapat 2 tahap didalamnya yaitu cleaning data dan transform data. pada tahapan cleaning data dimana data akan dicek apakah ada data yang hilang (missing data). jika ada data yang hilang atau kurang lengkap nantinya dapat digunakan pendekatan teknik imputansi selain itu dilakukan normalisasi data untuk memastikan bahwa variabel dengan skala yang berbeda dapat dibandingkan secara adil sedangkan pada tahapan transform data dimana pada tahapan ini, tipe jenis dari data akan diubah menjadi tipe data numerik agar bisa diproses.

Pentahapan selanjutnya yaitu proses pembagian data. Dimana dataset yang telah diproses dan dibersihkan sebelumnya akan dibagi menjadi dua bagian pada tahap pembagian data. Data pelatihan (training) akan digunakan untuk menentukan nilai kekuatan model, dan data tes (testing) akan digunakan untuk mengevaluasi hasil dari nilai kekuatan model [13]. Salah satu faktor yang memengaruhi seberapa baik model klasifikasi berfungsi pada algoritma pembelajaran mesin adalah model splitting data yang digunakan untuk mempartisi dataset[14]. Kami memilih pembagian data terbesar, 80:20, dalam proses penelitian ini. Untuk menghindari overfitting dan supaya memberikan estimasi performa model yang lebih stabil, maka dilakukan teknik k-fold validation.

Prediksi dilakukan dengan menggunakan algoritma Klasifikasi yaitu algoritma random forest. Metode ini digunakan untuk membangun pohon keputusan yang terdiri dari root node, internal node, dan leaf node dengan menggunakan data dan atribut secara acak sesuai dengan persyaratan. Simpul anak dibangun secara acak untuk setiap node, meningkatkan akurasi hasil. Pembentukan algoritma ini ditentukan oleh nilai vektor acak yang independen dan seragam di setiap pohon. Metode klasifikasi adalah yang paling akurat saat membuat prediksi. Ini dapat menemukan perbandingan antar hutan acak seperti properti metode ansambel dan dapat menangani banyak variabel input tanpa overfitting. Untuk evaluasi hasil prediksi dari model yang dihasilkan akan dilakukan dengan menggunakan matrix evaluasi. Dalam penelitian ini, lima komponen terdiri dari laporan metrix evaluasi: akurasi, ketepatan, recall, skor F1. Pada tahapan interpretasi hasil, algoritma random forest akan digunakan untuk menganalisis hasil dari proses pengklasifikasian risiko diabetes

3. Hasil dan Pembahasan

3.1. Hasil

Data diambil dari laman kaggle.com untuk data pasien diabetes yang berjumlah 768 pasien dari pima indians diabetes database. Data ini terdiri dari variabel :

1. Pregnancies
2. Glucose
3. BloodPressure
4. SkinThickness
5. Insulin
6. BMI (Body Mass Index)
7. DiabetesPedigreeFunction
8. Age

Pembagian data dengan komposisi 80 % data training dan 20 % data testing. Hasil dari pembagian data ini menghasilkan data training sebesar 615 dan data testing sebesar 153 dari total data keseluruhan yang berjumlah 768. Hasil dari klasifikasi ini menghasilkan model klasifikasi pasien penderita diabetes dengan tingkat akurasi sebesar 76%, dengan nilai mean squares error sebesar 0,24.

Dengan memperhatikan nilai dari tabel confusion metrik untuk evaluasi model maka model yang dikembangkan menghasilkan tabel confusion metrik sebagai berikut :

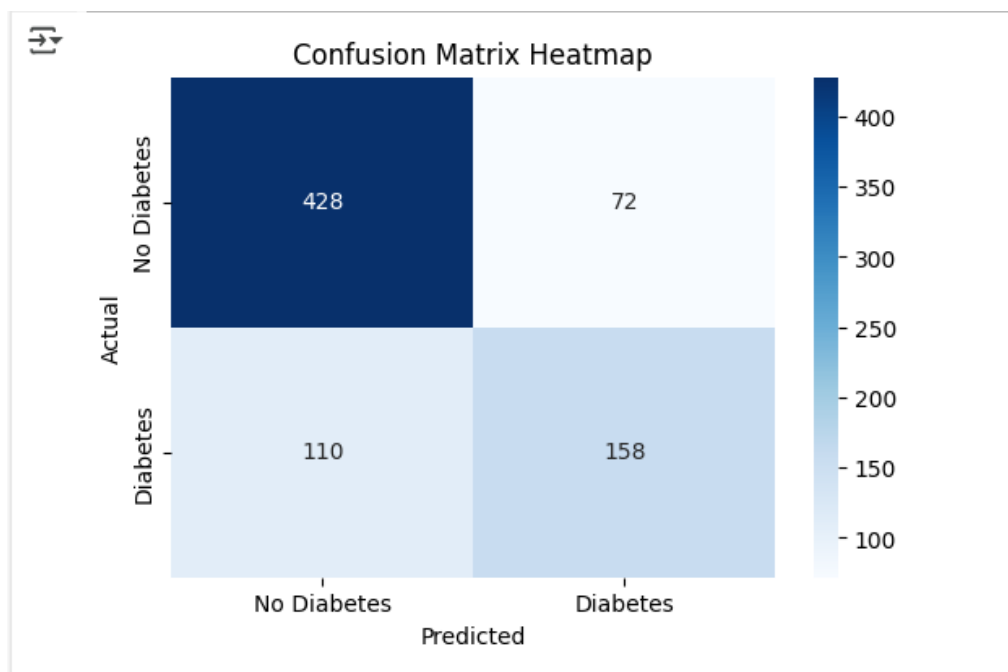
Tabel 1: Tabel Confusion Matrix

		Kenyataan Positif	Kenyataan Negatif
Prediksi Positif		158	72
Prediksi Negatif		110	428

Dimana hasil dari interpretasi tabel diatas yaitu :

- a. True Positives(TP) : Model memprediksi positif dengan benar sebanyak 158.
- b. False Positives (FP) : Model memprediksi diabetes, padahal sebenarnya tidak diabetes sebanyak 72.
- c. False Negatives (FN) : Model memprediksi tidak diabetes, padahal sebenarnya diabetes sebanyak 110.
- d. True Negatives (TN) : Model memprediksi negatif dengan benar sebanyak 428.

Visualisasi dari tabel confusion matrix diatas sebagai berikut :



Gambar 2: Visualisasi Tabel Confusion Matri

a. Akurasi, dimana dengan rumus perhitungan sebagai berikut :

$$\begin{aligned} \text{Akurasi} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ \text{Akurasi} &= \frac{158 + 428}{158 + 428 + 72 + 110} \\ &= 0,763021 \\ &= 0,763021 \times 100\% = 76\% \end{aligned}$$

b. Presisi

$$\begin{aligned}\text{Presisi} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ &= \frac{158}{158 + 72} \\ &= 0,6869 \times 100\% \\ &= 69 \%\end{aligned}$$

c. Recall (Sensivitas/ tingkat deteksi)

$$\begin{aligned}\text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ &= \frac{158}{158 + 110} \\ &= 0,5895 \times 100 \% \\ &= 59 \%\end{aligned}$$

d. F1-Score

$$\begin{aligned}\text{F1-Score} &= 2 \times \left(\frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \right) \\ &= 2 \times \left(\frac{0,687 \times 0,59}{0,687 + 0,59} \right) \\ &= 0.6348 \times 100\% \\ &= 63 \%\end{aligned}$$

Keterangan :

- a. True positive (TP)
- b. True negative (TN)
- c. False positive (FP)
- d. False negative (FN)

Hasil dari tampilan sistem informasi prediksi penderita diabetes sebagai berikut :

The screenshot displays a web browser window with the title 'Prediksi Diabetes'. The address bar shows the URL '127.0.0.1:5000/predict'. The main content area is titled 'Form Prediksi Diabetes' and contains several input fields with the following values: Pregnancies: 6, Glucose: 148, Blood Pressure: 72, Skin Thickness: 35, Insulin: 0, BMI: 33.6, Diabetes Pedigree Function: 0.627, and Age: 50. Below these fields is a 'Prediksi' button. Underneath the button, the text 'Hasil Prediksi: Diabetes' is displayed. In the bottom right corner of the browser window, there is a watermark that says 'Activate Windows Go to Settings to activate Windows.' The Windows taskbar is visible at the bottom of the screen, showing the time as 10:45 AM on 01/17/25.

Gambar 2: Form Sistem Informasi Prediksi Penderita Diabetes

Gambar 2 diatas merupakan halaman form entri untuk aplikasi prediksi penyakit diabetes. Variabel yang di entri datanya yaitu pregnancies (kehamilan), glucose(kadar gula darah dalam tubuh),blood pressure (tekanan darah diastolik), skin thickness (ketebalan kulit bagian trisep), insulin (tingkat dalam satuan mu U/ml), BMI (indeks massa tubuh), age (umur), diabetes pedigree function (indikator riwayat diabetes dalam keluarga). Variabel diatas perlu diisi nilainya sesuai dengan hasil pengukuran yang telah dilakukan oleh laboratorium. Pasien diprediksi menderita penyakit diabetes atau tidak berdasarkan inputan variabel diatas dimana penentuan prediksi pasien tersebut menderita penyakit diabetes atau tidak menggunakan model hasil klasifikasi menggunakan algoritma random forest.

3.2. Pembahasan

Pada preprocessing data, 2 tahapan yang dikerjakan yaitu cleaning data dan transform data. Pada tahapan cleaning data, data akan dicek apakah ada data yang hilang (missing data). jika ada data yang hilang atau kurang lengkap nantinya dapat digunakan pendekatan teknik imputansi. Teknik imputansi yang digunakan yaitu dengan menggunakan nilai rata-rata/ mean. Dalam penelitian ini penggantian nilai yang kosong dengan nilai rata-rata/ mean. Pada penelitian ini juga dilakukan normalisasi data untuk memastikan bahwa variabel dengan skala yang berbeda dapat dibandingkan secara adil. Seperti penelitian yang dilakukan oleh Noval Firmansyah dkk[15]. Untuk tahapan transform data, dimana proses transformasi data dalam hal ini melakukan perubahan tipe data ke dalam tipe data numerik tidak dilakukan, karena datasetnya telah berisi format data numerik berupa tipe data float untuk masing-masing kolom dalam dataset database pima indian datasetnya.

Tahapan pembagian data, dataset yang telah diproses dan dibersihkan sebelumnya akan dibagi menjadi dua bagian pada tahap pembagian data. Data pelatihan (training) akan digunakan untuk menentukan nilai kekuatan model, dan data tes (testing) akan digunakan untuk mengevaluasi hasil dari nilai kekuatan model[13]. Salah satu faktor yang memengaruhi seberapa baik model klasifikasi berfungsi pada algoritma pembelajaran mesin adalah model splitting data yang digunakan untuk mempartisi dataset[14]. Pada tahap ini dilakukan pembagian data dengan komposisi 80 data training dan 20 % data testing. Hasil dari pembagian data ini menghasilkan data training sebesar 615 dan data testing sebesar 153 dari total data keseluruhan yang berjumlah 768. Untuk menghindari overfitting dan supaya memberikan estimasi performa model yang lebih stabil, maka dilakukan teknik k-fold validation.

Pemodelan Klasifikasi penderita diabetes menggunakan algoritma random forest dimana Metode ini digunakan untuk membangun pohon keputusan yang terdiri dari root node, internal node, dan leaf node dengan menggunakan data dan atribut secara acak sesuai dengan persyaratan. Simpul anak dibangun secara acak untuk setiap node, meningkatkan akurasi hasil. Pembentukan algoritma ini ditentukan oleh nilai vektor acak yang independen dan seragam di setiap pohon. Metode klasifikasi adalah yang paling akurat saat membuat prediksi. Ini dapat menemukan perbandingan antar hutan acak seperti properti metode ansambel dan dapat menangani banyak variabel input tanpa overfitting. Metode Random Forest merupakan salah satu metode algoritma klasifikasi yang dibuat untuk memprediksi dari sekumpulan prediksi pohon keputusan dan juga termasuk dari metode pembelajaran terarah. Metode ini dipilih karena atas dasar dari kelebihanannya dalam akurasi klasifikasi yang sangat bagus dengan error yang lebih rendah[9]. Hasil dari klasifikasi pada penelitian ini menghasilkan model klasifikasi pasien penderita diabetes dengan tingkat akurasi sebesar 76%, dengan nilai mean squares error sebesar 0,24. Melakukan evaluasi model untuk mengetahui kemampuan model yang dibuat untuk membuat klasifikasi yang akurat. Dalam penelitian ini, lima komponen terdiri dari laporan metrix evaluasi: akurasi, ketepatan, recall, skor F1, dukungan, dan konsistensi matriks. Melakukan evaluasi

model untuk mengetahui kemampuan model yang dibuat untuk membuat klasifikasi yang akurat. Dengan memperhatikan nilai dari tabel confusion metrik.

4. Kesimpulan

Penelitian ini berhasil membuat model prediksi untuk penderita penyakit diabetes dengan menginputkan variabel yang menyebabkan penyakit diabetes. Tingkat akurasi yang dihasilkan oleh model ini yaitu sebesar 76% dengan nilai mean square error sebesar 0,24. Dimana model ini menghasilkan nilai yang cukup baik dalam memprediksi pasien masuk dalam kategori diabetes atau tidak. Hasil evaluasi kinerja model yang dihasilkan menghasilkan nilai akurasi sebesar 76%, Presisi sebesar 69%, Recall sebesar 59% dan F1-Score sebesar 63%. Dimana hasil dari interpretasi tabel confusion matrix didapatkan bahwasannya model memprediksi positif dengan benar sebanyak 158. Model memprediksi diabetes, padahal sebenarnya tidak diabetes sebanyak 72 dan model memprediksi tidak diabetes, padahal sebenarnya diabetes sebanyak 110 serta model memprediksi negatif dengan benar sebanyak 428

Untuk penelitian selanjutnya bisa menggunakan algoritma SMOTE (Synthetic Minority Oversampling Technique) yang merupakan algoritma pembangkitan sampel baru yang berasal dari kelas minoritas untuk membuat proporsi data menjadi lebih seimbang. Sehingga tingkat akurasi prediksi yang dihasilnya meningkat dari tingkat akurasi sebelumnya.

Pustaka

- [1] S. Jagadeesan, A. Kapoor, and S. Ghosh, "Diabetes Prediction Using Machine Learning," *AIP Conf. Proc.*, vol. 3075, no. 1, pp. 140–146, 2024, doi: 10.1063/5.0217181.
- [2] D. Mahatma, Gangga. Widiastuti, Wisda. Nova, Riki. Darmayanti, Anita. Abdullah, "Prevalensi Diabetes Mellitus Tipe 2 Pada Anak Dan Remaja (Studi Literatur)," *Nusant. Hasana J.*, vol. 4 no 6, 2024.
- [3] Y. A. Jatmiko, S. Padmadisastra, and A. Chadidjah, "Analisis Perbandingan Kinerja Cart Konvensional, Bagging Dan Random Forest Pada Klasifikasi Objek: Hasil Dari Dua Simulasi," *Media Stat.*, vol. 12, no. 1, p. 1, 2019, doi: 10.14710/medstat.12.1.1-12.
- [4] G. Sanhaji, A. Febrianti, and H. Hidayat, "Aplikasi DIATECT Untuk Prediksi Penyakit Diabetes Menggunakan SVM Berbasis Web," *J. Tekno Kompak*, vol. 18, no. 1, p. 150, 2024, doi: 10.33365/jtk.v18i1.3643.
- [5] L. Putri, Nurliana Rizka. Goeirmento, "Aplikasi Sistem Pakar Untuk Diagnosa Penyakit Diabetes Melitus Dengan Algoritma Certainty Factor Berbasis Web," *Apl. dan Inov. Ipteks*, 2020.
- [6] A. Az, R. Septian, M. A. Saktiawan, and R. A. Saputra, "PREDIKSI PENYAKIT HIPERTENSI MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA REGRESI LOGISTIK," vol. 8, no. 6, pp. 12062–12068, 2024.
- [7] N. H. Adiningrum, Nur Tri Ramadhanti . Harani, "Analisis Perbandingan Ensemble Machine Learning Dengan Teknik Smote Untuk Prediksi Diabetes," *J. Elektro Inform. Swadharma*, vol. 05, 2025.
- [8] Suharni, "Eksplorasi Metode Pengolahan Big Data Untuk Pemodelan Prediktif Dalam Bidang Kesehatan," *J. Rev. Pendidik. dan Pengajaran*, vol. 7 Nomer 1, pp. 1353–1360, 2024.

- [9] F. S. Wijaya, M. F. Arotsid, and R. Adriano, "Literatur Review: Penerapan Algoritma Random Forest untuk Klasifikasi Penyakit Diabetes," vol. 2, no. 3, pp. 474–481, 2024.
- [10] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.
- [11] S. Devella, Y. Yohannes, and F. N. Rahmawati, "Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 310–320, 2020, doi: 10.35957/jatisi.v7i2.289.
- [12] Suci Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 121–127, 2022, doi: 10.35580/variasiunm31.
- [13] A. Setiawan, Z. H. Nst, Z. Khairi, and L. Efrizoni, "KLASIFIKASI TINGKAT RISIKO DIABETES MENGGUNAKAN ALGORITMA," vol. 7, no. 2, pp. 263–271, 2024.
- [14] N. Nurdiana and A. Algifari, "Studi Komparasi Algoritma Id3 Dan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *INFOTECH J.*, vol. 6, no. 2, pp. 18–23, 2020, [Online]. Available: <https://ejournal.unma.ac.id/index.php/infotech/article/view/816>.
- [15] N. Firmansah and I. Uce, "Prediction of Credit Eligibility Using the Random Forest Method [Prediksi Kelayakan Pemberian Kredit Menggunakan Metode Random Forest]," *Univ. Muhammadiyah Sidoarjo*, pp. 1–10, 2023.