Teknika 17 (1): 163 - 172

JURNAL TEKNIKA
ISSN: 0854-3143
e-ISSN: 2622-3481
Journal homepage: http://jurnal.polsri.ac.id/index.php/teknika
Journal Email: teknika@polsri.ac.id

Prediksi Penyakit Diabetes Menggunakan Algoritma Random Forest

Sriyanto*[1] , Agiska Ria Supriyatna[2]

[*1]Magister Teknik Informatika; Institut Informatika dan Bisnis Darmajaya; Jl. Zainal Abidin Pagar Alam No.93, Kec. Labuhan Ratu, Kota Bandar Lampung, Lampung 35141 Telepon: (0721) 787214
[2] Teknologi Rekayasa Internet; Politeknik Negeri Lampung; Jl. Sukarno Hatta No. 10 Rajabasa Raya Kec. Rajabasa, Kota Bandar Lampung, Lampung 35142 Telepon: (0721) 703995
*Email Penulis Korespondensi: sriyanto@darmajaya.ac.id

Abstrak

Diabetes adalah penyakit kronis degeneratif yang disebabkan oleh produksi insulin yang tidak mencukupi di dalam pankreas. Banyak faktor yang diduga menjadi penyebab diabetes. diantaranya adalah: keturunan, kadar gula darah yang tinggi, usia, hipertensi, sariawan terus menerus, gatal-gatal, penglihatan yang buram, berat badan berlebihan, penurunan pendengaran, kesemutan, dan faktor lainnya. Sangat penting untuk mengetahui faktor utama yang menjadi pemicu penyakit diabetes. Diperlukan diagnosa dokter melalui pemeriksaan darah untuk memastikan apakah seseorang mengidap diabetes atau tidak. Selain itu, melalui penerapan ilmu pengetahuan data mining dapat dikembangkan sebuah model untuk memprediksi penyakit diabetes. Model prediksi dapat digunakan sebagai alat bantu bagi tenaga medis dan masyarakat awam untuk memperkirakan apakah seseorang mengidap diabetes atau tidak. Penelitian ini bertujuan mengembangkan sebuah model prediksi penyakit diabetes dengan menerapkan algoritma random forest. Hasil penelitian bermanfaat untuk membantu dokter dan tenaga kesehatan serta masyarkat umum untuk mendeteksi penyakit diabetes sejak dini. Tahapan penelitian adalah mengumpulkan dataset diabetes, implementasi algoritma, dan evaluasi kinerja algoritma random forest. Hasil penelitian menunjukkan bahwa algoritma random forest dapat melakukan prediksi penyakit diabetes dengan kinerja yang baik. Nilai-nilai evaluasi kinerja algoritma random forest untuk prediksi penyakit diabetes adalah: akurasi sebesar 99.3 %, recall sebesar 99.5%, presisi sebesar 99.1%, F1-score sebesar 99.%, dan AUC sebesar 100%.

Kata kunci*—Akurasi, Data Mining, Diabetes, Penyakit, Random Forest*

Abstract

Diabetes is a degenerative chronic disease caused by insufficient insulin production in the pancreas. Many factors are thought to be the cause of diabetes. including: heredity, high blood sugar levels, age, hypertension, continuous canker sores, itching, blurry vision, excessive body weight, decreased swelling, tingling, and others. It is very important to know the main factors that trigger diabetes. A doctor's diagnosis is needed through a blood test to determine

whether a person has diabetes or not. In addition, through the application of data mining science a model can be developed to predict diabetes. The prediction model can be used as a tool for medical personnel and the general public to predict whether a person has diabetes or not. This study aims to develop a diabetes prediction model by applying the random forest algorithm. The research results are useful to help doctors and health workers as well as the general public to detect diabetes early. The stages of the research were collecting the diabetes dataset, implementing the algorithm, and evaluating the performance of the random forest algorithm. The results showed that the random forest algorithm can predict diabetes with good performance. The performance evaluation values of the random forest algorithm for predicting diabetes are: accuracy of 99.3%, recall of 99.5%, precision of 99.1%, F1-score of 99.%, and AUC of 100%.

Keywords—*Accuracy, Data Mining, Diabetes, Disease, Random Forest*

1. PENDAHULUAN

Diabetes menurut *World Health Organization* (WHO), adalah penyakit degeneratif kronis yang disebabkan oleh produksi insulin yang tidak mencukupi di pankreas. Penyakit ini disebabkan gangguan metabolisme glukosa akibat kekurangan insulin baik secara absolut maupun relatif [1]. Pada penderita diabetes, pankreas tidak dapat memproduksi insulin sesuai dengan kebutuhan tubuh. Sedangkan tanpa insulin, sel-sel tubuh tidak dapat menyerap dan mengubah glukosa menjadi energi. Penyakit diabetes memiliki 2 tipe, yaitu: tipe 1 dan tipe 2. Namun secara umum, penderita diabetes tipe 1 dan tipe 2 akan mengalami beberapa gejala, seperti sering merasa haus, frekuensi buang air kecil yang meningkat, rasa lelah, gangguan penglihatan, keputihan, dan luka atau infeksi yang lama sembuh [1],[2].

WHO menyatakan bahwa pada 2020 terdapat 422 juta orang di dunia mengidap diabetes, mayoritas tinggal di negara berpenghasilan rendah dan menengah. Pengidap diabetes pada 2015 sejumlah 415 juta jiwa, dan diperkirakan meningkat menjadi 642 juta jiwa pada 2040. Hal ini berarti 1 dari 11 orang menderita diabetes di 2015 dan 1 dari 10 orang dewasa menderita diabetes di 2040. Setiap negara mengalami peningkatan jumlah pasien diabetes, pada umumnya usia orang yang mengalami diabetes berada diantara 40-59 tahun [1], [2], [3].

Dari 2010 sampai 2030, kerugian dari *gross domestic product* (GDP) di seluruh dunia karena diabetes sekitar 1,7 triliun dolar. Pada 2013, salah satu beban pengeluaran kesehatan terbesar di dunia adalah diabetes yaitu sekitar 612 miliar dolar, diperkirakan sekitar 11% dari total pembelanjaan untuk kesehatan dunia [1]. Diabetes merupakan penyebab utama untuk kebutaan, serangan jantung, stroke, gagal ginjal, dan amputasi kaki [3]. Pada tahun 2012, diabetes merupakan penyebab kematian ke delapan pada kedua jenis kelamin dan penyebab kematian kelima pada perempuan [1],[3]. Pada tahun 2012, sekitar 1 juta orang dewasa di wilayah regional Asia Tenggara meninggal karena diabetes, maupun kematian karena komplikasi dan konsekuensi dari diabetes, seperti gagal ginjal, penyakit jantung, dan pembuluh darah [3]. Sebanyak 1.6 juta kematian setiap tahun karena diabetes [1]. Dari data tersebut tidak salah jika disebut diabetes adalah salah satu penyebab utama kematian di dunia.

Gejala penyakit diabetes sangat bervariasi pada setiap pasien, sehingga sulit dikenali. Menurut [3] 1 diantara 2 orang penyandang diabetes masih belum terdiagnosis dan belum menyadari bahwa dirinya diabetes, karena gejalanya mirip dengan kondisi sakit biasa, shingga banyak orang yang tidak menyadari bahwa mereka mengidap penyakit diabetes dan bahkan sudah mengarah pada komplikasi. Diperlukan diagnosa dokter melalui pemeriksaan darah untuk memastikan apakah seseorang mengidap diabetes atau tidak. Upaya lain yang dapat dilakukan untuk membantu mengatasi penyakit diabetes adalah mengembangkan sebuah model melalui penerapan teknologi komputer yang mampu melakukan prediksi penyakit diabetes, sebagai upaya mendeteksi penyakit diabetes sejak dini. Dilihat dari angka kematian yang tinggi yang diakibatkan oleh diabetes, prediksi begitu penting dilakukan untuk menekan angka

kematian. Model prediksi dapat digunakan sebagai alat bantu bagi tenaga medis dan masayarakat awam untuk memperkirakan apakah seseorang mengidap diabetes atau tidak.

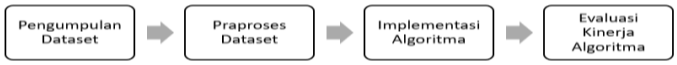
Beberapa model prediksi penyakit telah berhasil dikembangkan dengan menerapkan algoritma data mining, diantaranya klasifikasi penyakit stroke dengan algoritma *decision tree* C4.5 [4] dan klasifikasi malaria menggunakan algoritma *naïve bayes* [5]. Berbagai algoritma telah diimplementasikan untuk prediksi atau klasifikasi penyakit diabetes, diantaranya: algoritma C4.5 [6],[7],[8], *random forest* [9], *decision tree* [10], SVM [11], [12], *ensemble adaboost* dan *bagging* [13], naïve bayes [14], adaboost [15].

Tujuan dari penelitian ini adalah mengimplementasikan model prediksi penyakit diabetes menggunakan algoritma *random forest*. Algoritma *random forest* memiliki keungggulan karena model yang dikembangkan lebih konsisten [16]. Hal ini terjadi karena pada proses pembangunan pohon keputusan dilakukan tanpa melakukan pemangkasan pohon [17] dan secara independen dengan menggunakan data acak, sehingga mengurangi *overfitting* [18],[19]. Karena hal tersebut *random forest* memiliki akurasi yang tinggi [20]. Penelitian terdahulu telah membuktikan keunggulan dari algoritma *random forest*, diantaranya: penelitian [9] memiliki tingkat akurasi 97.88% pada klasifikasi penyakit diabetes dan pada klasifikasi penderita Covid 19 sebesar 96.6% [21].

Penelitian ini bermanfaat memberikan pengetahuan tentang penerapan algoritma *random forest* dalam mengembangkan sebuah model prediksi penyakit diabetes, sehingga dapat digunakan untuk referensi ilmiah dalam penelitian penerapan data mining selanjutnya. Hasil penelitian dapat digunakan untuk membantu tenaga medis dalam melakukan identifikasi penyakit diabetes pada pasien.

2. METODE PENELITIAN

Pada Gambar 1 disajikan alur penelitian, dimulai dengan mengumpulkan dataset, implementasi algoritma, dan evaluasi kinerja algoritma.



Gambar 1 Tahapan Penelitian

2.1 Pengumpulan Dataset

Dataset yang digunakan berasal dari https://www.kaggle.com yang bersifat *open source*. Dataset berjumlah 520 *record* dan memiliki 17 attribut, yaitu: Age, Gender, Polyuria, Polydipsia, Sudden Weight Loss, Weakness, Polyphagia, Genital Thrush, Visual Blurring, Itching, Irritability, Delayed Healing, Partial Paresis, Muscle Stiffness, Alopecia, Obesity, Class.

2.2 Praproses Dataset

Kegiatan analisis data menggunakan software *orange* versi 3.32.0. Praproses data dilakukan untuk mengolah dataset menjadi bentuk data yang dimengerti oleh *tools orange*.

2. 3 Implementasi Algoritma

Penelitian ini menggunakan algoritma *random forest* yang diperkenalkan oleh Leo Breiman pada 2001, seorang profesor bidang statistik dan ahli dalam bidang *machine learning* yang berasal dari University of California. Konsep *random forest* adalah menggabungkan banyak pohon keputusan. Algoritma ini menggabungkan hasil dari beberapa pohon keputusan yang dibangun secara acak, sehingga model yang dihasilkan lebih akurat [19].

Cara kerja *random forest* adalah membangun beberapa pohon keputusan secara paralel dan menentukan prediksi berdasarkan suara terbanyak. Tahapan pembuatan model prediksi menggunakan algoritma *random forest* dengan *n* observasi dan *p* peubah penjelas adalah sebagai berikut [19]:

1. Melakukan proses *bootstrap* dengan cara memilih sampel secara acak berukuran *n* pada data training.
2. Mulai membangun pohon keputusan tunggal dengan memakai data *training* yang dihasilkan dari proses *bootstrap*. Kemudian memilih peubah penjelas secara acak dengan $m < p$. Selanjutnya dari *m* peubah penjelas dipilih peubah penjelas terbaik sebagai pemisah dan dilanjutkan dengan pemisahan menjadi dua simpul baru.
3. Proses ini terus berlanjut sampai ukuran minimum dari pengamatan pada simpul tercapai. Nilai m yang direkomendasikan yaitu $\sqrt{p}$.
4. Mengulangi tahapan sebanyak k kali untuk mendapatkan k pohon keputusan. Setiap pohon keputusan menghasilkan satu suara dan kelas ditentukan oleh suara terbanyak dari k buah suara.

2. 4 Evaluasi Kinerja Algoritma

Evaluasi berfungsi untuk mengetahui kinerja algoritma yang digunakan, misalnya untuk mengetahui besarnya tingkat akurasi yang dihasilkan. Evaluasi dan validasi dilakukan secara mendalam dengan tujuan agar hasil prediksi sesuai dengan sasaran yang ingin dicapai. Evaluasi dan validasi menggunakan matrik konfusi. Matrik konfusi adalah tabel yang digunakan untuk menggambarkan kinerja model klasifikasi atau prediksi pada suatu set data *testing* yang nilai-nilai sebenarnya sudah diketahui. Matrik konfusi merupakan tabulasi silang antara data kelas positif dan kelas negatif yang masuk dalam kelas prediksi dan kelas aktual. Matrik konfusi terdiri dari *True positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN) [22]. Tabel matrik konfusi disajikan pada Tabel 1.

Tabel 1 Matrik Konfusi

Label Aktual	Label Prediksi	Label Prediksi
	1	2
1	TP	FN
2	FP	TN

Pada penelitian ini, kelas positif adalah kelas mengidap diabetes dan kelas negatif adalah tidak mengidap diabetes. Berdasarkan matrik konfusi dilakukan perhitungan nilai akurasi, *recall*, presisi, F1-score, dan *Area Under Curve* (AUC) [22].

i. Akurasi
Akurasi adalah nilai akhir yang dihasilkan oleh sebuah model prediksi, yang merepresentasikan jumlah dataset yang benar dikenali dari keseluruhan data. Nilai akurasi dapat dihitung dengan cara membagi total dataset benar dikenali dengan total dataset dan data uji. Berikut merupakan rumus untuk menghitung nilai akurasi:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

ii. *Recall*
Menggambarkan jumlah data kelas positif yang diprediksi ke kelas positif, semakin baik kinerja model klasifikasi maka nilai *recall* mendekati 1.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

iii. Presisi
Menggambarkan jumlah data kelas positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif, semakin baik kinerja model klasifikasi maka nilai presisi mendekati 1.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

iv. F1-score

Perbandingan rata-rata dari nilai presisi dan nilai *recall*, rentang nilai F1 adalah antara 0 hingga 1, semakin baik kinerja model klasifikasi maka nilai F1 mendekati 1.

v. Nilai *Area Under Curve* (AUC) merupakan daerah di bawah kurva *Receiver Operating Characteristic* (ROC). Nilai AUC memiliki rentang antara 0,5 sampai dengan 1. Interpretasi nilai AUC dapat diklasifikasikan menjadi lima bagian yang berbeda, yaitu: 0,5 – 0,6 (akurasi salah), 0,6 – 0,7 (tingkat akurasi lemah), 0,7 – 0,8 (tingkat akurasi sedang), 0,8 – 0,9 (tingkat akurasi tinggi), dan 0,9 – 1 (tingkat akurasi sangat tinggi).

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan penelitian yang dilakukan adalah sebagai berikut.

3.1 Pengumpulan Dataset

Data yang digunakan bersumber dari https://www.kaggle.com. Pada Gambar 2 disajikan tangkapan layar dataset diabetes yang digunakan, dari baris 1 hingga 27.

	Age	Gender	Polyuria	Polydipsia	sudden weight loss	weakness	Polyphagia	Genital thrush	visual blurring	Itching	Irritability	delayed healing	partial paresis	muscle stiffness	Alopecia	Obesity	class
1	40	Male	No	Yes	No	Yes	No	No	No	Yes	No	Yes	No	Yes	Yes	Yes	Positive
2	58	Male	No	No	No	Yes	No	No	Yes	No	No	No	Yes	No	Yes	No	Positive
3	41	Male	Yes	No	No	Yes	Yes	No	No	Yes	No	Yes	No	Yes	Yes	No	Positive
4	45	Male	No	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No	No	No	Positive
5	60	Male	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Positive
6	55	Male	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Positive
7	57	Male	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	Yes	No	No	No	Positive
8	66	Male	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	No	No	Positive
9	67	Male	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Positive
10	70	Male	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	No	Positive
11	44	Male	Yes	Yes	No	Yes	No	Yes	No	No	Yes	Yes	No	Yes	Yes	No	Positive
12	38	Male	Yes	Yes	No	No	Yes	Yes	No	Yes	No	Yes	No	Yes	No	No	Positive
13	35	Male	Yes	No	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	No	Positive
14	61	Male	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Positive
15	60	Male	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	No	No	Positive
16	58	Male	Yes	Yes	No	Yes	Yes	No	No	No	No	Yes	Yes	Yes	No	No	Positive
17	54	Male	Yes	Yes	Yes	Yes	No	Yes	No	No	No	Yes	No	Yes	No	No	Positive
18	67	Male	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Positive
19	66	Male	Yes	Yes	No	Yes	Yes	No	Yes	No	No	No	Yes	Yes	No	No	Positive
20	43	Male	Yes	Yes	Yes	Yes	No	Yes	No	No	No	No	No	No	No	No	Positive
21	62	Male	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes	No	No	Positive
22	54	Male	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	Positive
23	39	Male	Yes	No	Yes	No	No	Yes	No	Yes	Yes	No	No	No	Yes	No	Positive
24	48	Male	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Positive
25	58	Male	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	No	Yes	Positive
26	32	Male	No	No	No	No	No	Yes	No	No	Yes	Yes	No	No	No	Yes	Positive
27	42	Male	No	No	No	Yes	Yes	No	No	No	Yes	No	No	Yes	No	No	Positive

Gambar 2 Tangkapan Layar Dataset Diabetes

3.1 Praproses Dataset

Pada penelitian ini dataset sudah baik dan tidak memiliki *missing value*. Kegiatan praproses data yang dilakukan adalah menentukan attribut yang akan dijadikan sebagai attribut fitur atau attribut label. Dalam hal ini atribut yang dijadikan label adalah attribut *Class*, selain itu dijadikan attribut fitur. Pada Tabel 1 disajikan rincian dataset yang digunakan.

Tabel 1 Attribut, Tipe Data, dan Nilai Kategori Dataset Diabetes

Attribut Label		
Attribut	**Tipe Data**	**Nilai**
Class	Binomial	Yes
		No
Attribut Fitur		
Attribut	**Tipe Data**	**Nilai**
Age	Numerik	16 - 90
Gender	Binomial	Yes
		No
Polyuria	Binomial	Yes
		No
Polydipsia	Binomial	Yes
		No
Sudden Weight Loss	Binomial	Yes
		No
Weakness	Binomial	Yes
		No
Polyphagia	Binomial	Yes
		No
Genital Thrush	Binomial	Yes
		No
Visual Blurring	Binomial	Yes
		No
Itching	Binomial	Yes
		No
Irritability	Binomial	Yes
		No
Delayed Healing	Binomial	Yes
		No
Partial Paresis	Binomial	Yes
		No
Muscle Stiffness	Binomial	Yes
		No
Alopecia	Binomial	Yes
		No
Obesity	Binomial	Yes
		No

3. 3 Implementasi Algoritma Random Forest

Nilai *skoring Info Gain*, *Gain Ratio*, dan *Gini Indeks* dari implementasi algoritma *random forest* untuk prediksi penyakit diabetes, disajikan pada Gambar 3.

		#	Info. gain	Gain ratio	Gini
1	Polyuria	2	0.362	0.362	0.210
2	Polydipsia	2	0.359	0.362	0.199
3	Gender	2	0.163	0.172	0.096
4	sudden weight loss	2	0.149	0.152	0.090
5	partial paresis	2	0.145	0.147	0.088
6	Polyphagia	2	0.088	0.088	0.056
7	Irritability	2	0.073	0.091	0.042
8	Alopecia	2	0.051	0.055	0.034
9	visual blurring	2	0.047	0.047	0.030
10	weakness	2	0.043	0.044	0.028
11	Age		0.023	0.011	0.015
12	muscle stiffness	2	0.011	0.011	0.007
13	Genital thrush	2	0.009	0.012	0.006
14	Obesity	2	0.004	0.006	0.002
15	delayed healing	2	0.002	0.002	0.001
16	Itching	2	0.000	0.000	0.000

Gambar 3 Skoring *Info Gain, Gain Ratio,* dan *Gini Indeks* Algoritma *Random Forest*

Berdasarkan Gambar 3 terdapat lima (5) attribut yang memiliki *skoring Info Gain, Gain Ratio,* dan *Gini Indeks* tertinggi, yaitu attribut Polyuria, selanjutnya diikuti oleh attribut Polydipsia, Gender, Sudden Weight Loss, dan Partial Paresis.

3. 3 Pengujian

Matrik konfusi yang dihasilkan disajikan pada Gambar 4.

		Predicted		
		Negative	**Positive**	**∑**
Actual	**Negative**	99.5 %	0.5 %	**200**
	Positive	0.9 %	99.1 %	**320**
	∑	**202**	**318**	**520**

Gambar 4 Matrik Konfusi

Berdasarkan matrik konfusi dapat dihitung akurasi, *recall*, presisi, F-1 *score*, dan AUC.

1. Nilai akurasi dapat dihitung dengan menggunakan persamaan (1), yaitu:

$$\text{Akurasi} = \frac{99.5 + 99.1}{99.5+99.1+0.5+0.9} = \frac{198.6}{200} = 0.993 = 99.3\ \%$$

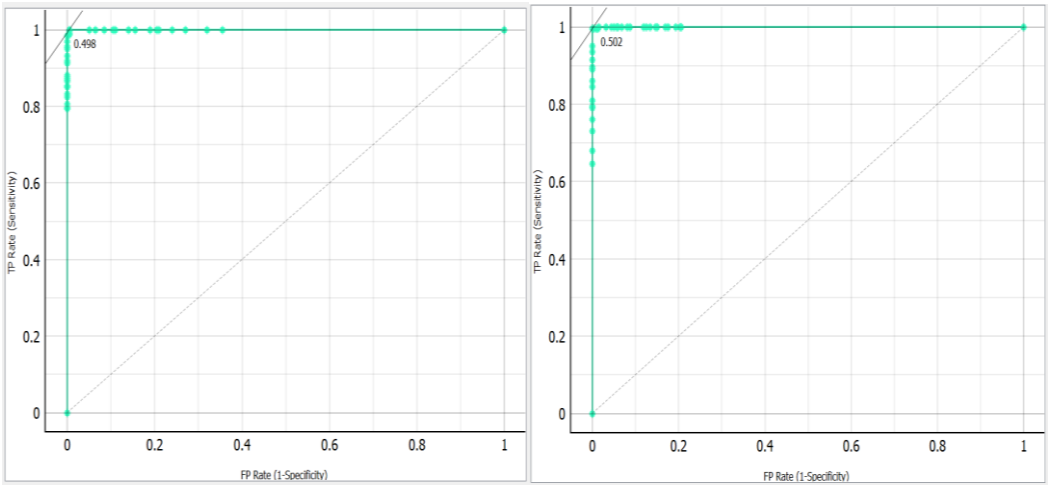
2. *Recall* $= \frac{99.5}{99.5+0.5} = \frac{99.5}{100} = 0.995 = 99.5\ \%$

3. Presisi $= \frac{99.5}{99.5+0.9} = \frac{99.5}{100.4} = 0.991 = 99.1\ \%$

4. F-1 *Score* $= \frac{99.1}{99.5} = 99.6\ \%$

5. Nilai AUC

   Hasil prediksi penyakit diabetes dapat dilihat dalam kurva ROC. Pada Gambar 5 disajikan kurva ROC Positif (a) pada titik 0.498 dan Negatif (b) pada titik 0.502.



(a) (b)

Gambar 5 Kurva ROC (a) Positif dan ROC (b) Negatif

Gambar 5.a memperlihatkan kurva ROC antara nilai *false positive rate* (sumbu x) sebesar 0,0 dengan *true positive rate* (sumbu y) sebesar 1.0 diperoleh titik potong optimal adalah 0,498 yang digunakan untuk menentukan kelas pasien, jika hasil prediksi diatas 0,498, maka kelas pasien adalah kelas diabetes selain itu kelas tidak diabetes. Gambar 5.b menunjukkan kurva ROC antara nilai *false positive rate* (sumbu x) sebesar 0,0 dengan *true positive rate* (sumbu y) sebesar 1.0 diperoleh titik potong optimal adalah 0,502 yang digunakan untuk menentukan kelas pasien, jika hasil prediksi diatas 0,502, maka kelas pasien adalah kelas diabetes selain itu kelas tidak diabetes. Berdasarkan kurva ROC diperoleh nilai AUC sebesar 1.000 yang didapatkan dari penjumlahan ROC positif (0.498) dan ROC Negatif (0.502), yang artinya hasil klasifikasi memiliki tingkat akurasi tinggi.

4. KESIMPULAN

Dari hasil penelitian dapat ditarik kesimpulan bahwa algoritma *random forest* mampu memprediksi penyakit diabetes dengan kinerja tinggi. Algoritma *Random forest* terbukti sangat handal untuk dijadikan rujukan pada pengembangan model prediksi pada kasus yang serupa. Hal ini dapat dilihat dari nilai AUC 100%. Model prediksi penyakit diabetes yang dikembangkan menggunakan algoritma *random forest* memiliki kinerja dengan nilai sebagai berikut: akurasi sebesar 99.3 %, *recall* sebesar 99.5%, presisi sebesar 99.1%, F1-*score* sebesar 99.%.

5. SARAN

Merujuk hasil algoritma *random forest* yang dihasilkan, maka disarankan untuk mengimplementasikan dalam bentuk perangkat lunak berbasis android untuk prediksi penyakit diabetes. Untuk penelitian selanjutnya perlu ditambahkan jumlah dataset agar model prediksi yang dikembangkan semakin handal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tim Redaksi Jurnal Teknika Politeknik Negeri Sriwijaya yang telah memberi memberi kesempatan, sehingga artikel ilmiah ini dapat diterbitkan.

DAFTAR PUSTAKA

[1] WHO, "Global Report on Adult Learning Executive Summary," *World Organ. Heal.*, hal. 3, 2016.

[2] IDF Diabetes Atlas Group, *IDF Diabetes Atlas Fourth Edition*. 2009.

[3] Kemenkes, "Diabetes," 2012.

[4] R. E. Pambudi, Sriyanto, dan Firmansyah, "Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C4.5," *J. Tek.*, vol. 16, no. 2, hal. 221–226, 2022.

[5] H. Halimah, D. Linda, dan F. Klaralia, "Penerapan Algoritma Naïve Bayes Untuk Memprediksi Penyakit Malaria Pada Puskesmas Hanura," *Teknika*, vol. 14, no. x, hal. 57–63, 2020.

[6] R. P. Fadhillah, R. Rahma, A. Sepharni, R. Mufidah, B. N. Sari, dan A. Pangestu, "Klasifikasi Penyakit Diabetes Mellitus Berdasarkan Faktor-Faktor Penyebab Diabetes menggunakan Algoritma C4.5," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 4, hal. 1265–1270, 2022.

[7] J. Sistem Komputer dan Sistem Informasi, P. Studi Teknologi Komputasi dan Informatika Stmik Bina Bangsa Kendari, F. Aris, D. Program Studi Sistem Komputer, P. Studi Sistem Komputer, dan S. Bina Bangsa Kendari, "Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi," *Router Res.*, vol. 1, no. 1, hal. 1–6, 2019.

[8] S. Putri, E. Irawan, dan F. Rizky, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5," *Januari*, vol. 2, no. 1, hal. 39–46, 2021.

[9] W. Apriliah, I. Kurniawan, M. Baydhowi, dan T. Haryati, "SISTEMASI: Jurnal Sistem Informasi Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *J. Sist. Inf.*, vol. 10, no. 1, hal. 163–171, 2021.

[10] A. Andriani, "Sistem Prediksi Penyakit Diabetes Berbasis Decision Tree," *J. Bianglala Inform.*, vol. I, no. 1, hal. 1–10, 2013.

[11] N. Maulidah, R. Supriyadi, D. Y. Utami, F. N. Hasan, A. Fauzi, dan A. Christian, "Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes," *Indones. J. Softw. Eng.*, vol. 7, no. 1, hal. 63–68, 2021.

[12] R. Pahlevi, K. Q. Fredlina, dan N. W. Utami, "Penerapan Algoritma ID3 Dan SVM Pada Klasifikasi Penyakit Diabetes Melitus Tipe 2," *Pros. Semin. Nas. Apl. Sains Teknol. 2021*, vol. 2, hal. 64–75, 2021.

[13] R. Rousyati, A. N. Rais, E. Rahmawati, dan R. F. Amir, "Prediksi Pima Indians Diabetes Database Dengan Ensemble Adaboost Dan Bagging," *EVOLUSI J. Sains dan Manaj.*, vol. 9, no. 2, hal. 36–42, 2021.

[14] H. Apriyani dan K. Kurniati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," *J. Inf. Technol. Ampera*, vol. 1, no. 3, hal. 133–143, 2020.

[15] G. Abdurrahman, "Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier," *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.*, vol. 7, no. 1, hal. 59–66, 2022.

[16] J. M. Klusowski, "Complete Analysis of a Random Forest Model," *arXiv*, vol. 13, hal. 1063–1095, 2018.

[17] A. Cutler dan D. R. Cutler, "Ensemble Machine Learning," *Ensemble Mach. Learn.*, no. February 2014, 2012.

[18] I. A. Rahmi, F. M. Afendi, dan A. Kurnia, "Metode AdaBoost dan Random Forest untuk Prediksi Peserta JKN-KIS yang Menunggak," *Jambura J. Math.*, vol. 5, no. 1, hal. 83–94, 2023.

[19] L. Breiman, "Random Forests - Random Features, Technical Report 567, Statistic Department, University of California, Berkeley," hal. 1–29, 1999.

[20] Y. A. Jatmiko, S. Padmadisastra, dan A. Chadidjah, "Analisis Perbandingan Kinerja Cart

Konvensional, Bagging Dan Random Forest Pada Klasifikasi Objek: Hasil Dari Dua Simulasi," *Media Stat.*, vol. 12, no. 1, hal. 1, 2019.

[21] Normah, B. Rifai, S. Vambudi, dan R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, hal. 174–180, 2022.

[22] M. Sokolova dan G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, hal. 427–437, 2009.