

Klasifikasi Pasien Diabetes Melitus Tipe 1 dengan Metode *Rotation Forest*

Carolyn Sonia Septeria*, Lisnur Wachidah

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*carolinsonia93@gmail.com, wachidah.lisnur07@gmail.com

Abstract. Practitioners of data mining developers today, using the method of classifier ensembles between regression trees and classification trees in modeling a decision that generates conjectures (predictions) of existing data quickly. However, in certain data that has many variables will result in many decision trees, making visualizing the given results complicated to be synthesized. To solve the problem, Rotation forest is one of the methods of the ensembles classifier that uses the main component analysis or principal component analysis (PCA) to rotate the axis of the change that will be built by the decision tree so that it can model fewer decision trees with better accuracy, sensitivity and specificity than other class ensembles. In this study, rotation forest method was used to classify the symptoms that often appear in patients diagnosed with Diabetes Mellitus type 1 in Diabetes Mellitus (DM) patients at Sylhet Hospital, Bangladesh in 2020. The results of the application of rotation forest method in diabetes mellitus (DM) type 1 patient data at Sylhet Hospital, Bangladesh in 2020 are known that patients who have symptoms X2(Gender), X5(Sudden weight loss), X4(Polydipsia), X13(*Partial paresis*), X3(Polyuria), X7(Polyphagia), X6(Weakness), X9(Visual blurring), X14 (*Muscle stiffness*), X16(Obesity), X11(Irritability), X1(Age), X12(*Delayed healing*) and X10(*Itching*). can be classified as type 1 diabetes patients with a modeling accuracy rate of 92.95%.

Keywords: type 1 diabetes mellitus, classification, Principal Component Analysis, Rotation Forest.

Abstrak. Praktisi pengembang data mining saat ini, menggunakan metode pohon gabungan (classifier ensembles) antara pohon regresi (regression tree) dengan pohon klasifikasi (classification tree) dalam melakukan pemodelan terhadap suatu keputusan yang menghasilkan dugaan (prediksi) dari data yang ada secara cepat. Namun, pada data tertentu yang memiliki banyak variabel akan menghasilkan banyak pohon keputusan sehingga membuat visualisasi hasil yang diberikan menjadi rumit untuk diinterpretasikan. Untuk mengatasi permasalahan tersebut maka Rotation forest adalah salah satu metode pohon gabungan (classifier ensembles) yang menggunakan analisis komponen utama atau principal component analysis (PCA) untuk merotasi sumbu peubah yang akan dibangun pohon keputusannya sehingga dapat memodelkan pohon keputusan yang lebih sedikit dengan nilai akurasi, sensitivitas dan spesifitas yang lebih baik dibandingkan metode pohon gabungan (classifier ensembles) lainnya. Pada penelitian ini, metode *rotation forest* digunakan untuk mengklasifikasi gejala-gejala yang sering muncul pada pasien terdiagnosis Diabetes Melitus tipe 1 pada pasien Diabetes Melitus (DM) di Rumah sakit Sylhet, Bangladesh tahun 2020. Hasil penerapan metode *rotation forest* pada data pasien Diabetes Melitus (DM) tipe 1 di Rumah sakit Sylhet, Bangladesh tahun 2020 diketahui bahwa pasien yang memiliki gejala X2(Gender), X5(Penurunan berat badan mendadak), X4(Mudah haus), X13(Setengah lumpuh), X3(Buang air kecil berlebih), X7(Makan berlebih), X6(Mudah lelah), X9(Pandangan kabur), X14(ketidakseimbangan kerja otot), X16(Obesitas), X11(Iritabilitas), X1(Umur), X12(Penyembuhan luka berlangsung lama) dan

X10(Gatal-gatal). dapat diklasifikasikan sebagai pasien diabetes tipe 1 dengan tingkat akurasi pemodelan sebesar 92,95%.

Kata Kunci: Diabetes melitus tipe 1, klasifikasi, Principal Component Analysis, Rotation Forest.

1. Pendahuluan

Rotation forest adalah metode pohon gabungan (*classifier ensembles*) yang menggunakan analisis komponen utama atau *principal component analysis* (PCA) untuk merotasi sumbu peubah yang akan dibangun pohon keputusannya sehingga dapat mengatasi salah satu permasalahan yang menjadi kendala seperti diungkapkan oleh Raharjo (2016) yaitu interpretasi yang rumit akibat terlalu banyak pohon keputusan yang dibangun. Menurut Kuncheva and Rodriguez (2007), *Rotation forest* dapat mencapai tingkat akurasi yang lebih baik dengan jumlah pohon yang lebih sedikit. Metode ini lebih baik dibandingkan dengan metode pohon gabungan lainnya, seperti *bagging*, *adaboost*, dan *random forest* dan pada Penelitian Raharjo (2016) juga menunjukkan *rotation forest* menghasilkan prediksi klasifikasi yang lebih baik dan konsisten serta efisien dalam waktu komputasi dibandingkan metode klasifikasi lain.

Pemodelan metode *rotation forest* dapat diimplementasikan di dunia kesehatan. Salah satunya pada data penyakit yang menjadi sorotan dunia kesehatan saat ini yaitu penyakit diabetes. Diabetes melitus umumnya diklasifikasi menjadi dua tipe yaitu diabetes melitus tipe 1, yang disebabkan keturunan dan diabetes melitus tipe 2 disebabkan *lifestyle* atau gaya hidup. Faktor risiko diabetes tipe 1 masih diteliti karena sulit untuk mengkonfirmasi seseorang terdiagnosis diabetes tipe 1 oleh karena itu, kasus faktor risiko diabetes tipe 1 menjadi fokus penelitian penulis.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana penerapan metode *rotation forest* untuk mengklasifikasi gejala-gejala yang sering muncul pada pasien terdiagnosis diabetes melitus tipe 1.”. Selanjutnya, Tujuan dalam penelitian ini adalah untuk mengetahui penerapan metode *rotation forest* dalam mengklasifikasi gejala-gejala yang sering muncul pada pasien terdiagnosis diabetes melitus tipe 1.

2. Metodologi

Rotation forest merupakan metode pohon gabungan (*classifier ensembles*) dengan menggunakan analisis komponen utama atau *principal component analysis* (PCA) untuk merotasi sumbu peubah yang akan dibangun pohon keputusannya. Pohon keputusan digunakan sebagai dasar pengklasifikasian karena sifatnya yang sensitif terhadap rotasi sumbu peubah namun tetap akurat. Meskipun menggunakan analisis komponen utama, semua komponen utama tetap digunakan untuk membangun pohon keputusan agar menjaga kelengkapan informasi data (Rodriguez et al. 2006).

Misalkan $x = [x_1, \dots, x_n]^T$ merupakan vektor amatan dengan p peubah, X merupakan gugus data dari gabungan vektor x berukuran $n \times p$ dan F merupakan gugus p peubah. Sedangkan $y = [y_1, \dots, y_n]^T$ merupakan vektor kelas dari peubah respon berukuran $n \times 1$. Seperti metode klasifikasi lainnya, dalam *rotation forest* perlu ditentukan jumlah pohon yang akan dibangun yaitu sebanyak L lalu seluruh pengklasifikasi dapat dilatih secara bersama. Tahapan-tahapan yang dilakukan untuk membentuk pohon keputusan $D_i; i = 1, 2, \dots, L$: (Rodriguez et al. 2006).

1. Bagi F secara acak menjadi k gugus peubah yang saling bebas dengan banyaknya peubah (m) yang hampir sama. $F_{i,j}$ merupakan gugus peubah untuk membangun pohon D_i , dengan m_j peubah asal, untuk $j = 1, 2, \dots, k$. $X_{i,j}$ merupakan gugus data X dengan peubah $F_{i,j}$.
2. Pada gugus data $X_{i,j}$ lakukan proses *bootstrap* (membangkitkan data). Gugus data hasil

bootstrap dinyatakan dengan $X_{i,j}^*$.

- Lakukan analisis komponen utama pada $X_{i,j}^*$ lalu simpan koefisien komponen utama sebagai $a_{i,j}^{(1)}, a_{i,j}^{(2)}, \dots, a_{i,j}^{(m_j)}$. Analisis komponen utama (*Principal Component Analysis/PCA*) merupakan teknik statistik yang dapat digunakan untuk mengurangi jumlah variabel atribut yang awalnya sangat banyak menjadi lebih sedikit sehingga memudahkan analisis pada tahap selanjutnya namun tetap dengan hasil yang representatif.
- Susun vektor-vektor koefisien utama ke dalam sebuah matriks rotasi R_i seperti Persamaan 2.1 dibawah ini:

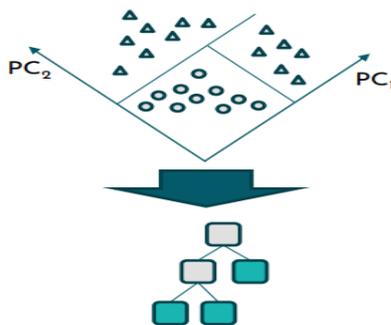
$$R_i = \begin{bmatrix} a_{i,1}^{(1)}, a_{i,1}^{(2)}, \dots, a_{i,1}^{(m_1)} & [0] & \dots & [0] \\ [0] & a_{i,1}^{(1)}, a_{i,1}^{(2)}, \dots, a_{i,2}^{(m_2)} & \dots & [0] \\ \dots & \vdots & \ddots & \vdots \\ [0] & [0] & \dots & a_{i,1}^{(1)}, a_{i,1}^{(2)}, \dots, a_{i,k}^{(m_k)} \end{bmatrix} \dots(2.1)$$

- Susun kembali kolom peubah R_i sehingga bersesuaian dengan susunan gugus peubah. Notasikan matriks rotasi yang telah tersusun kembali dengan simbol R_i^a , berukuran $p \times p$.
- Bangun pohon keputusan ke- i (D_i) dengan menggunakan data hasil perhitungan Persamaan 2.2 berikut:
 (XR_i^a, y)

\dots
 (2.2)

- Ulangi langkah 1 sampai 6 hingga diperoleh L pohon keputusan.

Dalam penelitian Rodriguez et al. 2006, *Rotation forest* akan optimal dengan membangun 10 pohon keputusan. Maka dalam penelitian ini akan dicobakan membangun 1 hingga 10 pohon keputusan agar mendapatkan pemodelan yang optimal. Ilustrasi mengenai rotation forest dapat dilihat pada Gambar 1.



Gambar 1. Ilustrasi Pemodelan Rotation Forest

Selanjutnya pemilihan pemilah terbaik dari masing-masing peubah bebas dilakukan dengan menggunakan nilai impuritas. Nilai impuritas mengukur keheterogenan suatu simpul. Pemilahan yang memaksimalkan nilai penurunan keheterogenan di dalam masing-masing simpul cabang relatif terhadap simpul utama merupakan pemilah terbaik. Indeks Gini merupakan istilah pendekatan yang dilakukan untuk mendefinisikan ukuran dari fungsi impuritas. Nilai indeks Gini pada simpul t , sebagai berikut:

$$i(t) = 1 - \sum_j^i p^2(j|t)$$

Dengan,

$$p(j|t) = \frac{N_j(t)}{N(t)}$$

Dimana, $i(t)$ adalah adalah nilai impuritas yang didapatkan dari indeks Gini, $p(j|t)$ adalah probabilitas pengamatan pada kategori ke- j dari simpul t , $N_j(t)$ adalah jumlah kategori j pada simpul t , $N(t)$ adalah jumlah amatan pada simpul t . Penurunan keheterogenan disebut juga dengan impuritas tereduksi. Peubah yang memiliki nilai impuritas tereduksi tertinggi akan menjadi simpul pemilah. Ukuran yang dapat digunakan pada pemodelan klasifikasi *rotation forest* yaitu melakukan perhitungan mean decrease gini (MDG) dikarenakan pemodelan klasifikasi *rotation forest* menghasilkan lebih dari satu pohon keputusan dalam pembentukan modelnya. Menghitung nilai MDG dapat menggunakan Persamaan 2.2 yaitu sebagai berikut:

$$MDG_s = \frac{1}{m} \sum_t [\Delta i(s, t) I(s, t)] \quad \dots (2.3)$$

Dimana,

m = banyaknya pohon yang terbentuk

$\Delta i(s, t)$ = nilai impuritas tereduksi untuk peubah penjelas X_s pada simpulan t

$I(s, t)$ = fungsi indicator yang bernilai 1 ketika X_s memilih simpul t dan 0 lainnya

Menurut Kuhn *et al* pada, Perhitungan Persamaan 2.1 dapat dipermudah dengan tersedia pada *packet Caret* di software R studio. Tingkatan urutan berdasarkan nilai peluang 0 hingga 1, peubah bebas yang dimaksud dalam pemodelan *rotation forest* merupakan komponen utama (PC) yang didapat dari Persamaan 2.2. Maka untuk melihat variabel yang membentuk model dilakukan analisis biplot.

Menurut (Jolliffe, 1986 & Rawlings 1988), analisis biplot adalah teknik statistika deskriptif yang dapat disajikan secara visual guna menyajikan secara simultan n obyek pengamatan dan p peubah dalam ruang bidang datar, sehingga ciri-ciri peubah dan obyek pengamatan serta posisi relatif antar obyek pengamatan dengan peubah dapat dianalisis. Sehingga hasil pemetaan menggunakan biplot dapat memberikan informasi mengenai hubungan antar variabel, kemiripan antar obyek pengamatan dan posisi relatif antar pengamatan dengan variabel.

Evaluasi Model

Evaluasi model merupakan tahapan penelitian yang bertujuan untuk memperoleh informasi yang terdapat pada hasil klasifikasi terhadap metode yang digunakan. Hasil klasifikasi yang diperoleh disertakan dengan alat ukur yang dalam penelitian ini berupa matriks konfusi dengan tujuan untuk mempermudah dalam menganalisis performa metode karena matriks konfusi memberikan informasi dalam bentuk angka sehingga dapat dihitung rasio keberhasilan klasifikasinya. Matriks konfusi merupakan tabel klasifikasi yang diperoleh dari jumlah ketepatan hasil prediksi dengan data aktual yang diambil dari salah satu titik π . Pada kasus dengan dua klasifikasi data keluaran, tiap kelas yang diprediksi memiliki empat kemungkinan keluaran yang berbeda, yaitu *true positif (TP)* pengklasikasian kategori 0 dengan benar saat prediksi sama dengan aktualnya, *true negative (TN)* pengklasikasian kategori 1 dengan benar saat prediksi sama dengan aktualnya, *false positif (FP)* terjadi pengklasikasian kategori 0 yang salah saat jumlah prediksi dengan aktual yang berbeda, dan *false negative (FN)* terjadi pengklasikasian kategori 1 yang salah saat jumlah prediksi dengan aktual yang berbeda. Dimana data prediksi diperoleh dari hasil model yang didapat dan data aktual merupakan data asli yang digunakan. Pengklasikasian kategori 0 atau dalam penelitian ini kategori pasien negative diabetes dengan benar disebut *true positive*, sedangkan jika pengklasifikasian salah disebut dengan *false negative*. Banyaknya pengklasifikasian kategori 1 atau dalam penelitian ini kategori pasien positif diabetes disebut *true negative*, sedangkan jika pengklasifikasian salah

disebut dengan *false positive*. Tabel matriks konfusi ditunjukkan pada Tabel 1.

Tabel 1. Matriks Konfusi

Prediksi	Aktual	
	0	1
0	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
1	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Nilai akurasi klasifikasi, sensitivitas, dan spesifisitas dihitung berdasarkan tabel matriks konfusi tersebut. Akurasi merupakan proporsi jumlah prediksi yang tepat dengan menghitung jumlah prediksi yang tepat dibagi total prediksi. Perhitungan nilai akurasi klasifikasi dirumuskan sebagai berikut (Fawcett, 2006):

$$Akurasi\ Klasifikasi = \frac{TP + TS}{TP + FP + FN + TN} \dots (2.4)$$

Adapun nilai sensitivitas yang berfungsi untuk mengukur tingkat keberhasilan klasifikasi ke dalam kelas 0, yaitu pasien *negative* Diabetes tipe 1 dan nilai spesifisitas mengukur tingkat keberhasilan klasifikasi ke dalam kelas 1, yaitu pasien positif Diabetes tipe 1. Nilai sensitivitas dan spesifisitas tersebut dirumuskan sebagai berikut:

$$Sensitivitas = \frac{TP}{TP + FN} \dots (2.5)$$

$$Spesifisitas = \frac{TN}{FP + TN} \dots (2.6)$$

Pada penelitian ini, tingginya nilai sensitivitas menunjukkan bahwa pemodelan klasifikasi pasien Diabetes tipe 1 dilakukan dengan baik sedangkan tingginya nilai spesifisitas menunjukkan bahwa pemodelan klasifikasi pasien positif Diabetes tipe 1 juga dilakukan dengan baik. Nilai sensitivitas dan spesifisitas yang semakin tinggi menunjukkan hasil prediksi klasifikasi yang semakin baik untuk kedua kelas respons. Besarnya nilai sensitivitas dan spesifisitas juga bergantung pada titik potong (*cut-point*) yang digunakan untuk memisahkan kedua kategori kelas respons. Bisa jadi model yang dihasilkan memiliki akurasi klasifikasi yang tinggi namun sensitivitas dan spesifisitasnya tidak seimbang ataupun sebaliknya model yang dihasilkan memiliki akurasi klasifikasi yang rendah namun sensitivitas dan spesifisitasnya seimbang (Hosmer et al. 2013).

Diabetes

Diabetes adalah penyakit kronis yang terjadi baik ketika pankreas tidak menghasilkan cukup insulin atau ketika tubuh tidak dapat secara efektif menggunakan insulin yang dihasilkannya. Insulin adalah hormon yang mengatur gula darah. *Hiperglikemia* atau gula darah yang tinggi adalah efek umum dari diabetes yang tidak terkontrol dan seiring waktu menyebabkan kerusakan serius pada banyak sistem tubuh, terutama saraf dan pembuluh darah. Diabetes memiliki dampak pada kesehatan organ lain karena seiring waktu diabetes dapat merusak jantung, pembuluh darah, mata, ginjal, dan saraf. Orang dewasa dengan diabetes memiliki peningkatan risiko serangan jantung dan stroke dua hingga tiga kali lipat dan akibat dari kerusakan akumulasi jangka panjang pada pembuluh darah kecil di retina menyebabkan *retinopati* diabetes, serta diabetes menjadi salah satu penyebab utama gagal ginjal. Dengan dampak yang sama diabetes dibagi menjadi 2 tipe klasifikasi penyebab terkena diabetes :

1. Diabetes tipe 2

Diabetes tipe 2 atau disebut tidak tergantung insulin merupakan akibat dari penggunaan

insulin yang tidak efektif oleh tubuh. Mayoritas penderita diabetes menderita diabetes tipe 2. Jenis diabetes ini sebagian besar merupakan hasil dari kelebihan berat badan dan ketidakaktifan fisik. Gejalanya mungkin mirip dengan diabetes tipe 1, tetapi seringkali kurang ditandai. Akibatnya, penyakit ini dapat didiagnosis beberapa tahun setelah onset, setelah komplikasi sudah muncul. Sampai saat ini, jenis diabetes ini hanya terlihat pada orang dewasa tetapi sekarang juga terjadi semakin sering pada anak-anak.

2. Diabetes tipe 1

Diabetes tipe 1 ditandai dengan produksi insulin yang kekurangan dan membutuhkan pemberian insulin setiap hari. Baik penyebab diabetes Tipe 1 maupun sarana untuk mencegahnya diketahui. Gejalanya termasuk ekskresi urin yang berlebihan (poliuria), haus (polidipsia), kelaparan konstan, penurunan berat badan, perubahan penglihatan, dan kelelahan. Gejala-gejala ini dapat terjadi tiba-tiba.

3. Pembahasan dan Diskusi

Data yang digunakan terdiri dari 520 pengamatan yang merupakan pasien di rumah sakit diabetes Sylhet, Bangladesh tahun 2020. Berdasarkan data yang ada, diketahui bahwa jumlah pasien terdiagnosis positif diabetes tipe 1 ada sebanyak 320 pasien sedangkan pasien terdiagnosis negatif ada sebanyak 200 pasien. Karakteristik pasien berdasarkan gejala yang dialami dapat dilihat pada Tabel 2 berikut:

Tabel 2. Karakteristik pasien berdasarkan gejala

Y	Diagnosis Diabetes	Nominal	0 = Negatif 1 = Positif
X1	<i>Age</i> (Umur pasien)	Rasio	Tahun
X2	<i>Gender</i> (Jenis Kelamin)	Nominal	0 = Laki-laki 1 = Perempuan
X3	<i>Polyuria</i> (Buang air kecil berlebihan)	Nominal	0 = Tidak 1 = Ya
X4	<i>Polydipsia</i> (Mudah Haus)	Nominal	0 = Tidak 1 = Ya
X5	<i>Sudden weight loss</i> (Penurunan berat badan mendadak)	Nominal	0 = Tidak 1 = Ya
X6	<i>Weakness</i> (Mudah Lelah)	Nominal	0 = Tidak 1 = Ya
X7	<i>Polyphagia</i> (Makan berlebih)	Nominal	0 = Tidak 1 = Ya
X8	<i>Genital thrush</i> (Sariawan Kelamin)	Nominal	0 = Tidak 1 = Ya
X9	<i>Visual blurring</i> (Pandangan kabur)	Nominal	0 = Tidak 1 = Ya
X10	<i>Itching</i> (Gatal-gatal)	Nominal	0 = Tidak 1 = Ya
X11	<i>Irritability</i> (Iritabilitas)	Nominal	0 = Tidak 1 = Ya
X12	<i>Delayed healing</i> (Penyembuhan luka berlangsung lambat)	Nominal	0 = Tidak 1 = Ya
X13	<i>Partial paresis</i> (setengah lumpuh)	Nominal	0 = Tidak 1 = Ya
X14	<i>Muscle stiffness</i> (ketidakseimbangan kerja otot)	Nominal	0 = Tidak 1 = Ya
X15	<i>Alopecia</i> (kebotakan)	Nominal	0 = Tidak 1 = Ya
X16	<i>Obesity</i> (Obesitas atau berat badan berlebih)	Nominal	0 = Tidak 1 = Ya

Maka berdasarkan uraian Tabel 2 diatas, dapat disimpulkan bahwa pasien yang mengalami gejala X3 (Buang air kecil berlebihan), X4 (Mudah haus), X5 (Penurunan berat badan mendadak), X6 (Mudah lelah), X7 (Makan berlebih), X9 (Pandangan kabur), X13 (setengah lumpuh) memiliki kemungkinan terdiagnosis diabetes tipe 1 lebih tinggi dibandingkan pasien yang tidak mengalami gejala tersebut.

Pemodelan *Rotation Forest*

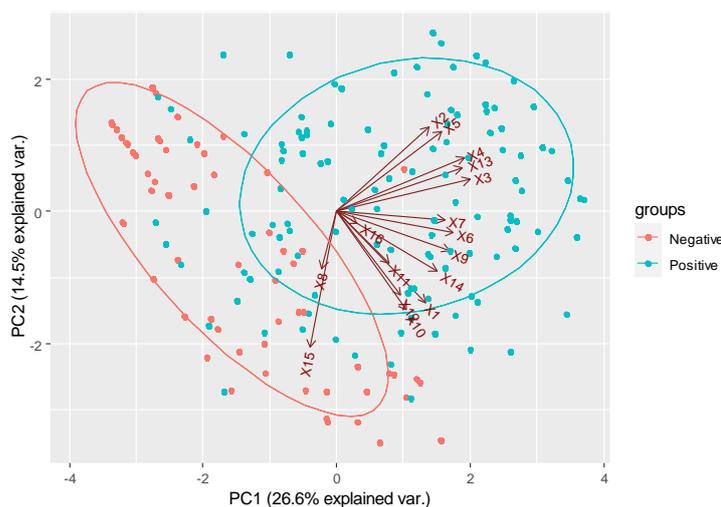
Dilakukan pemodelan *rotation forest* menggunakan data latih hasil perhitungan dengan koefisien komponen utama yang didapat dari Persamaan 2.1. Untuk mengetahui komponen mana yang berpengaruh dalam pemodelan *rotation forest* dalam mengklasifikasi pasien diabetes tipe 1 maka dilakukan perhitungan variabel penting menggunakan *Mean Decrease Gini* (MDG)

melalui Persamaan 2.3. Hasil perhitungan dapat dilihat pada Tabel 3 berikut.

Tabel 3. Nilai Mean Decrease Gini (MDG)

Komponen	Nilai Mean Decrease Gini
PC1	0,8717
PC2	0,6538
PC3	0,6493
PC4	0,6158
PC5	0,5019
PC6	0,5781
PC7	0,5243
PC8	0,5618
PC9	0,5573
PC10	0,5347
PC11	0,5448
PC12	0,5039
PC13	0,6215
PC14	0,5497
PC15	0,5057
PC16	0,5452

Dari Tabel 3 diketahui bahwa nilai *Mean Decrease Gini (MDG)* tertinggi adalah komponen utama ke-1 (PC1) yaitu sebesar 0,8717 artinya dalam pemodelan klasifikasi *rotation forest* menggunakan peubah PC1 sebesar 0,8717 atau 87,71%. Dimana nilai *Mean Decrease Gini (MDG)* tertinggi kedua adalah komponen utama ke-2 (PC2) yaitu sebesar 0,6538 atau 65,38%. Maka untuk mengetahui variabel yang digunakan dalam mengklasifikasi diabetes tipe 1 dipilih PC1 dan PC2 untuk dilakukan analisis lanjutan menggunakan analisis biplot. Gambar 4.3 memvisualisasikan hasil biplot.



Gambar 2. Visualisasi Biplot

Dari Gambar 2 diketahui, kelompok pasien terdiagnosis diabetes tipe 1 memiliki ciri-ciri gejala berikut X2(Gender), X5(Penurunan berat badan mendadak), X4(Mudah haus), X13(Setengah lumpuh), X3(Buang air kecil berlebih), X7(Makan berlebih), X6(Mudah lelah), X9(Pandangan kabur), X14(ketidakseimbangan kerja otot), X16(Obesitas), X11(Iritabilitas), X1(Umur), X12(Penyembuhan luka berlangsung lama) dan X10(Gatal-gatal). Dimana, variabel yang memiliki arah yang sama dan sudut yang lancip menandakan terjadinya hubungan positif

antar variabel dalam mengklasifikasi positif diabetes tipe 1 yaitu

1. Variabel X2(Gender), X5(Penurunan berat badan mendadak), X4(Mudah haus), X13(Setengah lumpuh), X3(Buang air kecil berlebih) artinya saat seorang pasien berjenis kelamin perempuan dengan gejala X4(Mudah haus), X13(Setengah lumpuh), X3(Buang air kecil berlebih) dapat diklasifikasikan kedalam pasien terdiagnosis positif diabetes tipe 1.
2. Variabel X7(Makan berlebih), X6(Mudah lelah), X9(Pandangan kabur), X14(ketidakeimbangan kerja otot), X16(Obesitas), X11(Iritabilitas), X1(Umur), X12(Penyembuhan luka berlangsung lama) dan X10(Gatal-gatal) artinya seorang pasien dengan gejala X7(Makan berlebih), X6(Mudah lelah), X9(Pandangan kabur), X14(ketidakeimbangan kerja otot), X16(Obesitas), X11(Iritabilitas), X1(Umur), X12(Penyembuhan luka berlangsung lama) dan X10(Gatal-gatal) dapat diklasifikasikan kedalam pasien terdiagnosis positif diabetes tipe 1.

Dari Gambar 2 juga dapat dilihat bahwa variabel X8(Sariawan Kelamin) dan X15(Kebotakan) tidak dalam kelompok ciri yang digunakan dalam pasien terdiagnosis positif diabetes tipe1 artinya X8(Sariawan Kelamin) dan X15(Kebotakan) bukan merupakan gejala yang digunakan untuk mengklasifikasi pasien diabetes tipe1

Untuk melihat kinerja model dalam mengklasifikasi dilakukan perhitungan evaluasi model menggunakan matriks konfusi pada data uji. Pada Tabel 4.4 merupakan matriks konfusi dari model yang didapat diketahui bahwa sebanyak 72 pasien negatif dan tepat diprediksi negatif, sebanyak 139 pasien positif dan tepat diprediksi positif, sebanyak 10 pasien negatif namun diprediksi positif, sebanyak 6 pasien positif namun diprediksi negatif.

Tabel 4. Matriks Konfusi Hasil Pemodelan

Prediksi	Aktual	
	Negatif	Positif
Negatif	72	6
Positif	10	139

Dari Tabel 4 dilakukan perhitungan nilai akurasi, sensitifitas dan spesifitas menggunakan Persamaan (2.4), Persamaan (2.5), dan Persamaan (2.6), sebagai berikut:

$$\text{Akurasi Klasifikasi} = \frac{72+139}{72+10+6+139} \times 100\% = 92,95\%$$

$$\text{Sensitivitas} = \frac{72}{72+10} \times 100\% = 87,80\%$$

$$\text{Spesifisitas} = \frac{139}{139+6} \times 100\% = 95,86\%$$

Hasil perhitungan di atas, menunjukkan bahwa dari model yang didapat memiliki ketepatan yang tinggi dalam mengklasifikasi pasien diabetes tipe 1 dapat dilihat dari nilai akurasi sebesar 92,95% menunjukkan bahwa model berhasil mengklasifikasi pasien diabetes tipe 1, sisanya sebesar 7,05% model gagal mengklasifikasi pasien diabetes tipe 1. Nilai sensitifitas sebesar 87,80% menunjukkan bahwa model berhasil mengklasifikasi pasien negatif diabetes tipe 1 sisanya sebesar 12,2% model gagal mengklasifikasi pasien negatif diabetes tipe 1. Serta nilai spesifisitas sebesar 95,86% menunjukkan bahwa model berhasil mengklasifikasi pasien positif diabetes tipe 1, sisanya sebesar 4,14% model gagal mengklasifikasi pasien positif diabetes tipe 1.

4. Kesimpulan

Dari penelitian ini dapat disimpulkan bahwa, Hasil penerapan metode *rotation forest* dalam mengklasifikasi pasien terdiagnosis diabetes tipe 1 berdasarkan gejala-gejala awal yang sering muncul yang dapat digunakan dalam mengklasifikasi pasien terdiagnosis diabetes melitus tipe 1 yaitu X2(Gender), X5(Penurunan berat badan mendadak), X4(Mudah haus), X13(Setengah lumpuh), X3(Buang air kecil berlebih), X7(Makan berlebih), X6(Mudah lelah), X9(Pandangan kabur), X14(ketidakeimbangan kerja otot), X16(Obesitas), X11(Iritabilitas), X1(Umur), X12(Penyembuhan luka berlangsung lama) dan X10(Gatal-gatal). Seorang pasien yang

memiliki gejala – gejala tersebut dapat diklasifikasikan sebagai penderita diabetes tipe 1. Dilihat dari hasil evaluasi model menunjukkan bahwa hasil pemodelan memiliki tingkat akurasi sebesar 92,95% yang berarti bahwa model berhasil mengklasifikasi pasien diabetes tipe 1, sisanya sebesar 7,05% model gagal mengklasifikasi pasien Diabetes tipe 1. Nilai sensitifitas sebesar 87,80% menunjukkan bahwa model berhasil mengklasifikasi pasien negatif Diabetes tipe 1, sisanya sebesar 12,20% model gagal mengklasifikasi pasien negatif Diabetes tipe 1. Nilai spesifisitas sebesar 95,86% menunjukkan bahwa model berhasil mengklasifikasi pasien positif Diabetes tipe 1, sisanya sebesar 4,14% model gagal mengklasifikasi pasien positif.

Acknowledge

Penelitian ini dapat dilaksanakan dengan baik berkat bantuan berbagai pihak, untuk itu peneliti mengucapkan terima kasih kepada ibu Dr. Lisnur Wachidah, Dra., M.Si. selaku pembimbing, Bapak Abdul Kudus, S.SI., M.SI., PH.D selaku Dekan FMIPA Universitas Islam Bandung, serta Bapak Dr. Aceng K. Mutaqin, M.T., M.Si. selaku Ketua Program Studi Statistika FMIPA Unisba, Dosen–dosen Program Studi Statistika Unisba yang telah banyak memberikan ilmu pengetahuannya. Serta seluruh keluarga, teman-teman sebimbingan dan sahabat-sahabat di program studi statistika angkatan 2017 Universitas Islam Bandung atas segala do'a dan kasih sayangnya.

Daftar Pustaka

- [1] Bartholomew, D. J. (2010). Principal Components Analysis. *International Encyclopedia of Education*, 374–377.
- [2] Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters* 27, 861–874.
- [3] Hosmer DW, Lemeshow S, Sturdivant RX. 2013. *Applied Logistic Regression Third Edition*. New Jersey (US): J Wiley.
- [4] Hair, J.F. Jr. , Anderson, R.E., Tatham, R.L., dan Black, W.C. 1998. *Multivariate Data Analysis, (5 th. Edition)*. Upper Saddle River, NJ: Prentice Hall.
- [5] International Diabetes Federation. 2019. *IDF DIABETES ATLAS : Ninth edition 2019*.
- [6] Kohavi R. 1995. A Study Of Cross Validation And Bootstrap For Accuracy Estimation And Model Selection. *Prosiding of the 14th International Joint Conference on Artificial Intelligence*. 2, 1137–1143.
- [7] M, Kuhn (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*. 28(5). 1-26
- [8] Rodriguez, J., L. Kuncheva, and C. Alonso (2006). Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(10), 1619–1630.
- [9] Sartono, B., & Syafitri, U. D. (2010). Metode Pohon Gabungan : Solusi Pilihan Untuk Mengatasi Kelemahan Pohon Regresi Dan Klasifikasi Tunggal (Ensemble Tree : An Alternative toward Simple Classification and Regression Tree). *Forum Statistika dan Komputasi*. 15(1), 1–7.
- [10] Sharma S. 1996. *Applied Multivariate Techniques*. New Jersey (US): J Wiley.
- [11] Wijaya, J., Soleh, A. M., & Rizki, A. (2018). Penanganan Data Tidak Seimbang pada Pemodelan Rotation Forest Keberhasilan Studi Mahasiswa Program Magister IPB. 2(2), 32–40.
- [12] Utama Muhammad Bangkit Riksa, Hajarisman Nusar. (2021). *Metode Pemilihan Variabel pada Model Regresi Poisson Menggunakan Metode Nordberg*. *Jurnal Riset Statistika*, 1(1), 35-42.