

# ***Modeling Early-stage Diabetes Mellitus using an Ensemble Learning Approach***

## **[Pemodelan Deteksi Dini Diabetes Mellitus menggunakan Pendekatan Ensemble Learning**

Syaikhina Usabili<sup>1)</sup>, Uce Indahyanti<sup>2)</sup>

<sup>1)</sup>Program Studi Manajemen Informasi Kesehatan, Universitas Muhammadiyah Sidoarjo, Indonesia

<sup>2)</sup>Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

\*Email Penulis Korespondensi: [191336300033@umsida.ac.id](mailto:191336300033@umsida.ac.id), [uceindahyanti@umsida.ac.id](mailto:uceindahyanti@umsida.ac.id)

**Abstract.** *Diabetes mellitus is characterized by hyperglycemia caused by the pancreas's inability to produce insulin properly. Diabetes has early-stage symptoms that can be used as a benchmark for determining whether a person has diabetes mellitus or not. Based on data from Sidoarjo Regional General Hospital, diabetes cases are the fourth most common of the 10 biggest diseases in Sidoarjo Regional General Hospital. The data used is data from 2022–2023 with selected attributes. The purpose of this research is to detect early symptoms of type 2 diabetes mellitus, where there are 5 signs that are detected very accurately. Data annotation is performed by proficient paramedics within their respective fields. This research uses the ensemble learning classification method with Rapidminer tools, conducts training and testing tests with a ratio of 60:40 on split data operators, and adds performance to produce accuracy values. The results obtained in the form of evaluation results with a Random Forest accuracy rate of 87.30%, where the accuracy level can be categorized as excellent classification.*

**Keywords – Prediction of early-stage, diabetes Mellitus, Rapidminer, Classification, Random Forest**

**Abstrak.** Diabetes Mellitus ditandai dengan hiperglikemia yang disebabkan oleh pankreas yang tidak dapat memproduksi insulin dengan baik. Diabetes memiliki gejala tahap awal yang dapat dijadikan sebagai tolak ukur seseorang terprediksi Diabetes Mellitus atau tidak. Berdasarkan data Rumah Sakit Umum Daerah Sidoarjo, kasus Diabetes menginjak urutan keempat terbanyak dari 10 penyakit terbesar yang ada di Rumah Sakit Umum Daerah Sidoarjo. Data yang digunakan merupakan data pada tahun 2022–2023 dengan atribut yang terpilih. Tujuan penelitian ini yaitu untuk mendeteksi dini gejala awal Diabetes Mellitus tipe 2 dimana terdapat 5 tanda yang terdeteksi sangat akurat. Anotasi data dilakukan oleh tenaga paramedis yang kompeten di bidangnya. Penelitian ini menggunakan metode klasifikasi ensemble learning dengan alat bantu Rapidminer untuk dan melakukan uji training dan uji testing dengan perbandingan 60:40 pada operator split data serta menambah performance untuk menghasilkan nilai akurasi. Hasil yang didapatkan berupa hasil evaluasi dengan tingkat akurasi Random Forest 87.30% yang dimana tingkat akurasi tersebut dapat dikategorikan sebagai excellent classification.

**Kata Kunci – Deteksi dini, Diabetes Mellitus, Rapidminer, Klasifikasi, Random Forest**

## **I. PENDAHULUAN**

Penyakit Diabetes Mellitus memiliki dampak negatif pada kinerja manusia yang dapat merugikan, dan menjadi penyebab masalah dalam kehidupan manusia. Penyakit ini tidak menular, tetapi berkembang paling cepat dan dikenal sebagai pembunuh paling cepat berkembang di dunia [1], saat ini penyakit Diabetes Mellitus menjadi keluhan Masyarakat karena penyakit tersebut mengalami peningkatan di setiap tahunnya. Menurut data International Diabetes Federation (IDF) pada tahun 2021, terdapat 537 juta penduduk hidup dengan mengidap Diabetes Mellitus dan Indonesia menempati peringkat kelima dari 10 terbesar negara yang terjangkau Penyakit Diabetes Mellitus, pada tingkatan IDF dengan catatan sebanyak 19,5 juta orang yang terjangkau penyakit Diabetes dan IDF memperkirakan Indonesia menjadi 28,6 juta penderita pada tahun 2045 mendatang [2], dan laporan Riskesdas tahun 2018 menunjukkan bahwa penderita Diabetes Mellitus dapat terjangkau mulai dari umur 15-74 tahun dan 75 tahun keatas dengan menunjukkan hasil 2% yang mengalami peningkatan dibandingkan dengan prevalensi diabetes melitus pada tahun 2013 yang mencapai angka 1,5%, namun prevalensi diabetes melitus jika menurut hasil pemeriksaan gula darahnya justru lebih meningkat dari 6,9% pada tahun 2013 menjadi 8,5% pada tahun 2018 [3].

Data terbaru di Rumah Sakit Umum Daerah Sidoarjo tahun 2022 menunjukkan bahwa penyakit Diabetes Mellitus termasuk penyakit dengan peringkat ke 4 dari 10 penyakit terbesar di Rumah Sakit Umum Daerah Sidoarjo dengan jumlah sebanyak 1392 penderita dan data Rumah Sakit Umum Daerah Sidoarjo menunjukkan bahwa penderita Diabetes tergolong menjadi beberapa kategori dalam kode penyakit, diantaranya yaitu dengan kode E11.9 (*type 2*

*Diabetes Mellitus without complications*), E11.8 (*type 2 Diabetes Mellitus with unspecified complications*), E11.5 (*type 2 Diabetes Mellitus with peripheral circulatory complications*), E11.3 (*type 2 Diabetes Mellitus with ophthalmic complications*), E11.2 (*type 2 Diabetes Mellitus with renal complications*).

Diabetes Mellitus dikategorikan menjadi 2 tipe, dimana definisi pada Diabetes Mellitus tipe 1 merupakan gangguan metabolisme kronis dapat ditandai dengan hiperglikemia yang berkembang Ketika pancreas tidak dapat memproduksi insulin dengan baik. Adapun Diabetes Mellitus tipe 2 yang disebabkan oleh pola makan yang tidak teratur sehingga dapat menghasilkan kadar gula darah yang tinggi dan mempengaruhi seberapa baik tubuh menggunakan insulin. Gejala awal diabetes meliputi berbagai gejala seperti sering buang air kecil, rasa haus yang berlebihan, kadar gula darah tinggi, penurunan badan, pemulihan tubuh yang buruk [4].

Tubuh akan berpacu untuk memproses lemak dan protein untuk diubah menjadi energi ketika tubuh tidak dapat memperoleh energi yang cukup dari gula karena kekurangan insulin. Disisi lain, kadar gula darah 180 mg/dl dianggap normal bagi penderita Diabetes Mellitus, namun jika kadar gula melebihi 200 mg/dl, Kadar gula darah normal untuk penyandang Diabetes Mellitus memiliki kadar gula 180 mg/dl. namun apabila kadar gula darah >200 mg/dl maka gula darah tersebut dianggap tinggi dan individu dengan diabetes melitus yang tidak terkontrol dapat kehilangan hingga 500 gram glukosa (atau 2.000 kalori per hari) melalui sistem urin dalam 24 jam. Kemudian, kaki kesemutan, gatal, atau luka yang tidak kunjung sembuh merupakan gejala tambahan yang mungkin muncul dan biasanya terkait dengan masalah. Pada wanita, pruritus vulva, atau gatal-gatal di area selangkangan, terkadang disertai rasa sakit, sementara pada pria, balanitis, atau nyeri di ujung penis, dialami. Kadar gula darah yang tidak terkontrol juga dapat menyebabkan diabetes dengan berbagai komplikasi, seperti penyakit jantung koroner, obesitas, stroke, gangguan mata, gangren pedis, ginjal, dan saraf [5].

Etiologi dari penyakit diabetes mellitus yaitu dari faktor genetik dan faktor lingkungan, yang mana etiologi tersebut dapat membuahkan gejala sehingga dapat dijadikan sebagai tahap awal prediksi terjangkitnya penyakit Diabetes [6] dengan penentuan tahap awal dari penyakit Diabetes Mellitus yang akan diterapkan menggunakan machine learning dengan teknik pengklasifikasian data mining [7]. Klasifikasi ini dapat digunakan pada penelitian dengan menggunakan data mining sebagai alat pengolahan data dan alat ukur yang dapat memberikan panduan untuk mengidentifikasi penyakit tertentu, dalam hal ini adalah Diabetes Mellitus[8].

Alasan menggunakan Klasifikasi data mining yaitu untuk mengetahui berapa banyaknya persentase tingkat keakuratan pada penyakit Diabetes Mellitus, menganalisis gejala tahap awal yang memungkinkan terjadinya Diabetes Mellitus di Rumah Sakit serta metode yang digunakan juga sangat mudah dipahami sehingga peneliti tidak kesusahan dalam mendeskripsikan hasil dalam bentuk pohon keputusan, *Random Forest* menggunakan beberapa gabungan pohon keputusan dengan kedua nya memiliki nilai akhir bersifat numerik, metode tersebut akan dijadikan sebagai metode *ensemble learning*, dimana *ensemble learning* merupakan penggunaan beberapa model secara bersamaan [9].

Pada penelitian terdahulu mengenai Klasifikasi data mining untuk diagnosa Diabetes Retinopathy dengan algoritma *Classification and Regression Trees (CART)* dan dataset berupa data public dari UCI Repository *Learning* yang diperoleh dari Universitas Debrecen, Hungaria menunjukkan hasil sebesar 63.4231%, dengan nilai precision 0.64%, nilai Recall 0.634%, dan nilai F-Measure 0,634% [8].

Penelitian lain mengenai Klasifikasi data mining untuk diagnosa Diabetes Mellitus menggunakan model Klasifikasi *Random Forest* dan mengambil dataset berupa data sekunder atau data public yang diambil dari UCI *Machine Learning* yang bersumber dari data di *Hospital in Sylhet, Bangladesh* memiliki hasil yaitu menyatakan bahwa diantara algoritma yang ada, *Random Forest* lebih unggul dengan hasil 97,88%. Temuan ini secara tepat dan metodis dikonfirmasi dengan menggunakan kurva *Receiver Operating Characteristic (ROC)* [9].

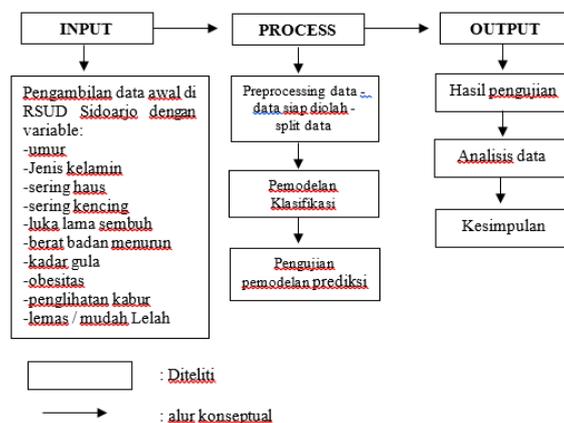
Penelitian lainnya mengenai Klasifikasi data mining untuk diagnosa Diabetes menggunakan metode C45, *Random Forest*, SVM dan dataset yang digunakan berupa data sekunder atau data public yang diperoleh dari UCI Pima Indians Diabetes Datasets di Kaggle Dataset Repository, model klasifikasi SVM menggunakan teknik resampling yang menggabungkan over dan under-sampling merupakan model yang memiliki performa paling baik dengan nilai AUC (*Area Under Curve*) sebesar 0.80 [10].

Penelitian ini bertujuan untuk mengetahui tingkat keakuratan prediksi tahap awal pada penyakit Diabetes Mellitus, Untuk melakukan uji coba kesesuaian penggunaan Rapidminer dalam melakukan prediksi resiko Diabetes dengan menerapkan metode *Random Forest (ensemble learning)*. Prediksi tahap awal Diabetes dilakukan dengan pemilihan atribut – atribut yang sesuai dengan Diabetes Mellitus dan akan diuji melalui pohon keputusan dan menghasilkan nilai akurasi, *Confusion matrix* yang akan menentukan kelayakan prediksi pada tahap awal penyakit Diabetes Mellitus tipe 2.

Berdasarkan permasalahan diatas, penelitian ini yang mengambil judul “Pemodelan deteksi dini Diabetes Mellitus menggunakan pendekatan *Ensemble Learning*” dilakukan dengan tujuan untuk mengetahui tingkat keakuratan prediksi, serta menganalisis deteksi dini yang memungkinkan terjadinya Diabetes Mellitus dengan tipe 2.

## II. METODE

Dalam penelitian ini menggunakan metode penelitian kuantitatif, secara khusus data rekam medis Rumah Sakit Umum Daerah Sidoarjo diobservasi sebagai bagian dari instrumen penelitian. Tahapan penelitian dimulai dari identifikasi masalah, tahap pengumpulan data, anotasi dataset oleh pakar, tahap penambangan data menggunakan metode *ensemble learning*, pengujian pemodelan prediksi menggunakan teknik *Confusion Matrix*, tahap analisis data dan kesimpulan. Tahapan – tahapan yang akan dilakukan dalam penelitian ini dilakukan sesuai dengan urutan dibawah ini :



Gambar 1. tahapan persiapan dan pengolahan data penelitian

### Persiapan data

Tahap ini menjelaskan langkah persiapan data termasuk *preprocessing* sampai dengan data siap diolah. Data penelitian menggunakan data rekam medis di Rumah Sakit Umum Sidoarjo pada bulan Desember 2022 – Maret 2023. Penderita penyakit Diabetes Mellitus di Rumah Sakit Umum Daerah Sidoarjo pada bulan Desember 2022 sampai dengan Maret 2023 sebanyak 668 penderita dengan melihat anamnesa, catatan perkembangan pasien terintegrasi, dan ringkasan pasien pulang. Atribut data menggambarkan gejala tahap awal penyakit Diabetes Mellitus, yang merupakan aspek yang diteliti apakah pasien tersebut terjangkit penyakit Diabetes Mellitus atau tidak.

Data dianotasi oleh tenaga paramedis yang kompeten di bidangnya, yaitu oleh dr. Dian Samudra, Sp. PD. Atribut yang akan diteliti berupa umur, jenis kelamin, sering haus, sering kencing, berat badan menurun, luka lama sembuh, kadar gula darah meningkat, obesitas, penglihatan kabur, lemas/mudah Lelah. Atribut tersebut dipilih karena sangat signifikan dengan gejala awal pada penyakit Diabetes Mellitus. Data yang diambil dari rekam medis berupa hardfile dan csv, dan perlu diubah dalam format excel sehingga lebih mudah dalam proses pada machine learning [11].

Pada proses dilakukan *preprocessing* data dimana data yang telah diambil akan di sortir dengan cara penghapusan data yang tidak digunakan guna meningkatkan dan memastikan data tersebut dapat diproses dengan baik dengan melalui *preprocessing* menghasilkan data dengan 399 Class *Positive* dan 269 Class *Negative* dengan catatan, class *negative* memiliki gejala untuk Diabetes Mellitus namun tidak terjangkit penyakit tersebut. Data yang telah melewati tahapan *preprocessing*, akan dibagi menjadi dua bagian dengan menggunakan split data berupa data training dengan rasio desimal 0.6 dan data testing dengan rasio desimal 0.4 yang akan menghasilkan data berupa performansi akurasi *Confusion Matrix*, *True Positive* dan *True Negative* dalam bentuk persentase [12].

### Data Mining

Data mining adalah salah satu proses untuk menciptakan suatu pengetahuan dengan menggunakan database sebagai Upaya untuk mengumpulkan suatu informasi yang bermanfaat, data mining adalah proses untuk mengekstraksi informasi suatu informasi yang tersembunyi, pola, dan koneksi data spesifik dari ide suatu prediksi [13] Data mining juga merupakan serangkaian proses untuk mengumpulkan data atau informasi yang berguna dengan melakukan penambangan data dari berbagai aplikasi yang diterapkan yang juga memiliki berbagai dampak di berbagai bidang, khususnya di bidang Kesehatan [14], beberapa strategi, salah satunya yaitu Teknik Klasifikasi, data mining dapat digunakan dalam penggalian data yang dapat menangani Kumpulan data berukuran besar [15].

### Klasifikasi

Model yang ditemukan dengan memperjelas ide kelas data adalah proses dari klasifikasi yang bertujuan untuk memetakan sebuah data kedalam beberapa kelas yang telah terpilih [16]. Klasifikasi juga merupakan proses untuk

menciptakan properti – properti pada suatu proyek di sebuah basis data dan mengklasifikasikan ke dalam kelas – kelas yang sudah ditetapkan dengan membedakan atribut kedalam beberapa kelompok tingkatan [17]. Klasifikasi digunakan secara meluas dalam bidang Kesehatan, dimana Teknik Klasifikasi ini mampu memproses data dengan jumlah yang sangat besar sehingga dapat diterapkan secara efisien dalam ruang lingkup medis dan juga sebagai hasil pembelajaran mesin dalam suatu penelitian [18].

### **Random Forest**

*Random Forest* merupakan metode pengklasifikasian yang digunakan untuk memperbaiki proses dari metode sebelumnya dengan jenis yang sama, dengan menggunakan tools yaitu beberapa akar pohon keputusan dan memiliki hasil yang lebih luas dibandingkan dengan pohon keputusan lainnya. *Random Forest* memiliki beberapa parameter yaitu *myr* dan *n tree* untuk proses pemilahan jumlah pohon yang akan digunakan [19]. *Random Forest* juga merupakan pohon keputusan yang dimasukkan secara eksplisit dalam proses pembangunan model dari setiap pohon nya, dan *Random Forest* dapat dilihat juga sebagai generalisasi dari metode bagging yang secara tidak langsung menambahkan pohon keputusan secara acak dalam pembuatan modelnya [20].

### **Imbalanced data**

*Imbalanced data* dapat mengumpulkan data dengan proporsi kelas yang tidak seimbang dengan kelas yang disebut mayoritas dan minoritas sehingga hasil yang dihasilkan dari suatu model tidak rata dan menyebabkan suatu model cenderung overfitting dan performa tidak baik [21]. Salah satu Teknik yang digunakan untuk menyetarakan data dengan menggunakan teknik *balancing* menggunakan SMOTE Upsampling. SMOTE Upsampling merupakan teknik yang digunakan untuk mengatasi ketidakseimbangan atau *imbalance* data dalam kelas klasifikasi guna meningkatkan jumlah *class* minoritas dan membantu meningkatkan kinerja model klasifikasi [22].

### **Evaluasi model**

*Confusion matrix* merupakan kinerja pengklasifikasian atau evaluasi model dari klasifikasi multi class, yang dimana data yang telah diolah dapat menjadi bagian dari satu class pada titik tertentu, *confusion matrix* dijadikan sebagai alat yang efektif untuk membantu memberikan informasi terkait hasil class yang memiliki performa yang berupa *True Positive*, *True Negative* atau model yang memiliki jumlah objek untuk meramal dengan benar dan salah dengan menghitung berbagai ukuran evaluasi seperti *Precision* dan *recall* [23]. *Confusion Matrix* memanfaatkan informasi yang disajikan dari hasil yang telah diolah oleh pengklasifikasi untuk memprediksi, mengevaluasi dan merepresentasikan nilai akurasi pada hasil yang ada [24].

## **III. HASIL DAN PEMBAHASAN**

### **Pengumpulan data**

Data yang digunakan dalam penelitian ini berupa data yang diambil dari Rumah Sakit Umum Daerah Sidoarjo pada bulan Desember 2022 – Maret 2023 sebanyak 668 data dan data akan diproses menggunakan atribut – atribut berikut :

**Table 1.** Atribut Data

No	Nama atribut	Keterangan
1	Umur	21 - 86 tahun
2	Jenis kelamin	L / P
3	Sering haus	Ya / Tidak
4	Sering kencing	Ya / Tidak
5	Luka lama sembuh	Ya / Tidak
6	Berat badan menurun	Ya / Tidak
7	Kadar gula meningkat	Ya / Tidak
8	Obesitas	Ya / Tidak
9	Penglihatan kabur	Ya / Tidak
10	Lemas/ mudah lelah	Ya / Tidak
11	Class	Positif / Negatif

Sumber : Data Rumah Sakit Umum Daerah Sidoarjo Desember 2022 – Maret 2023

### **Preprocessing Data**

Tahap *preprocessing* data menggunakan alat bantu *Rapidminer*. Sebelum dimasukkan ke dalam pemodelan, data dalam format xls dibersihkan dan diseimbangkan terlebih dahulu. Dataset terdiri dari 399 data *Class Positive* dan 269 data *Class Negative* sehingga terdapat ketidakseimbangan kelas (*imbalanced*). Selanjutnya dilakukannya, dan memilih

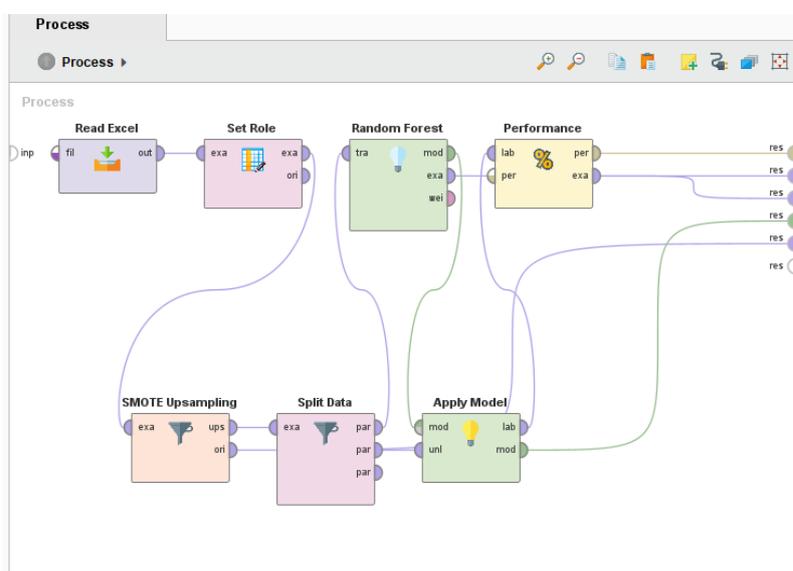
tipe dan juga memberikan label pada tabel yang akan menjadi parameter pada hasil dari perhitungan klasifikasi, serta menambahkan split data untuk uji pemodelan prediksi dengan menggunakan uji *training* dengan rasio 0.6 dan uji *testing* dengan rasio 0.4, kemudian memilih *apply model* sebagai bentuk evaluasi dari kinerja suatu model yang telah terpilih.

### Hasil dan Analisis Data

Pengolahan data dilakukan dengan laptop personal *Intel Core i3*, 8GB RAM, sistem operasi *Windows 10* dan *Rapidminer 10.0.000*

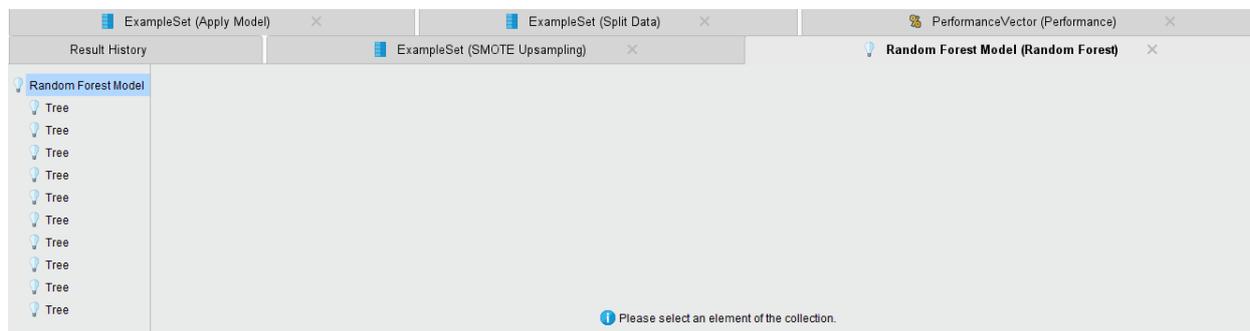
#### Pemodelan klasifikasi menggunakan Random Forest

Pada Gambar 2 ditampilkan proses klasifikasi menggunakan algoritma Random Forest. Beberapa fitur yang perlu diatur pada Rapid Miner antara lain: Read Excel, Set role untuk menargetkan atribut dengan role label pada Class, SMOTE Upsampling, Split Data, Random Forest, Apply Model, dan Performance.



**Gambar 2.** Desain proses pengklasifikasian *Random Forest*

Adapun hasil dari pengklasifikasian *Random Forest* dalam bentuk pohon keputusan dapat dilihat pada beberapa gambar dibawah :



**Gambar 3.** Hasil 10 pohon keputusan *Random forest*

Gambar diatas merupakan model dari *random forest* dengan 10 kali *maximal depth* atau 10 kali perputaran untuk *variable* secara acak, dan berikut adalah salah satu hasil terbaik dari 10 model *random forest*



1. Jika Umur <36, lemas/mudah lelah = ya, penglihatan kabur = ya, Jenis Kelamin = L, maka class menunjukkan hasil Positif
2. Jika Umur >27, berat badan menurun = ya, sering kencing = ya, kadar gula meningkat = ya, maka class menunjukkan hasil positif.

Deskripsi pertama menyatakan bahwa pada umur >36 dengan melihat jenis kelamin yaitu Laki – laki , dapat menjadikan seseorang berpeluang mengidap penyakit Diabetes Mellitus tipe 2 yang diikuti dengan lemas/mudah lelah, dan penglihatan kabur yang dimiliki pasien dengan gejala tersebut sehingga dikategorikan kedalam kategori positif. Sedangkan pada deskripsi ke 2 menyatakan bahwa pasien dengan umur >27 dan mengalami berat badan menurun, sering kencing, kadar gula meningkat >200 mg/dl menjadikan seseorang berpeluang menderita Diabetes Mellitus tipe 2 dikategorikan kedalam kategori positif.

Adapun hasil dari confusion matrix atau uji tingkat akurasi dari klasifikasi *Random Forest* dapat dilihat pada tabel 3 dibawah ini :

**Table 2.** Matrik Konfusi

	True Positive	True Negative	Class Precision
<b>Pred. Positive</b>	205	14	93.61%
<b>Pred. Negative</b>	51	242	82.59%
<b>Class Recall</b>	80.08%	94.53%	

Tabel diatas menunjukkan bahwa nilai akurasi sebesar 87.30%, yang dimana jumlah data benar atau true positif pada prediksi positif sebesar 205 dan data yang tidak sesuai sebanyak 14 data pada *true negative*, sedangkan pada prediksi *negative* terdapat data benar sebanyak 51 data dan juga 242 data lainnya pada data yang tidak sesuai sehingga menghasilkan *class recall* pada data benar yaitu 93.61% dan data tidak sesuai sebesar 82.59%. Tingkat akurasi yang ditujukan untuk menilai hasil terhadap klasifikasi bernilai 0.80- 0.90 = *good classification*, 0.90 – 1.00 = *excellent classification* [25].

#### IV. SIMPULAN

Pemodelan untuk deteksi dini diabetes mellitus membuahkan hasil dengan model klasifikasi ensemble learning dan *imbalanced data* menggunakan SMOTE Upsampling pada penyakit Diabetes Mellitus tipe 2 klasifikasi *Random Forest* berdasarkan Confusion Matrix didapatkan hasil sebesar 87.30% yang dimana nilai yang dihasilkan dari *Random Forest* dikategorikan sebagai kategori klasifikasi sangat baik atau *excellent classification*. Adapun bagi Rumah Sakit Umum Daerah Sidoarjo, determinan pasien terdiagnosa Diabetes Mellitus tipe 2 terdeteksi dini dengan gejala tahap awal berupa luka lama sembuh, sering kencing, lemas/mudah lelah, kadar gula meningkat dan umur diatas 30 tahun. Hasil dari Penelitian ini dapat dijadikan sebagai informasi dan masukan bagi Rumah Sakit Umum Daerah Sidoarjo untuk mempermudah melakukan perkiraan atau prediksi awal pada penyakit Diabetes Mellitus tipe 2 menggunakan metode Klasifikasi Ensemble Learning, namun perlu diadakannya validasi tingkat lanjut untuk memberikan generalisabilitas terhadap penyakit Diabetes Mellitus tipe 2. Untuk penelitian selanjutnya untuk bisa ditingkatkan kembali agar prediksi dari gejala awal Diabetes Mellitus dengan 2 tipe yang berbeda dengan metode dan model yang lebih akurat dan lebih detail serta meluas dan sebagai referensi untuk menambah informasi dan memperluas pengetahuan agar terciptanya informasi baru.

#### UCAPAN TERIMA KASIH

Puji dan Syukur kami panjatkan kepada Tuhan Yang Maha Esa, atas rahmat dan karunia-Nya, sehingga peneliti dapat menyelesaikan penelitian ini dengan baik. Peneliti mengucapkan rasa terimakasih sebesar – besarnya kepada Dekan Fakultas Ilmu Kesehatan Muhammadiyah Sidoarjo, Dosen pembimbing Prodi Manajemen Informasi Kesehatan Muhammadiyah Sidoarjo, serta seluruh informan dan kerabat dari Rumah Sakit Umum Daerah Sidoarjo yang telah memberikan bimbingan, dorongan, dukungan selama proses penelitian ini.

## REFERENSI

- [1] B. E. Nyarko, R. S. Amoah, and A. Crimi, "Boosting diabetes and pre-diabetes detection in rural Ghana [version 2; peer review: 2 approved]," *F1000 Res.*, vol. 8, p. 19, Aug. 2019, doi: <https://doi.org/10.12688/f1000research.18497.2>.
- [2] D. Magliano and E. J. Boyko, *IDF diabetes atlas*, 10th edition. Brussels: International Diabetes Federation, 2021.
- [3] S. Bessy, "Hasil Riskesdas 2018 Kemestrian Kesehatan," presented at the hasil riskesdas, Kemkes, 02112018. [Online]. Available: [https://kesmas.kemkes.go.id/assets/upload/dir\\_519d41d8cd98f00/files/Hasil-riskesdas-2018\\_1274.pdf](https://kesmas.kemkes.go.id/assets/upload/dir_519d41d8cd98f00/files/Hasil-riskesdas-2018_1274.pdf)
- [4] W. Yusnaeni and W. Widiarina, "Penerapan Algoritma C4.5 Dalam Prediksi Resiko Diabetes Tahap Awal (Early Stage Diabetes)," *J. Tek. Komput.*, vol. 8, no. 1, pp. 56–60, Jan. 2022, doi: 10.31294/jtk.v8i1.11566.
- [5] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, Jul. 2020, doi: 10.33096/ijodas.v1i2.11.
- [6] O. Ozougwu, "The pathogenesis and pathophysiology of type 1 and type 2 diabetes mellitus," *J. Physiol. Pathophysiol.*, vol. 4, no. 4, pp. 46–57, Sep. 2013, doi: 10.5897/JPAP2013.0001.
- [7] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. SISKOM-KB Sist. Komput. Dan Kecerdasan Buatan*, vol. 4, no. 1, pp. 15–21, Oct. 2020, doi: 10.47970/siskom-kb.v4i1.169.
- [8] P. Subarkah, "Penerapan Algoritme Klasifikasi Classification And Regression Trees (Cart) Untuk Diagnosis Penyakit Diabetes Retinopathy," *MATRIK J. Manaj. Tek. Inform. Dan Rekayasa Komput.*, vol. 19, no. 2, pp. 294–301, May 2020, doi: 10.30812/matrik.v19i2.676.
- [9] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *SISTEMASI*, vol. 10, no. 1, p. 163, Jan. 2021, doi: 10.32520/stmsi.v10i1.1129.
- [10] W. Nugraha and R. Sabaruddin, "Teknik Resampling untuk Mengatasi Ketidakseimbangan Kelas pada Klasifikasi Penyakit Diabetes Menggunakan C4.5, Random Forest, dan SVM," *Techno.Com*, vol. 20, no. 3, pp. 352–361, Aug. 2021, doi: 10.33633/tc.v20i3.4762.
- [11] B. T. R. Doni, S. Susanti, and A. Mubarak, "PENERAPAN DATA MINING UNTUK KLASIFIKASI PENYAKIT HEPATOCELLULAR CARCINOMA MENGGUNAKAN ALGORITMA NAÏVE BAYES," *J. Responsif Ris. Sains Dan Inform.*, vol. 3, no. 1, pp. 12–19, Feb. 2021, doi: 10.51977/jti.v3i1.403.
- [12] A. K. F. Aidia, P. J. Amelia, and V. R. Setyaning Nastiti, "Prediksi Jumlah Pasien Covid-19 Dengan Menggunakan Klasifikasi Algoritma Machine Learning," *SINTECH Sci. Inf. Technol. J.*, vol. 5, no. 2, pp. 165–172, Oct. 2022, doi: 10.31598/sintechjournal.v5i2.1163.
- [13] R. Ghorbani and R. Ghousi, "Predictive data mining approaches in medical diagnosis: A review of some diseases prediction," *Int. J. Data Netw. Sci.*, pp. 47–70, 2019, doi: 10.5267/j.ijdns.2019.1.003.
- [14] F. Aris, "Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi," vol. 1, no. 1, 2019.
- [15] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, p. 640, Apr. 2021, doi: 10.30865/mib.v5i2.2937.
- [16] F. Elfaladonna and A. Rahmadani, "ANALISA METODE CLASSIFICATION-DECISSION TREE DAN ALGORITMA C.45 UNTUK MEMREDIKSI PENYAKIT DIABETES DENGAN MENGGUNAKAN APLIKASI RAPID MINER," *SINTECH Sci. Inf. Technol. J.*, vol. 2, no. 1, pp. 10–17, Apr. 2019, doi: 10.31598/sintechjournal.v2i1.293.
- [17] D. R. Ente, S. A. Thamrin, S. Arifin, H. Kuswanto, and A. Andreza, "KLASIFIKASI FAKTOR-FAKTOR PENYEBAB PENYAKIT DIABETES MELITUS DI RUMAH SAKIT UNHAS MENGGUNAKAN ALGORITMA C4.5," *Indones. J. Stat. Its Appl.*, vol. 4, no. 1, pp. 80–88, Feb. 2020, doi: 10.29244/ijsa.v4i1.330.
- [18] M. Tarawneh and O. Embarak, "Hybrid Approach for Heart Disease Prediction Using Data Mining Techniques," in *Advances in Internet, Data and Web Technologies*, vol. 29, L. Barolli, F. Xhafa, Z. A. Khan, and H. Odhabi, Eds., in Lecture Notes on Data Engineering and Communications Technologies, vol. 29. , Cham: Springer International Publishing, 2019, pp. 447–454. doi: 10.1007/978-3-030-12839-5\_41.
- [19] M. Syukron, R. Santoso, and T. Widiharih, "PERBANDINGAN METODE SMOTE RANDOM FOREST DAN SMOTE XGBOOST UNTUK KLASIFIKASI TINGKAT PENYAKIT HEPATITIS C PADA IMBALANCE CLASS DATA," *J. Gaussian*, vol. 9, no. 3, pp. 227–236, Aug. 2020, doi: 10.14710/j.gauss.v9i3.28915.
- [20] C. C. Aggarwal, *Data Mining: The Textbook*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.

- [21] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Inform. J. Ilmu Komput.*, vol. 18, no. 3, p. 239, Dec. 2022, doi: 10.52958/iftk.v18i3.4694.
- [22] F. D. Astuti and F. N. Lenti, "Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN," vol. 13, no. 1, 2021.
- [23] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label Classifier Performance Evaluation with Confusion Matrix," *Comput. Sci.*.
- [24] P. Cavalin and L. Oliveira, "Confusion Matrix-Based Building of Hierarchical Classification," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, vol. 11401, R. Vera-Rodriguez, J. Fierrez, and A. Morales, Eds., in Lecture Notes in Computer Science, vol. 11401. , Cham: Springer International Publishing, 2019, pp. 271–278. doi: 10.1007/978-3-030-13469-3\_32.
- [25] F. Gorunescu, *Data Mining*, vol. 12. in Intelligent Systems Reference Library, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-19721-5.

**Conflict of Interest Statement:**

*The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.*