

## PENERAPAN ALGORITMA *ITERATIVE DICHOTOMISER 3* (ID3) DALAM KLASIFIKASI FAKTOR RISIKO PENYAKIT DIABETES MELITUS

### *Application of The Iterative Dichotomizer 3 (ID3) Algorithm in The Classification of Risk Factors for Diabetes Mellitus*

Ferdina<sup>1</sup>, Neva Satyahadewi<sup>2\*</sup>, Dadan Kusnandar<sup>3</sup>

<sup>1,2,3</sup>Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Tanjungpura  
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, Kalimantan Barat, Indonesia

*E-mail Corresponding Author: neva.satya@math.untan.ac.id*

---

**Abstrak:** Algoritma *Iterative Dichotomizer 3* (ID3) adalah sebuah metode yang digunakan untuk membuat pohon keputusan. Pohon keputusan merupakan salah satu metode klasifikasi dengan model prediksi menggunakan struktur pohon. *International Diabetes Federation* pada tahun 2021 menyatakan bahwa Indonesia menduduki posisi kelima dalam kasus diabetes terbanyak, dengan jumlah penyandang diabetes sebanyak 19,47 juta penduduk. Tujuan penelitian ini adalah menerapkan Algoritma ID3 dan menentukan akurasi dalam klasifikasi faktor risiko diabetes melitus. Data yang digunakan dalam penelitian ini adalah data hasil tes kesehatan karyawan di Kota Banyuwangi, Jember, Kediri, Madiun, Malang, Sidoarjo dan Surabaya. Data kemudian dibagi menjadi data *training* dan data *testing*. Atribut klasifikasi yang digunakan dalam penelitian ini adalah jenis kelamin, usia, gula darah sewaktu (GDS), *High Density Lipoprotein* (HDL), *Low Density Lipoprotein* (LDL), dan trigliserida. Berdasarkan hasil pengujian klasifikasi Algoritma ID3 menggunakan data *training* dan *software* R Studio, diperoleh variabel dengan nilai *information gain* tertinggi adalah GDS. Berdasarkan hasil perhitungan, nilai akurasi yang diperoleh dari metode Algoritma ID3 dan data *testing* adalah sebesar 90%. Berdasarkan nilai akurasi yang diperoleh, dapat disimpulkan bahwa keakuratan Algoritma ID3 tergolong baik dalam klasifikasi faktor risiko penyakit diabetes melitus.

**Kata Kunci:** Algoritma ID3, *Confusion Matrix*, Diabetes, Klasifikasi.

**Abstract:** *Algorithm Iterative Dichotomizer 3 (ID3) is used to create decision trees. Create a decision tree. A decision tree is one of the classification methods with a prediction model using a tree structure. International Diabetes Federation in 2021 stated that Indonesia occupied the fifth position in the most cases of diabetes, with a total of 19.47 million people with diabetes. as many as 19.47 million people. The purpose of this research is to apply the ID3 algorithm and determine its accuracy in the classification of risk factors for diabetes mellitus. The data used in this research is data from employee health examinations in the cities of Banyuwangi, Jember, Kediri, Madiun, Malang, Sidoarjo, and Surabaya. The data is then divided into training data and testing data. The classification attributes used in this study are gender, age, temporary blood sugar (GDS), high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides. Based on the test results classification using training data and the R Studio software, the variables with the highest information gain values are GDS. Based on the calculation results, the accuracy value obtained from the ID3 Algorithm method and testing data is 90%. Based on the accuracy values obtained, it can be concluded that the accuracy of the ID3 Algorithm is relatively good in the classification of risk factors for diabetes mellitus. Classification of risk factors for diabetes mellitus.*

**Keywords:** *Classification, Confusion Matrix, Diabetes, ID3 Algorithm.*

---

## 1. PENDAHULUAN

Kesehatan merupakan aspek penting dalam kehidupan. Permasalahan yang dihadapi dalam meningkatkan derajat kesehatan masyarakat banyak berkaitan dengan gaya hidup yang kurang sehat. Gaya hidup yang kurang sehat dapat memicu munculnya berbagai macam penyakit. Masalah yang sering muncul dalam gaya hidup masyarakat tersebut adalah penyakit diabetes melitus [1]. Diabetes melitus merupakan penyakit kronis yang ditandai dengan peningkatan kadar gula darah [2]. Diabetes Melitus atau biasa disebut dengan kencing manis merupakan suatu kelainan metabolisme tubuh yang disebabkan oleh tidak berfungsinya hormon insulin dalam tubuh secara efektif dalam mengatur keseimbangan gula darah. Hal ini menyebabkan peningkatan konsentrasi kadar gula di dalam darah (hiperglikemia) [3]. Pada tahun 2022, *International Diabetes Federation* (IDF)

menyatakan bahwa Indonesia menduduki posisi kelima dalam kasus diabetes terbanyak setelah China, India, Pakistan dan Amerika Serikat, dengan jumlah penyandang diabetes sebanyak 19,47 juta penduduk dengan prevalensi sebesar 10,6 persen [4].

Faktor risiko penyakit diabetes melitus meliputi faktor risiko yang bisa diubah dan faktor risiko yang tidak dapat diubah. Faktor risiko yang bisa diubah diantaranya adalah kadar kolesterol dan trigliserida, sedangkan usia di atas 40 tahun merupakan salah satu faktor risiko yang tidak dapat diubah [5]. Untuk mengidentifikasi penyakit diabetes melitus, perlu dilakukan pengecekan kesehatan di laboratorium. Namun, tidak semua orang melakukan pengecekan kesehatan secara rutin.

Beragam penelitian dan pengembangan data *mining* banyak diaplikasikan pada bidang kesehatan [6]. Data *mining* dibagi menjadi estimasi, klasifikasi, pengklasteran, asosiasi dan prediksi [7]. Pohon keputusan atau *decision tree* merupakan salah satu metode yang digunakan untuk klasifikasi dan prediksi model. Metode pohon keputusan merupakan suatu metode yang dapat mengubah data yang sangat besar menjadi sebuah pohon keputusan yang mewakili aturan-aturan yang mudah dipahami [8]. Algoritma *Iterative Dichotomiser 3* (ID3) merupakan sebuah metode yang digunakan untuk membuat pohon keputusan [9]. Algoritma ID3 merupakan algoritma yang paling banyak digunakan dalam pohon keputusan. Metode ID3 adalah fungsi klasifikasi yang luas dan memiliki beberapa keunggulan, diantaranya aturan penilaian yang mudah dipahami dan model yang sederhana [10]. Algoritma ID3 digunakan untuk mengklasifikasikan faktor risiko penyakit diabetes melitus yang ada, untuk identifikasi awal seseorang terhadap penyakit diabetes melitus secara mandiri. Dengan demikian, seseorang yang diduga memiliki penyakit diabetes melitus dapat mendapatkan pengecekan lebih lanjut dan perawatan lebih awal.

Penelitian sebelumnya telah menggunakan pendekatan algoritma ID3 untuk menyelesaikan beberapa permasalahan, antara lain: penelitian untuk mengidentifikasi data rekam medis dengan studi kasus penyakit diabetes melitus di Balai Kesehatan Kementerian Perindustrian, Jakarta [1]. Penelitian lainnya adalah untuk mengevaluasi calon nasabah, kemudian memutuskan apakah seorang nasabah layak atau tidaknya mendapatkan kredit [11]. Penelitian berikutnya adalah untuk menentukan diagnosis kesehatan selama kehamilan, dengan fokus pada penentuan kesehatan pada ibu hamil [12]. Berdasarkan latar belakang yang ada, penelitian ini bertujuan untuk menerapkan metode algoritma ID3 dalam mengklasifikasikan faktor risiko penyakit diabetes melitus dan menentukan akurasi hasil klasifikasi algoritma ID3 dalam mengidentifikasi penyakit diabetes melitus.

## 2. METODOLOGI

### 2.1. Variabel Penelitian

Dalam penelitian ini, data yang digunakan adalah data sekunder yang diperoleh dari hasil pemeriksaan kesehatan pada karyawan [13]. Prevalensi diabetes melitus di Jawa Timur menempati peringkat kelima tertinggi di Indonesia. Data yang digunakan dalam penelitian ini merupakan data dari beberapa kota di Jawa Timur, diantaranya adalah Kota Banyuwangi, Jember, Kediri, Madiun, Malang, Sidoarjo dan Surabaya. Jumlah sampel yang akan digunakan adalah sebanyak 348 data. Atribut independen yang digunakan antara lain: jenis kelamin ( $X_1$ ), usia ( $X_2$ ), gula darah sewaktu (GDS) ( $X_3$ ), *High Density Lipoprotein* (HDL) ( $X_4$ ), *Low Density Lipoprotein* (LDL) ( $X_5$ ), dan trigliserida ( $X_6$ ), yang merupakan faktor risiko penyakit diabetes melitus. Atribut dependen yang digunakan adalah kadar Gula Darah Puasa (GDP) ( $Y$ ).

### 2.2. Tahapan Penelitian

Tahapan dalam penelitian setelah data diperoleh adalah dilakukan analisis statistik deskriptif pada data. Kemudian, dilakukan pembagian data *training* dan data *testing*. Data *training* digunakan untuk mendapatkan model klasifikasi berdasarkan faktor risiko penyakit diabetes melitus. Data *testing* digunakan untuk menghitung tingkat akurasi pada model klasifikasi. Setelah memperoleh data *training* dan data *testing*, dilakukan pemodelan klasifikasi dengan menerapkan Algoritma ID3 pada data *training*.

Langkah pertama dalam pemodelan klasifikasi menggunakan algoritma ID3 adalah menghitung nilai *entropy* total. Secara matematis *entropy* dirumuskan sebagai berikut [14]:

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (1)$$

dengan  $i = 1, 2, \dots, n$  dan  $p_i$  merupakan probabilitas yang diperoleh dari jumlah kasus dari kelas ke- $i$  dibagi total kasus.  $\log_2$  pada Persamaan (1) dapat digunakan untuk *output class* atau label kelasnya berupa dua kategori minimal (*binary classification*). Sedangkan untuk kondisi *multi-class*, maka  $\log$  yang digunakan adalah sesuai dengan jumlah label kelasnya [15]. Pada penelitian ini terdapat 3 label kelas, sehingga digunakan  $\log_3$ .

Setelah mendapatkan nilai *entropy* total, dilakukan perhitungan *entropy* setiap atribut dan perhitungan nilai *information gain*. *Information gain* digunakan untuk mengukur efektivitas suatu atribut dalam klasifikasi data. perhitungan *information gain* dituliskan sebagai berikut [16]:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2)$$

dimana  $Entropy(S_i)$  merupakan nilai *entropy* dari kelas ke- $i$  dalam atribut A. Atribut dengan nilai *information gain* tertinggi menjadi node akar.

Langkah selanjutnya adalah membuat *node* cabang yang didapat dari perhitungan *entropy* dan *information gain* selanjutnya. Tahap ini diulangi hingga membentuk pohon keputusan. Pohon keputusan adalah salah satu metode klasifikasi yang paling umum karena mudah diinterpretasi manusia, yang menggunakan struktur pohon atau struktur berhirarki [17]. Setelah pohon keputusan diperoleh, dilakukan evaluasi tingkat akurasi model klasifikasi menggunakan *confusion matrix* dan data *testing*. *Confusion matrix* digunakan untuk mengevaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah [18]. Suatu matriks dari prediksi yang akan dibandingkan dengan data aktual dari data atau dengan kata lain berisi informasi nilai aktual dan nilai prediksi dalam klasifikasi [19].

*Confusion matrix* dapat digunakan untuk mengukur berbagai *performance metric* untuk mengukur kinerja model yang telah dibuat. *Performance metric* yang sering digunakan meliputi *accuracy*, *precision*, dan *recall* atau *sensitivity*. *Precision* menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan model. *Recall* atau *sensitivity* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. *Accuracy* menggambarkan seberapa akurat klasifikasi suatu model [20]. Semakin tinggi akurasi, semakin baik model klasifikasi yang dihasilkan. Besarnya nilai akurasi terhadap klasifikasi kualitas model ditulis dalam Tabel 1.

Akurasi	Kualitas Model
90,01% - 100%	Sangat Baik
80,01% - 90%	Baik
70,01% - 80%	Cukup
60,01% - 70%	Buruk
≤60%	Gagal

Sumber: [21]

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Statistik Deskriptif

Analisis statistik deskriptif digunakan untuk memperoleh gambaran umum mengenai karakteristik suatu data. Adapun statistik deskriptif disajikan pada Tabel 2.

Atribut	Kategori	Jumlah	
Gula Darah Puasa (Y)	<110 mg/dl	Normal	286
	≥110 mg/dl - <126 mg/dl	Prediabetes	19
	≥ 126 mg/dl	Diabetes Melitus	43
Jenis Kelamin ( $X_1$ )	Laki-laki		201
	Perempuan		147
Usia ( $X_2$ )	≤40 tahun		165
	>40 tahun		183
Gula Darah Sewaktu ( $X_3$ )	<140 mg/dl	Normal	294
	≥140 mg/dl - <200 mg/dl	Prediabetes	20
	≥200 mg/dl	Diabetes	34

Atribut	Kategori	Jumlah	
HDL ( $X_4$ )	$\geq 35$ mg/dl	Rendah	47
	$< 35$ mg/dl	Normal	301
LDL ( $X_5$ )	$< 100$ mg/dl	Normal	78
	$\geq 100$ mg/dl	Tinggi	270
Trigliserida ( $X_6$ )	$< 250$ mg/dl	Normal	289
	$\geq 250$ mg/dl	Tinggi	59

### 3.2 Pembagian Data Training dan Data Testing

Pembagian data *training* dan data *testing* dilakukan dengan menggunakan bantuan *software* R Studio. Rasio pembagian data pada penelitian ini adalah 80% data *training* dengan total 278 data dan 20% data *testing*, dengan total 70 data [1].

### 3.3 Pembentukan Algoritma ID3

Perhitungan Algoritma ID3 dilakukan dengan bantuan *software* Microsoft Excel dan R Studio. Perhitungan untuk atribut GDS ( $X_3$ ) adalah sebagai berikut:

- a. Berdasarkan Persamaan (1) dihitung nilai *entropy* total.

$$Entropy(S) = \sum_{i=1}^3 -p_i \log_3 p_i$$

$$Entropy(total) = \left(-\frac{229}{278} \log_3 \left(\frac{229}{278}\right)\right) + \left(-\frac{15}{278} \log_3 \left(\frac{15}{278}\right)\right) + \left(-\frac{34}{278} \log_3 \left(\frac{34}{278}\right)\right)$$

$$= 0,5227$$

Nilai *entropy* total adalah sebesar 0,5227.

- b. Langkah selanjutnya adalah menghitung *entropy* setiap kelas.

Nilai *entropy* pada atribut GDS ( $X_3$ ) dengan kelas Normal adalah:

$$Entropy(X_3, N) = \sum_{i=1}^3 -p_i \log_3 p_i$$

$$= \left(-\frac{221}{233} \log_3 \left(\frac{221}{233}\right)\right) + \left(-\frac{8}{233} \log_3 \left(\frac{8}{233}\right)\right) + \left(-\frac{4}{233} \log_3 \left(\frac{4}{233}\right)\right)$$

$$= 0,2145$$

Nilai *entropy* untuk atribut  $X_3$ , kelas Normal adalah sebesar 0,2145.

Nilai *entropy* pada atribut GDS ( $X_3$ ) dengan kelas Prediabetes adalah:

$$Entropy(X_3, P) = \sum_{i=1}^3 -p_i \log_3 p_i$$

$$= \left(-\frac{8}{18} \log_3 \left(\frac{8}{18}\right)\right) + \left(-\frac{5}{18} \log_3 \left(\frac{5}{18}\right)\right) + \left(-\frac{5}{18} \log_3 \left(\frac{5}{18}\right)\right)$$

$$= 0,9758$$

Nilai *entropy* untuk atribut  $X_3$ , kelas Prediabetes adalah sebesar 0,9758.

Nilai *entropy* pada atribut GDS ( $X_3$ ) dengan kelas Diabetes adalah:

$$Entropy(X_3, D) = \sum_{i=1}^3 -p_i \log_3 p_i$$

$$= \left(-\frac{0}{27} \log_3 \left(\frac{0}{27}\right)\right) + \left(-\frac{2}{27} \log_3 \left(\frac{2}{27}\right)\right) + \left(-\frac{25}{27} \log_3 \left(\frac{25}{27}\right)\right)$$

$$= 0$$

Nilai *entropy* untuk atribut  $X_3$ , kelas Diabetes adalah sebesar 0.

- c. Berdasarkan Persamaan (2) dihitung nilai *information gain* GDS ( $X_3$ )

$$Gain(S, X_3) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i)$$

$$= 0,5227 - \left(\left(\frac{233}{278} \times 0,2145\right) + \left(\frac{18}{278} \times 0,9758\right) + \left(\frac{27}{278} \times 0\right)\right)$$

$$= 0,2797$$

Nilai *information gain* menunjukkan bahwa ukuran efektifitas atribut GDS ( $X_3$ ) dalam mengklasifikasi data adalah sebesar 0,2797.

**Tabel 3. Nilai Entropy dan Information Gain**

Atribut	Kategori	Entropy	Gain
Jenis Kelamin ( $X_1$ )	Laki-laki	0,3779	0,0123
	Perempuan	0,6110	
Usia ( $X_2$ )	≤40 tahun	0,2863	0,0305
	>40 tahun	0,6784	
GDS ( $X_3$ )	Normal	0,2145	0,2797
	Prediabetes	0,9758	
	Diabetes	0,0000	
HDL ( $X_4$ )	Rendah	0,8176	0,0159
	Normal	0,4575	
LDL ( $X_5$ )	Normal	0,3780	0,0043
	Tinggi	0,5596	
Trigliserida ( $X_6$ )	Normal	0,4531	0,0150
	Tinggi	0,7762	

Berdasarkan Tabel 3, atribut yang memiliki nilai *information gain* terbesar adalah atribut  $X_3$ , yaitu sebesar 0,2797 yang artinya atribut GDS memiliki pengaruh paling besar dalam klasifikasi penderita diabetes melitus. Atribut  $X_3$  memiliki nilai gain terbesar, sehingga  $X_3$  ditetapkan sebagai node akar.

Langkah selanjutnya adalah mengulangi perhitungan *entropy* dan *information gain* untuk membuat node cabang. Langkah ini diulangi hingga membentuk pohon keputusan. Perhitungan nilai *information gain* untuk node akar  $X_3$  kelas Prediabetes adalah sebagai berikut:

- a. Menghitung *entropy* total.

$$\begin{aligned} Entropy(X_3, P) &= \sum_{i=1}^3 -p_i \log_3 p_i \\ &= \left(-\frac{8}{18} \log_3 \left(\frac{8}{18}\right)\right) + \left(-\frac{5}{18} \log_3 \left(\frac{5}{18}\right)\right) + \left(-\frac{5}{18} \log_3 \left(\frac{5}{18}\right)\right) \\ &= 0,9758 \end{aligned}$$

Nilai *entropy* untuk *root node*  $X_3$  kelas Prediabetes adalah sebesar 0,9758.

- b. Kemudian dihitung nilai *entropy* setiap kelas.

Nilai *entropy* pada atribut Triglisierida dengan kelas ( $X_6$ ) Normal adalah:

$$\begin{aligned} Entropy(X_6, N) &= \sum_{i=1}^3 -p_i \log_3 p_i \\ &= \left(-\frac{3}{11} \log_3 \left(\frac{3}{11}\right)\right) + \left(-\frac{3}{11} \log_3 \left(\frac{3}{11}\right)\right) + \left(-\frac{5}{11} \log_3 \left(\frac{5}{11}\right)\right) \\ &= 0,9713 \end{aligned}$$

Nilai *entropy* untuk atribut  $X_3$ , kelas Normal adalah sebesar 0,9713.

Nilai *entropy* pada atribut Triglisierida dengan kelas ( $X_6$ ) Tinggi adalah:

$$\begin{aligned} Entropy(X_6, T) &= \sum_{i=1}^3 -p_i \log_3 p_i \\ &= \left(-\frac{5}{7} \log_3 \left(\frac{5}{7}\right)\right) + \left(-\frac{2}{7} \log_3 \left(\frac{2}{7}\right)\right) + \left(-\frac{0}{7} \log_3 \left(\frac{0}{7}\right)\right) \\ &= 0 \end{aligned}$$

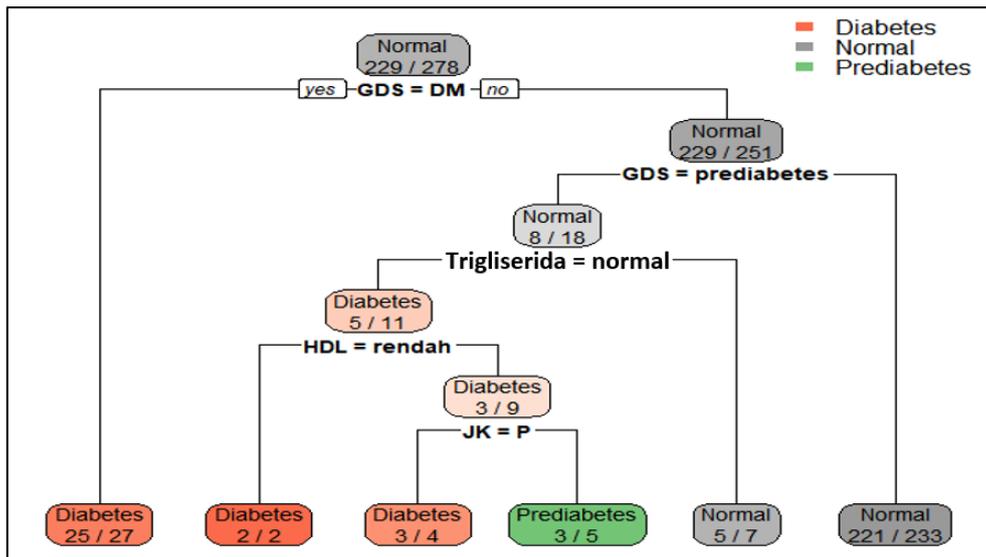
Nilai *entropy* untuk atribut  $X_3$ , kelas Tinggi adalah sebesar 0.

- c. Selanjutnya dihitung nilai *information gain* Triglisierida ( $X_6$ ).

$$\begin{aligned} Gain(X_3, X_6) &= Entropy(X_3, P) - \sum_{i=1}^2 \frac{|S_i|}{|S|} \times Entropy(S_i) \\ &= 0,9758 - \left(\left(\frac{11}{18} \times 0,9713\right) + \left(\frac{7}{18} \times 0\right)\right) \\ &= 0,3822 \end{aligned}$$

Nilai *information gain* Triglicerida ( $X_6$ ) adalah sebesar 0,3822.

Hasil model klasifikasi faktor risiko penyakit diabetes melitus menggunakan Algoritma ID3 disajikan pada Gambar 1.



**Gambar 1. Pohon Keputusan Algoritma ID3 Menggunakan Data Training**

Berdasarkan Gambar 1 diketahui bahwa:

1. Dari total 278 data *training*, terdapat 229 data normal.
2. Dari total 27 data gula darah sewaktu diabetes, terdapat 25 data diabetes.
3. Dari 251 data gula darah sewaktu selain diabetes, terdapat 229 data normal.
4. Dari 223 data gula darah sewaktu normal, terdapat 221 data normal.
5. Dari 18 data gula darah sewaktu prediabetes, terdapat 8 data normal.
6. Terdapat 5 data normal dari 7 data dengan trigeserida tinggi dan gula darah sewaktu prediabetes.
7. Terdapat 5 data diabetes dari 11 data dengan trigliserida normal dan gula darah sewaktu prediabetes.
8. Terdapat 2 data diabetes dari 2 data dengan HDL rendah dan gula darah sewaktu prediabetes.
9. Terdapat 3 data diabetes dari 9 data dengan HDL normal, trigliserida normal dan gula darah sewaktu prediabetes.
10. Terdapat 3 data diabetes dari 4 data dengan jenis kelamin perempuan, HDL normal, trigliserida normal dan gula darah sewaktu prediabetes.
11. Terdapat 3 data diabetes dari 5 data dengan jenis kelamin laki-laki, HDL normal, trigliserida normal dan gula darah sewaktu prediabetes.

Berdasarkan Gambar 1 dapat disimpulkan bahwa jika seseorang diklasifikasikan diabetes melitus, maka faktor risiko yang mungkin adalah:

1. Faktor risiko dengan GDS diabetes.
2. Faktor risiko dengan GDS prediabetes, trigliserida normal dan HDL rendah.
3. Faktor risiko dengan GDS prediabetes, trigliserida normal, HDL normal dan jenis kelamin perempuan.

Jika seseorang diklasifikasikan prediabetes, maka faktor risiko yang mungkin adalah faktor risiko dengan GDS prediabetes, trigliserida normal, HDL normal dan jenis kelamin laki-laki.

Jika seseorang diklasifikasikan normal, maka faktor risiko yang mungkin adalah:

1. Faktor risiko dengan GDS prediabetes dan trigliserida tinggi.
2. Faktor risiko dengan GDS normal.

### 3.4 Evaluasi Model Pohon Keputusan

Setelah model pohon keputusan diperoleh, selanjutnya dilakukan evaluasi model klasifikasi menggunakan *confusion matrix*. *Data testing* merupakan data yang digunakan untuk evaluasi model. *Data testing* adalah sebanyak 70 data. Adapun hasil evaluasi model klasifikasi algoritma ID3 disajikan pada Tabel 4 di bawah ini.

**Tabel 4. Confusion Matrix Algoritma ID3**

		<i>Predicted</i>		
		Normal	Prediabetes	Diabetes
<i>Actual</i>	Normal	56	1	0
	Prediabetes	4	0	0
	Diabetes	1	1	7

Pada Tabel 4. diperoleh hasil dari *confusion matrix* Algoritma ID3, sehingga dihitung tingkat ketepatan akurasi model adalah sebagai berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{\sum_{i=1}^m x_{ii}}{\text{jumlah data testing}} \times 100\% \\
 &= \frac{56 + 0 + 7}{70} \times 100\% \\
 &= 90\%
 \end{aligned}$$

Dengan menggunakan bantuan *confusion matrix*, hasil perhitungan yang diperoleh memiliki nilai akurasi adalah sebesar 90%, artinya model tersebut tergolong baik sehingga model dapat digunakan untuk mengklasifikasikan faktor risiko penyakit diabetes melitus.

### 3. KESIMPULAN

Dalam pengklasifikasian faktor risiko penyakit diabetes melitus diperoleh nilai akurasi sebesar 90% yang berarti model baik sehingga dapat digunakan untuk mengklasifikasi faktor risiko penyakit diabetes melitus. Dari hasil analisis diperoleh klasifikasi faktor risiko untuk diabetes adalah: faktor risiko dengan GDS diabetes, faktor risiko dengan GDS prediabetes, trigliserida normal dan HDL rendah, dan faktor risiko dengan GDS prediabetes, trigliserida normal, HDL normal dan jenis kelamin perempuan. Klasifikasi faktor risiko untuk Prediabetes adalah dengan GDS prediabetes, trigliserida normal, HDL normal dan jenis kelamin laki-laki. faktor risiko untuk Normal adalah: Faktor risiko dengan GDS prediabetes dan trigliserida tinggi, dan faktor risiko dengan GDS normal. Nilai akurasi faktor risiko GDS ( $X_3$ ) adalah yang paling berpengaruh dalam diagnosis penyakit diabetes melitus. sedangkan faktor risiko LDL ( $X_5$ ) tidak terlalu berpengaruh dalam diagnosis penyakit diabetes melitus.

### DAFTAR PUSTAKA

- [1] A. E. Tyasti, D. Ispriyanti and A. Hoyyi, "Algoritma Iterative Dichotomiser 3 (ID3) untuk Mengidentifikasi Data Rekam medis," *Jurnal Gaussian*, vol. 4, no. 2, pp. 237-246, 2015.
- [2] R. Bilous and R. Donnelly, *Buku Pegangan Diabetes*, Jakarta: Bumi Medika, 2015.
- [3] R. P. Febrinasari, T. A. Sholikhah, D. N. Pakha and S. E. Putra, *Buku Saku Diabetes Melitus Untuk Awam*, Surakarta: UNS Press, 2020.
- [4] International Diabetes Federation, "IDF Diabetes Atlas 2021," 2022. [Online]. Available: <https://diabetesatlas.org/atlas/tenth-edition/>. [Accessed 7 November 2022].
- [5] P2PTM Kemenkes RI, "Faktor risiko Penyakit Diabetes Melitus (DM)-Faktor Risiko yang Bisa Diubah," 2019. [Online]. Available: <http://p2ptm.kemendes.go.id/infographic-p2ptm/penyakit-diabetes-melitus/faktor-risiko-penyakit-diabetes-melitus-dm-faktor-risiko-yang-bisa-diubah>. [Accessed 6 April 2022].

- [6] M. A. Muslim, B. Prasetyo, E. L. H. Mawarni, A. J. Herowati, Mirqotussa'adah, S. H. Rukmana and A. Nurzahputra, Data Mining Algoritma C4.5 Disertai Contoh Kasus dan Penerapannya dengan Program Komputer, Semarang: FMIPA UNNES, 2019.
- [7] A. A. Fajrin and A. Maulana, "Penerapan Data Mining untuk Analisis Pola Pembelian Konsumen dengan Algoritma FP-GROWTH pada Data Transaksi Penjualan Spare Part Motor," *Kumpulan jurnal Ilmu Komputer (KLIK)*, vol. 5, no. 1, pp. 27-36, 2018.
- [8] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Edik Informatika*, vol. 2, no. 2, pp. 213-219, 2017.
- [9] V. Sihombing, "Klasifikasi Algoritma Iterative Dichotomizer (ID3) untuk Tingkat Kepuasan pada Sarana Laboratorium Komputer," *Jurnal Teknologi dan Ilmu Komputer Prima (JUTIKOMP)*, vol. 1, no. 2, pp. 180-187, 2018.
- [10] S. A. Hameed, "An Efficient of Classification the Gestational Diabetes Using ID3 Classifier," *Al-Nahrain Journal of Science*, vol. 25, no. 1, pp. 51-58, 2022.
- [11] H. D. Purnomo and T. Wahyono, "Sistem Pendukung Keputusan Pemberian Kredit menggunakan ID3," *Aiti : Jurnal Teknologi Informasi*, vol. 14, no. 1, pp. 13-30, 2017.
- [12] Hikmatulloh, A. Rahmawati, D. Wintana and D. A. Ambarsari, "Penerapan Algoritma Iterative Dichotomiser Three (ID3) dalam Mendiagnosa Kesehatan Kehamilan," *Kumpulan jurnal Ilmu Komputer (KLIK)*, vol. 6, no. 2, pp. 116-127, 2019.
- [13] S. A. Pujiati and N. Andayani, Penggunaan R dalam Farmasi, Surabaya: Berbagi.NET Publisher, 2015.
- [14] J. Suntoro, Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP, Jakarta: Elex Media Komputindo, 2019.
- [15] A. M. Khalimi, "Menghitung Entropy Kondisi Multi-Class," 2022. [Online]. Available: [http://www.pengalaman-edukasi.com/2022/06/cara-hitung-entropy-lebih-dari-2-kelas.html?m=1#google\\_vignette](http://www.pengalaman-edukasi.com/2022/06/cara-hitung-entropy-lebih-dari-2-kelas.html?m=1#google_vignette). [Accessed 27 Mei 2023].
- [16] D. Nofriansyah and G. W. Nurcahyo, Algoritma Data Mining dan Pengujiannya, Yogyakarta: Deepublish, 2019.
- [17] I. Choina, R. Aulia and A. Zakir, "Penerapan Algoritma ID3 untuk Menyeleksi Pegawai Kontrak di Kantor Pengadilan Kota Langsa," *CESS (Journal of Computer Engineering, System and Science)*, vol. 5, no. 1, pp. 47-52, 2020.
- [18] D. K. Widiyati, M. Wati and H. S. Pakpahan, "Penerapan Algoritma ID3 Decision Tree pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara," *J. Rekayasa teknol. Inf*, vol. 2, no. 2, pp. 125-134, 2018.
- [19] A. Budiman, A. Setyanto and F. W. Wibowo, "Prediksi Tingkat Kelulussan Mahasiswa Menggunakan Algoritma C4.5," *TEKNOMATIKA*, vol. 11, no. 2, pp. 83-93, 2019.
- [20] K. S. Nugroho, "Confusion Matrix untuk Evaluasi Model pada Supervised Learning," 13 November 2019. [Online]. Available: <https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f>. [Accessed 31 Juli 2023].
- [21] M. S. Maulana, R. Sabarudin and W. Nugraha, "Prediksi Ketepatan Lulusan Mahasiswa dengan Komparasi Algoritma Klasifikasi," *JUSTIN: Jurnal Sistem dan Teknologi Informasi*, vol. 7, no. 3, pp. 202-206, 2019.