

## Penerapan Algoritma Klasifikasi *Classification And Regression Trees* (CART) untuk Diagnosis Penyakit Diabetes Retinopathy

Pungkas Subarkah

Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto  
E-Mail : [subarkah@amikompurwokerto.ac.id](mailto:subarkah@amikompurwokerto.ac.id)

### ABSTRAK

Penyakit *diabetic retinopathy* atau DR adalah salah satu komplikasi penyakit diabetes yang bisa menyebabkan kematian bagi penderitanya. Komplikasi tersebut berupa kerusakan pada bagian retina mata. Tingginya kadar glukosa dalam darah adalah penyebab pembuluh darah kapiler kecil menjadi pecah dan dapat menyebabkan kebutaan. Retinopati diabetes diawali dengan melemah atau hancurnya kapiler kecil di retina, darah bocor yang kemudian menyebabkan penebalan jaringan, pembengkakan, dan pendarahan yang luas. Penelitian ini bertujuan untuk menganalisis diagnosis penyakit diabetes *retinopathy*. Algoritma *Classification And Regression Trees* (CART) merupakan salah satu algoritma klasifikasi dengan menggunakan *dataset* diambil dari UCI Repository Learning diperoleh dari Universitas Debrecen, Hongaria, yang terdiri dari data pasien terindikasi penyakit diabetes *retinopathy* dan normal penyakit diabetes *retinopathy*. Metode yang digunakan dalam penelitian ini yaitu identifikasi masalah, pengumpulan data, tahap *pre-processing*, metode klasifikasi, validasi dan evaluasi serta penarikan kesimpulan. Adapun metode validasi dan evaluasi yang digunakan yaitu *10-cross validation* dan *confusion matrix*. Dari hasil perhitungan yang telah dilakukan, maka didapatkan hasil akurasi pada algoritma CART sebesar 63.4231%, dengan nilai *precision* 0.64%, nilai *Recall* 0.634%, dan nilai *F-Measure* 0,634%.

**Kata Kunci:** *Diabetes Retinopathy, Klasifikasi, CART.*

### ABSTRACT

*Diabetic Retinopathy or DR is a complication of diabetes that can cause death for sufferers. These complications are damage to the retina of the eye. High blood sugar levels cause small capillaries to rupture and can cause blindness. Diabetic retinopathy begins with a decrease or destruction of small capillaries in the retina, blood leak which then causes tissue thickening, swelling, and extensive bleeding. This study attempts to analyze the diagnosis of diabetic retinopathy. Classification Algorithms And Regression Trees (CART) is one algorithm that uses a dataset taken from the UCI Learning Repository obtained from Debrecen University, which consists of patient data indicated by diabetic retinopathy and normal retinopathy diabetes. The method used in this research is problem collection, data collection, pre-processing, collection methods, validation and evaluation as well as gathering conclusions. The validation and evaluation methods used are 10-cross validation and confusion matrix. From the results of calculations that have been done, the accuracy of the CART algorithm is obtained at 63.421%, with a precision value of 0.64%, a Recall value of 0.634%, and a value of F- Measures 0.634%.*

**Keywords:** *Diabetes Retinopathy, Classification, CART.*

Author Korespondensi (Pungkas Subarkah)  
Email : [subarkah@amikompurwokerto.ac.id](mailto:subarkah@amikompurwokerto.ac.id)

### I. PENDAHULUAN

Penyakit *diabetic retinopathy* atau DR adalah salah satu komplikasi penyakit diabetes yang bisa menyebabkan kematian bagi penderitanya. Komplikasi tersebut berupa kerusakan pada bagian retina mata. Tingginya kadar glukosa dalam darah adalah penyebab pembuluh darah kapiler kecil menjadi pecah dan dapat menyebabkan kebutaan. Retinopati diabetes diawali dengan melemah atau hancurnya kapiler kecil di retina, darah bocor yang kemudian

menyebabkan penebalan jaringan, pembengkakan, dan pendarahan yang luas [1]. Penyakit ini pada umumnya akan berdampak secara langsung pada terganggunya penglihatan penderita dan apabila terlambat ditangani maka akan menyebabkan penderita mengalami kebutaan secara permanen. DR hingga saat ini masih sulit disembuhkan karena mayoritas penderita melakukan pemeriksaan di saat kondisi penyakit telah memasuki tahap berbahaya. Hal ini dikarenakan sifat dari penyakit DR ini yang

tidak menunjukkan gejala yang terlihat bila masih pada tahap awal [2].

Penyakit ini merupakan salah satu komplikasi mikrovaskular diabetes melitus dengan angka prevalensi yang cukup tinggi. Berdasarkan data penelitian yang dilakukan di beberapa daerah di Indonesia, diperkirakan prevalensi retinopati diabetik sebesar 42,6%. Setidaknya akan ditemukan 24.600 orang dengan retinopati diabetik dan sekitar 10% dari jumlah tersebut mengalami kebutaan. Jumlah ini diperkirakan akan meningkat pada tahun 2030 dengan estimasi 98.400 orang menderita diabetes dan sekitar 11.000 orang diantaranya mengalami kebutaan. Mengingat sebagian besar kebutaan akibat diabetik retinopati merupakan kondisi yang permanen dan tidak dapat diobati, maka pencegahan dan pengobatan menjadi hal yang penting [3]. Efek visual dari kebocoran digambarkan dengan adanya perubahan pada retina seperti perubahan venous, *cotton wool spots*, mikroaneurisma, *hard exudate* dan pendarahan [4].

Menurut International Diabetes Mellitus Federation (IDF) Atlas 2017: epidemi Diabetes Melitus di Indonesia masih menunjukkan kecenderungan meningkat. Indonesia adalah negara peringkat ke-enam di dunia setelah Tiongkok, India, Amerika Serikat, Brazil dan Meksiko dengan jumlah penderita Diabetes Melitus usia 20–79 tahun sekitar 10,3 juta orang. Riset Kesehatan Dasar (Riskesdas) juga memperlihatkan peningkatan angka prevalensi Diabetes Melitus yang cukup signifikan, yaitu dari 6,9 % di tahun 2013 menjadi 8,5 % di tahun 2018; sehingga estimasi jumlah penderita Diabetes Melitus di Indonesia mencapai 22 juta orang [5]. Diagnosis dini pada penyakit diabetes retinopathy dapat dilakukan dengan metode *data mining*. Dimana proses pada metode *data mining* bertujuan untuk menguraikan temuan baru di dalam *dataset* dan menjelaskan suatu proses yang menggunakan teknik statistik, matematis, *artificial intelligence*, dan *machine learning* untuk melakukan ekstrak dan identifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* tersebut.

Salah satu fungsi utama *data mining* adalah klasifikasi. Klasifikasi banyak digunakan untuk menentukan keputusan sesuai pengetahuan baru yang didapat dari pengolahan data lampau menggunakan perhitungan suatu algoritma. Teknik klasifikasi dapat diterapkan dalam semua bidang misalnya dalam bidang kesehatan [6],

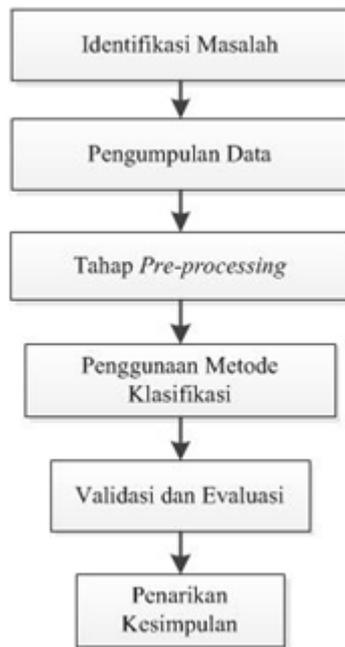
serta banyak digunakan dalam bidang lain. Dalam *dataset* klasifikasi terdapat satu atribut tujuan atau dapat pula disebut dengan atribut label. Atribut inilah yang akan dicari dari data baru dengan dasar atribut lain pada data lampau. Banyaknya atribut dapat mempengaruhi performansi suatu algoritma [7].

Algoritma *Classification And Regression Trees* (CART) adalah satu metode atau algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan [8]. Metode ini dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen dan Charles J. Stone sekitar tahun 1980. CART merupakan metodologi statistik non-parametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk variabel respon kategorik maupun kontinu. Metode ini merupakan metode yang biasa diterapkan suatu himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan biner [9]. Kelebihan dari algoritma CART adalah metode yang bersifat non-parametrik atau cocok digunakan untuk data yang berjenis *numeric*. CART tidak memerlukan variabel yang akan dipilih terlebih dahulu dan menghasilkan invariant untuk transformasi variabel yang independen [10]. Pada klasifikasi algoritma CART, sebuah *record* akan diklasifikasikan ke dalam salah satu dari sekian klasifikasi yang tersedia pada variabel tujuan berdasarkan nilai-nilai variabel prediktornya [11].

Penelitian ini bertujuan untuk menganalisis penyakit diabetes *retinopathy* menggunakan *dataset* yang diperoleh dari UCI *repository machine learning* dan dibuat oleh Universitas Debrecen, Hongaria, untuk melakukan klasifikasi diagnosis penyakit diabetes *retinopathy*. Peneliti menggunakan algoritma klasifikasi yaitu CART.

## II. METODELOGI

Tahapan penelitian dilakukan untuk mempermudah dalam melakukan penelitian. Desain penelitian analisis algoritma klasifikasi CART untuk diagnosis penyakit diabetes *retinopathy* dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

### 2.1. Identifikasi Masalah

Pada tahapan ini dilakukan untuk mengetahui permasalahan serta metode yang sesuai sehingga dapat ditentukan poin-poin untuk target diagnosis penyakit diabetes *retinopathy*.

### 2.2. Pengumpulan Data

Dalam tahapan ini pengumpulan data menggunakan data sekunder yang digunakan, diambil dari *database UCI Repository*. *Dataset* terdiri dari 1151 *record*, yang memiliki 19 *attribute* (18 *attribute* dan 1 *attribute* target). *Attribute* target terdiri dua *output* yaitu normal diabetes *retinopathy* dan terindikasi diabetes *retinopathy* [12]. Kedua *output* pada *class attribute* ini akan digunakan sebagai *output* data penyakit diabetes *retinopathy*. Penjelasan mengenai atribut diabetes *retinopathy*, dapat dilihat pada tabel 1 dan tabel 2.

Tabel 1. Atribut Pada *Dataset Diabetes Retinopathy*

Nama	Informasi
<i>Attribute 0</i>	Hasil biner pada kualitas penilaian
<i>Attribute 1</i>	Hasil biner pada pra-skrining
<i>Attribute 2</i>	Hasil deteksi MA
<i>Attribute 3</i>	Hasil deteksi MA
<i>Attribute 4</i>	Hasil deteksi MA
<i>Attribute 5</i>	Hasil deteksi MA
<i>Attribute 6</i>	Hasil deteksi MA
<i>Attribute 7</i>	Hasil deteksi MA
<i>Attribute 8</i>	Hasil normalisasi pada eksudat
<i>Attribute 9</i>	Hasil normalisasi pada eksudat
<i>Attribute 10</i>	Hasil normalisasi pada eksudat

<i>Attribute 11</i>	Hasil normalisasi pada eksudat
<i>Attribute 12</i>	Hasil normalisasi pada eksudat
<i>Attribute 13</i>	Hasil normalisasi pada eksudat
<i>Attribute 14</i>	Hasil normalisasi pada eksudat
<i>Attribute 15</i>	Hasil normalisasi pada eksudat
<i>Attribute 16</i>	Jarak Euclidean pada pusat manula dan pusat cakram optik
<i>Attribute 17</i>	Diameter cakram optik
<i>Attribute 18</i>	Hasil biner pada klasifikasi berbasis AM/FM
<i>Target Attribute 19</i>	Label kelas. 1 = mengandung tanda-tanda DR (label akumulatif untuk kelas Messidor 1, 2, 3), kelas. 0 = tidak ada tanda-tanda DR.

Tabel 2. Jumlah *Dataset* Berdasarkan *Class*

No	Jenis Klasifikasi	Jumlah <i>Record Dataset</i>
1	<i>Contains Signs</i>	611
2	<i>No Signs</i>	540
<b>Jumlah</b>		<b>1151</b>

### 2.3. Tahap Pre-Processing

Pada tahapan ini dilakukan suatu proses seleksi data dengan tujuan untuk mendapatkan data yang bersih dan siap digunakan sebagai bahan penelitian. Tahap *pre-processing* data meliputi identifikasi dan pemilihan atribut (*attribute identification and selection*), penanganan nilai atribut yang hilang atau atribut yang tidak lengkap (*handling missing values*) dan proses diskretisasi nilai.

### 2.4. Penggunaan metode klasifikasi

Metode yang peneliti gunakan yaitu algoritma CART, dari hasil *confusionmatrix* dapat dihitung nilai *precision*, *recall* dan *F-Measure* [13]. Rincian *confusion matrix* dapat dilihat pada tabel 3.

Tabel 3. *Confusion Matrix*

<i>Correct classification</i>	<i>Classification as</i>	
	+	-
+	<i>True positives</i>	<i>False positives</i>
-	<i>False positives</i>	<i>True negatives</i>

Adapun rumus perhitungan dari sebuah *precision*, *recall* dan *F-Measure* dari sebuah *accuracy* [13], dapat dilihat pada persamaan dibawah ini.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

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

$$F\text{-Measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

### 2.5.1. Validasi dan Evaluasi

Dalam tahap ini dilakukan validasi dan evaluasi serta pengukuran keakuratan hasil yang dicapai oleh model menggunakan teknik yang terdapat dalam aplikasi Weka yaitu *confusion matrix* dan *cross-validation*.

## 2.5. Penarikan Kesimpulan

Pada tahapan ini yaitu menyimpulkan hasil yang diperoleh dari penelitian dengan menggunakan Algoritma CART yang memberikan hasil akurasi terbaik untuk mendiagnosis penyakit diabetes mellitus berdasarkan nilai *precision*, *recall*, *F-Measure* dari tiap masing-masing algoritma dengan tingkat diagnosis sebagai berikut:

- Excellent classification* = 0.90 – 1.00
- Good classification* = 0.80 – 0.90
- Fair classification* = 0.70 – 0.80
- Poor classification* = 0.60 – 0.70
- Failure* = 0.50 – 0.60

## III. HASIL DAN PEMBAHASAN

### 3.1. Identifikasi Masalah

Dalam tahapan ini penulis melakukan beberapa studi pendahuluan dengan mempelajari literatur-literatur yang berkaitan dengan penelitian penulis yaitu untuk mendiagnosis penyakit diabetes *retinopathy* serta pemilihan algoritma yang sesuai untuk penelitian penulis ini. Dari hasil tersebut penulis melakukan penelitian dengan menggunakan algoritma CART untuk menganalisa *dataset*.

### 3.2. Pengumpulan Data

Dalam penelitian penulis ini data yang akan digunakan dalam penelitian yaitu dengan mengambil dari repository diambil dari *database UCI repository*. *Dataset* terdiri dari 1151 *record*, yang memiliki 19 *attribute* (18 *attribute* dan 1 *attribute* target). *Attribute* target terdiri dua *output* yaitu normal diabetes *retinopathy* dan terindikasi diabetes *retinopathy* [12]. kedua *output* pada *class attribute* ini akan digunakan sebagai *output* data penyakit diabetes *retinopathy*. Data ini harus diolah terlebih dahulu yaitu melalui tahap *pre-processing*, dimana tahapan ini bertujuan untuk menyesuaikan atribut

– atribut yang akan digunakan dalam mengolah *dataset* tersebut karena setiap atribut pada *dataset* ini akan memberikan hasil yang berbeda – beda pada hasil akurasi dalam mendiagnosa penyakit diabetes *retinopathy* dan hal ini berkaitan ada atau tidaknya kelengkapan nilai pada setiap atribut. Atribut-atribut yang dimaksud adalah :

- Attribute 0*  
Hasil biner dari penilaian yang berkualitas
  - Attribute 1*  
Hasil biner dari *pre-screening*
  - Attribute 2-7*  
Hasil dari deteksi MA.
  - Attribute 8-15*  
Eksudat yang diwakili oleh seperangkat, yang menghasilkan jumlah piksel yang membentuk lesi pada bola mata.
  - Attribute 16*  
Jarak euclidean dari pusat makula dan pusat cakram optik untuk memberikan informasi penting mengenai kondisi pasien.
  - Attribute 17*  
Diameter cakram optik.
  - Attribute 18*  
Hasil biner dari klasifikasi berbasis AM / FM
  - Class\_Attribute*  
Label kelas. 1 = mengandung tanda-tanda DR (label akumulatif untuk kelas Messidor 1, 2, 3), 0 = tidak ada tanda-tanda DR.
- Atribut diatas yang digunakan peneliti untuk bahan penelitian dan terlebih dahulu disesuaikan dengan tipe data yang dapat dibaca oleh aplikasi Weka.

### 3.3. Tahap Pre-Processing

Dalam tahapan ini, dilakukan identifikasi dan penyesuaian atribut, serta penyeleksian dari *dataset* diabetes *retinopathy* agar data yang diperoleh ialah data yang benar-benar siap untuk digunakan. Tabel 4, yang merupakan *dataset* diabetes *retinopathy* yang telah dilakukan penyesuaian atribut untuk aplikasi Weka.

**Tabel 4.** Data *Pre-Processing*

Data Asli	Data Hasil <i>Pre-Processing</i>	Keterangan
1	1	Hasil biner dari penilaian yang berkualitas
1	1	Hasil biner dari <i>pre-screening</i>
22	22	Hasil dari deteksi MA
19	19	
19	19	
15	15	
12	12	
9	9	
70.192324	70.192324	Eksudat yang diwakili oleh seperangkat, yang menghasilkan jumlah piksel yang membentuk lesi pada bola mata.
4		
24.971335	24.971335	
4.403699	4.403699	
0.713671	0.713671	
0.394666	0.394666	
0.207558	0.207558	
0.089976	0.089976	
0.039876	0.039876	Jarak euclidean dari pusat makula dan pusat cakram optik untuk memberikan informasi penting mengenai kondisi pasien.
0.518668	0.518668	
0.098155	0.098155	Hasil biner dari klasifikasi berbasis AM / FM
1	1	Label kelas. 1 = mengandung tanda-tanda DR (label akumulatif untuk kelas Messidor 1, 2, 3), 0 = tidak ada tanda-tanda DR
<i>Contains Signs</i>	0	Class, <i>Contains Signs</i>
<i>Signs</i>	1	Class, <i>Signs</i>

Dari hasil analisis tabel 4, tentang hasil data *pre-processing dataset* diabetes *retinopathy* yang telah dilakukan, bahwa semua atribut nilainya lengkap atau tidak *missing value*. Berikut ini

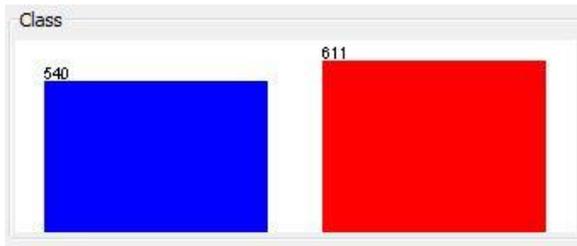
merupakan tabel sebelum dan sesudah melalui proses penanganan *missing value*, dapat dilihat pada tabel 5.

**Tabel 5.** Perbandingan Tabel *Dataset* Sebelum dan Sesudah Melalui Proses Penanganan *Missing Value*

Atribut <i>Dataset</i> asli	Atribut <i>dataset</i> setelah melalui proses penanganan <i>missing values</i>
<i>Attribute 0</i>	<i>Attribute 0</i>
<i>Attribute 1</i>	<i>Attribute 1</i>
<i>Attribute 2</i>	<i>Attribute 2</i>
<i>Attribute 3</i>	<i>Attribute 3</i>
<i>Attribute 4</i>	<i>Attribute 4</i>
<i>Attribute 5</i>	<i>Attribute 5</i>
<i>Attribute 6</i>	<i>Attribute 6</i>
<i>Attribute 7</i>	<i>Attribute 7</i>
<i>Attribute 8</i>	<i>Attribute 8</i>
<i>Attribute 9</i>	<i>Attribute 9</i>
<i>Attribute 10</i>	<i>Attribute 10</i>
<i>Attribute 11</i>	<i>Attribute 11</i>
<i>Attribute 12</i>	<i>Attribute 12</i>
<i>Attribute 13</i>	<i>Attribute 13</i>
<i>Attribute 14</i>	<i>Attribute 14</i>
<i>Attribute 15</i>	<i>Attribute 15</i>
<i>Attribute 16</i>	<i>Attribute 16</i>
<i>Attribute 17</i>	<i>Attribute 17</i>
<i>Attribute 18</i>	<i>Attribute 18</i>
<i>Target_Attribute 19</i>	<i>Target_Attribute 19</i>

### 3.3. Penggunaan Metode Klasifikasi

Setelah melalui tahapan *pre-processing*, selanjutnya *dataset* mulai diolah dengan menggunakan aplikasi Weka. Tahapan ini bertujuan untuk menghasilkan *confusion matrix* berdasarkan metode evaluasi *10-fold cross validation*, dimana *dataset* tersebut dibagi menjadi 10 *subsets* (9 *subsets* sebagai *trainingsets* dan 1 *subsets* sebagai *testingset*) dengan jumlah 10 kali iterasi. Adapun *classifier* yang digunakan untuk uji coba pada *dataset* adalah *SimpleCart* (algoritma *Classification And Regression Tress*). Gambar 3 menunjukkan visualisasi data secara keseluruhan yang dijabarkan sesuai dengan atribut setelah melewati tahapan *pre-processing*. Selanjutnya data-data tersebut diklasifikasi menggunakan algoritma CART (*SimpleCart* dalam penggunaan pada aplikasi Weka) yang menghasilkan pola klasifikasi. Berikut gambar 3 tentang visualisasi data keseluruhan, sebagai berikut :



Gambar 3. Visualisasi data keseluruhan

Keterangan dari gambar 3, diatas sebagai berikut:

- Gambar grafik dengan warna biru terindikasi diabetes *retinopathy*.
- Gambar grafik dengan warna merah normal atau tidak terindikasi diabetes *retinopathy*.

Hasil perhitungan dengan menggunakan aplikasi Weka didapatkan *rule* pohon keputusan dari *dataset* diabetes *retinopathy* yang dapat dilihat pada gambar 4. Pada pohon keputusan tersebut, dapat diketahui bahwa atribut 14 merupakan akar dari *rule* pohon keputusan. Jika pada atribut 14 nilainya kurang dari 0.0203705, pada atribut 2 kurang dari 55.5, selanjutnya pada atribut 8 jika nilainya kurang dari 127.805703 maka *class* atributnya tidak terindikasi diabetes *retinopathy*.

CART Decision Tree

```

14 < 0.0203705
| 2 < 55.5
| | 8 < 127.805703: 0(402.0/243.0)
| | 8 >= 127.805703
| | | 2 < 11.0: 0(20.0/9.0)
| | | 2 >= 11.0: 1(55.0/25.0)
| 2 >= 55.5: 1(137.0/62.0)
14 >= 0.0203705: 1(167.0/31.0)
    
```

Gambar 4. *Rule* pohon keputusan dari *dataset* Diabetes *Retinopathy*

Kemudian jika nilai pada atribut 8 nilainya lebih dari 127.805703, selanjutnya pada atribut 2 nilainya kurang dari 11.0 maka *class* atributnya yaitu tidak terindikasi diabetes *retinopathy*. Sedangkan jika pada atribut 2 nilainya lebih dari sama dengan 11.0 maka *class* atributnya adalah terindikasi diabetes *retinopathy*. Selanjutnya pada atribut 2 nilainya 55.5 maka *class* atributnya adalah terindikasi diabetes *retinopathy*. Kemudian jika pada atribut 14 nilainya lebih dari sama dengan 0,203705 maka *class* atributnya terindikasi diabetes *retinopathy*. Pada algoritma pohon keputusan yang dihasilkan menggunakan algoritma CART yang terlihat pada gambar 5,

menghasilkan nilai akurasi sebesar 63, 4231%. Nilai akurasi tersebut didapatkan dari hasil perhitungan dari *precision*, *recall* dan *F-Measure*. Proses perhitungannya dapat dilihat pada tabel 6.

Tabel 6. *Confusion* Kelas “0”

368 ( <i>True positive</i> )	172 ( <i>False Negative</i> )
249 ( <i>False Positive</i> )	362 ( <i>True Negative</i> )

Pada tabel 6. *Confusion* Kelas “0”, merupakan penjabaran dari *confusion matrix* yang dapat dilihat pada tabel 6. *Confusion* Kelas “0”.

Kelas “0” :

$$Precision = \frac{TP}{TP+FP} = \frac{368}{368+249} = 0,596$$

$$Recall = \frac{TP}{TP+FN} = \frac{368}{368+172} = 0,681$$

$$F-Measure = \frac{2x\ precision \times recall}{precision + recall} = \frac{2x0,596 \times 0,681}{0,596 + 0,681} = \frac{2x0,405876}{1,277} = \frac{0,811752}{1,277} = 0,636$$

Tabel 7. *Confusion* Kelas “1”

362 ( <i>True Positive</i> )	249 ( <i>False Negative</i> )
172 ( <i>False Positive</i> )	368 ( <i>True Negative</i> )

Kelas “1” :

$$Precision = \frac{TP}{TP+FP} = \frac{362}{362+272} = 0,678$$

$$Recall = \frac{TP}{TP+FN} = \frac{362}{362+172} = 0,592$$

$$F-Measure = \frac{2x\ precision \times recall}{precision + recall} = \frac{2x0,678 \times 0,592}{0,678 + 0,592} = \frac{2x0,401376}{1,27} = \frac{0,802752}{1,27} = 0,632$$

Dari hasil perhitungan diatas antara lain *precision*, *recall* dan *F-Measure* yang dihasilkan dari kelas 0 dan kelas 1 dapat dihitung nilai rata-rata dari kelas-kelas yang ada (*Weighted Avg*) dengan terlebih dahulu menjumlahkan nilai A = (368 + 172) = 540 dan nilai B = (249 + 362) = 611. Rumus untuk menghitung *Weighted Avg*, sebagai berikut :

$$Weighted\ Avg\ (precision) = \frac{0,596 \times A + 0,688 \times B}{0,596 \times 540 + 0,688 \times 611} = \frac{321,84 + 420,368}{1151} = \frac{742,208}{1151} = 0,64$$

$$\text{Weighted Avg (recall)} = \frac{0,681xA + 0,592 \times B}{0,681 \times 540 + 0,592 \times 611} = \frac{367,74 + 361,712}{1151} = \frac{729,452}{1151} = 0,634$$

$$\text{Weighted Avg (F-Measure)} = \frac{0,636xA + 0,632 \times B}{0,636 \times 540 + 0,632 \times 611} = \frac{343,44 + 386,152}{1151} = \frac{729,592}{1151} = 0,634$$

Hasil diatas merupakan perhitungan nilai akurasi berdasarkan *confusion matrix* yang disajikan pada tabel 8. Nilai akurasi berdasarkan *confusion matrix*, sebagai berikut:

**Tabel 8.** Nilai Akurasi Berdasarkan *Confusion Matrix*

Class	Precision	Recall	F-Measure
0	0,596	0,681	0,636
1	0,678	0,592	0,632
Weighted Avg	0,64	0,634	0,634

Pengklasifikasian untuk menguji *dataset* diabetes *retinopathy* dengan algoritma CART menggunakan aplikasi Weka 3.6 membutuhkan waktu 0,78 *second*.

### 3.4. Validasi dan Evaluasi

Untuk mengukur tingkat akurasi dari algoritma klasifikasi yang digunakan yaitu menggunakan *confusion matrix* yang disajikan pada tabel 9. Tabel tersebut merupakan tabel yang dihasilkan oleh *confusion matrix* dari pengujian *dataset* menggunakan algoritma CART dengan metode *10-fold cross validation*.

**Tabel 9.** *Confusion Matrix*

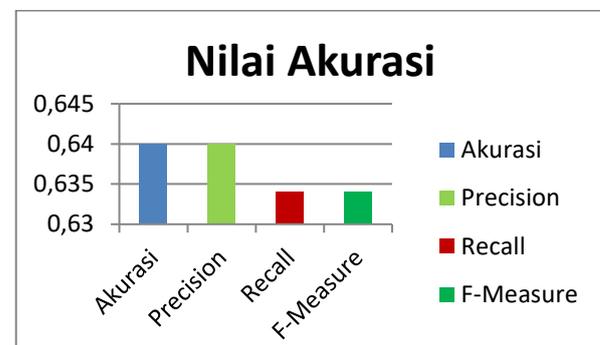
	Normal Diabetes Retinopathy	Terindikasi Diabetes Retinopathy
Normal Diabetes Retinopathy	368	172
Terindikasi Diabetes Retinopathy	249	362
1151	617	534

Dari tabel 9, terlihat bahwa jumlah data hasil bentukan *rule* yang terkena penyakit diabetes *retinopathy* yang sama dengan data *testing* yang

juga terkena diabetes *retinopathy* sebanyak 368. Kemudian, jumlah data hasil bentukan *rule* yang tidak terkena diabetes *retinopathy* dengan data *testing* yang terkena diabetes *retinopathy* sebanyak 172. Selanjutnya, jumlah data hasil bentukan *rule* yang terkena diabetes *retinopathy* dan data *testing* yang tidak terkena diabetes *retinopathy* sebanyak 249. Terakhir, jumlah data hasil bentukan *rule* yang tidak terkena diabetes *retinopathy* yang sama dengan data *testing* yang juga tidak terkena diabetes *retinopathy* sebanyak 362.

### 3.5. Penarikan Kesimpulan

Dari hasil perhitungan yang telah dilakukan pada algoritma CART didapatkan nilai akurasi yaitu 63.4231%, dengan nilai *precision* 0.64%, nilai *Recall* 0.634%, dan nilai *F-Measure* 0,634%. Grafik hasil akurasi dari perhitungan pada algoritma CART disajikan pada gambar 6.



## IV. SIMPULAN DAN SARAN

### 4.1. Kesimpulan

Dari hasil penelitian ini, peneliti mengambil kesimpulan bahwa hasil analisis klasifikasi algoritma *Classification And Regression Trees* (CART) pada *dataset* diabetes *retinopathy*, berdasarkan dari nilai *confusion matrix* didapatkan hasil akurasi sebesar 63,4321% dengan nilai *precision* 0.64%, nilai *Recall* 0.634%, dan nilai *F-Measure* 0,634%.

### 4.2. Saran

1. Menambahkan algoritma klasifikasi untuk mengetahui bagaimana hasil akurasi yang lebih baik pada diagnosis penyakit diabetes *retinopathy*.
2. Melakukan seleksi fitur *data mining* pada algoritma dalam klasifikasi penyakit diabetes *retinopathy*.

3. Peneliti berharap untuk penelitian selanjutnya, hasil analisis ini dapat diimplementasikan ke sebuah sistem.

#### DAFTAR PUSTAKA

- [1] B. S. A. Ayu, Ratna Gitasari. Hidayat, "Klasifikasi Penyakit Diabetes Retinopati Berdasarkan Citra Digital dengan menggunakan Metode Wavelet dan Support Vector Machine.," Universitas Telkom, 2013.
- [2] I. M. Dewi, "Mengenal Lebih Jauh Retinopathy Diabetik," 2018. [Online]. Available: <https://www.columbiaasia.com/indonesia/health-articles/mengenal-lebih-jauhretinopathy-diabetik>.
- [3] "Diabetik Retina : Pergeseran Paradigma Kebutaan pada Era Milenial." [Online]. Available: <http://www.yankes.kemkes.go.id/read-retinopati-diabetik-pergeseran-paradigma-kebutaan-pada-era-milenial-5984.html>. Diakses Hari Selasa, 03 Maret 2020. Pukul 14.57 WIB.
- [4] S. Faust, O., Acharya, R. U., Ng, E. Y. K., Ng, K. H. and J.S, "Algorithms for the Automated Detection of Diabetic Retinopathy Using Digital Fundus Image : A Review," *JaMed Syst*, 2010.
- [5] IDF., "Diabetes Melitus Atlas," *Sixth Edition*, 2013. [Online]. Available: <http://www.idf.org/Diabetes>
- [6] Melitus Atlas/download-book. Di akses Pada Hari Selasa, 03 Maret 2020. Pukul 15.04 WIB.
- [7] and D. . S. Christobel, Angeline, "An Empirical Comparison of Data Mining Classification Methods," 201AD, pp. 24– 28.
- [8] E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset, 2012.
- [9] W. B. Komalasari, "Metode Pohon Regresi Untuk Eksplorasi Data Dengan Peubah Yang Banyak Dan Kompleks," *J. Infomatika Pertan.*, vol. Volume 16, 2007.
- [10] Nuriyah, "Perbandingan Metode chi-square automatic interaction detection (chaid) dan classification and regression tree (cart) Dalam Menentukan Klasifikasi Alumni UIN Sunan Kalijaga Berdasarkan Masa Studi," Universitas Islam Negeri Sunan Kalijaga, 2013.
- [11] R. Timofeev, *Classification and Regression Trees (CART) Theory and Applications*. Berlin: Humboldt University, 2004.
- [12] D. Susanto, S., dan Suryadi, *Pengantar Data Mining*. Yogyakarta: Andi Offset, 2010.
- [13] "Machine Learning Repository," 2018. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set#>.
- [14] M. Han, J., & Kamber, *Data Mining Concepts, Model and Techniques 2nd Edition*. San Fransisco: Elsevier, 2006.