

Penggunaan Metode *Classification Analysis Regression Trees* dan *Iterative Dichotomizer 3* Dalam Mengklasifikasikan Pasien Hipertensi Di Rumah Sakit Umum Daerah Dr. Pirngadi Kota Medan

¹Jeni Yulinda, ²Riri Syafitri Lubis, ³Rima Aprilia

¹²³Program Studi Matematika, Universitas Islam Negeri Sumatera Utara, Indonesia

¹jeni0703192031@uinsu.ac.id, ²riri_syafitri@uinsu.ac.id, ³rima_aprilia@uinsu.ac.id

ARTICLE INFO

Article History:

Diterima : 16-11-2023

Disetujui : 11-12-2023

Keywords:

Hypertension ;
Classification Analysis
Regression Trees ;
ID3 ;



ABSTRACT

Abstract: Hypertension is a deadly disease without any symptoms. The high cases of hypertension in Indonesia poses a challenge for health development in Indonesia because treating people with hypertension takes a long time and costs a lot. To find out the factors that influence hypertensive patients and reduce the death rate, an intelligent system is needed to classify hypertensive patients early. The methods used for *Classification Analysis Regression Trees* and *Iterative Dichotomizer 3*. The results of this study show that the *Classification Analysis Regression Trees* method produces an accuracy of 93,75%. Meanwhile, the *Iterative Dichotomizer 3* method produces a low accuracy of 87,5%. So the *Classification Analysis Regression Trees* method is better at classifying hypertensive patients than the *Iterative Dichotomizer 3* method. Variables that effect it are produced by the classification analysis regression trees method namely diastolic blood pressure and iterative dichotomizer 3 namely heredity as parent node.

Abstrak: Hipertensi merupakan salah satu penyakit yang mematikan tanpa disertai gejala. Tingginya kasus hipertensi di Indonesia menyebabkan tantangan bagi pembangunan kesehatan di Indonesia karena perawatan penderita hipertensi membutuhkan waktu yang lama dan biaya yang tidak sedikit. Untuk mengetahui faktor yang mempengaruhi pasien hipertensi serta mengurangi angka kematian diperlukan sistem cerdas untuk mengklasifikasikan pasien hipertensi secara dini. Metode yang digunakan untuk pengklasifikasian yaitu *Classification Analysis Regression Trees* dan *Iterative Dichotomizer 3*. Adapun hasil dari penelitian ini diperoleh bahwa metode *Classification Analysis Regression Trees* menghasilkan akurasi sebesar 93,75%. Sedangkan metode *Iterative Dichotomizer 3* menghasilkan akurasi yang rendah yaitu 87,5%. Sehingga Metode *Classification Analysis Regression Trees* lebih baik dalam mengklasifikasikan pasien hipertensi dibandingkan metode *Iterative Dichotomizer 3*. Variabel yang mempengaruhinya yang dihasilkan oleh metode *Classification Analysis Regression Trees* yaitu tekanan darah diastolik dan *Iterative Dichotomizer 3* yaitu Keturunan sebagai *parent node*.



<https://doi.org/10.31764/justek.vXIY.ZZZ>



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

A. LATAR BELAKANG

Hipertensi adalah suatu keadaan dimana tekanan darah meningkat melebihi batas normal. Seseorang dikatakan hipertensi bila memiliki tekanan darah sistolik ≥ 140 mmHg dan atau tekanan darah diastolik ≥ 90 mmHg, pada pemeriksaan yang berulang (Hastuti, 2022). Hipertensi atau penyakit tekanan darah tinggi adalah kondisi kronis yang ditandai dengan peningkatan tekanan darah pada dinding arteri. Kondisi ini menyebabkan jantung bekerja lebih keras untuk mengedarkan darah melalui pembuluh darah ke seluruh tubuh. Hal ini dapat mengganggu aliran darah, merusak pembuluh darah bahkan menyebabkan penyakit degeneratif, hingga kematian. Hipertensi sering kali disebut sebagai pembuluh gelap (*silent Killer*), karena termasuk penyakit yang mematikan tanpa disertai dengan gejala lebih dahulu (Yudha, *et al.*, 2018).

Hipertensi merupakan penyakit yang sampai sekarang banyak ditemukan di dunia, bahkan hingga saat ini kasus hipertensi terus meningkat seiring kemajuan zaman yaitu dengan perubahan pola hidup yang tidak sehat (Arum, 2019). Hipertensi sampai saat ini masih menjadi tantangan utama pembangunan kesehatan di Indonesia karena tingginya angka prevalensi hipertensi. Hipertensi memiliki faktor risiko yang menyebabkan beban utama bagi negara berkembang, karena perawatan penderita hipertensi membutuhkan waktu yang lama dan biaya yang tidak sedikit. Kejadian dan prevalensi hipertensi terus meningkat di seluruh dunia.

Berdasarkan prediksi WHO pada tahun 2025 prevalensi hipertensi di seluruh dunia pada orang dewasa mencapai 29,2% (Depkes RI, 2007). Hipertensi merupakan satu dari sepuluh penyakit yang menyebabkan mortalitas Asia. Kasus hipertensi meningkat disebabkan karena masih terbatasnya akses dalam pengobatan dan sedikitnya pengetahuan akan hipertensi. Menurut Kurnia, Anih dan Ners dalam bukunya *self management* hipertensi, banyak faktor risiko yang mempengaruhi pasien hipertensi diantaranya yaitu jenis kelamin, usia, tekanan darah sistolik, tekanan darah diastolik, obesitas, keturunan, dan penyakit penyerta.

Berdasarkan penjelasan di atas, diperlukan suatu sistem cerdas yang dapat mengklasifikasikan hipertensi secara dini berdasarkan kondisi pasien. Klasifikasi berkaitan erat dengan Data *Mining* karena ada banyak metode klasifikasi di dalamnya. Untuk melakukan pengklasifikasian pada pasien hipertensi digunakan metode *Classification Analysis Regression Trees* (CART) dan *Iterative Dechotomizer 3* (ID3) untuk mengatasi adanya dugaan hubungan antara faktor-faktor yang mempengaruhi pasien hipertensi.

Classification Analysis Regression Trees (CART) pertama kali diperkenalkan oleh oleh Breiman, Friedman, Olshen dan Stone pada tahun 80-an. *Classification Analysis Regression Trees* (CART) dikembangkan untuk melakukan analisis klasifikasi terhadap variabel respon nominal, ordinal maupun kontinu. CART juga dapat menyeleksi variabel yang paling penting dan variabel interaksi saat menentukan hasil atau variabel prediktor (Misshuari, *et al.*, 2020).

Algoritma CART memiliki beberapa kelebihan dibandingkan dengan algoritma yang lain seperti, variabel dependen ataupun independen dalam CART tidak terikat oleh

asumsi distribusi tertentu, dapat mengetahui interaksi antar variabel independen dan hasil klasifikasi yang lebih mudah dipahami, dan tidak berlaku adanya transformasi data (Jatmiko, *et al.*, 2019). Jika variabel dependen kontinu, pohon yang dihasilkan disebut pohon regresi (*regression tree*), tetapi jika variabel dependennya kategoris, pohon yang dihasilkan disebut pohon klasifikasi (*classification tree*).

Terdapat 3 tahapan dalam pembangunan Algoritma *CART* (Sauddin, *et al.*, 2019) yaitu sebagai berikut:

1. Pembentukan pohon Klasifikasi

Pemilahan pemilah disetiap simpul digunakan untuk mendapatkan simpul yang tingkat kehomogenan nilai variabelnya paling besar. Pada proses pemilahan sampel yang dipilih didasarkan pada aturan kriteria pemilihan terbaik (*goodness-of-split*). Untuk menghitung tingkat keheterogenan suatu kelas melalui suatu *node* tertentu dalam pohon klasifikasi menggunakan *impurity measure* $i(t)$ yang mampu mendapatkan fungsi pemilah yang optimal. Fungsi *impurity measure* $i(t)$ yang digunakan yaitu *indeks gini* (Distante, *et al.*, 2020). Jika kelas objek dinyatakan dengan k , $k=1,2,\dots,m$, dimana m adalah jumlah kelas untuk variabel/output respon y (Misshuari, *et al.*, 2020). Bentuk Umum dari Indeks Gini pada simpul t kelas ke- j adalah sebagai berikut :

$$i(t) = 1 - \sum_{j=1}^k p^2(j|t) \quad (1)$$

Selanjutnya pembentukan pohon akan berhenti jika sudah mencapai batas maksimum level yang telah ditentukan pada kriteria pohon maksimal (Insan, *et al.*, 2020). Pemisahan pohon pada simpul t akan menjadi t_R dan t_L dengan menggunakan persamaan berikut:

$$R(t) > R(t_R) + R(t_L) \quad (2)$$

dengan, $R(t)$ adalah jumlah simpangan kuadrat dari simpul t , $R(t_R)$ adalah jumlah simpangan kuadrat dari simpul kanan pada simpul t , dan $R(t_L)$ adalah Jumlah simpangan kuadrat dari simpul kiri pada simpul t .

Tahap selanjutnya dalam pembentukan pohon klasifikasi *CART* adalah proses pelabelan kelas (*class assignment*). Pelabelan pada terminal *node* berdasarkan aturan jumlah anggota kelas terbanyak yang ada pada variabel respon, (Ikhsan, *et al.*, 2022) jika:

$$P(j_0|t) = \max_j P(j|t) = \max_j \frac{N_t(j)}{N(t)} \quad (3)$$

dengan j_0 adalah label kelas untuk simpul terminal t yang memberikan nilai dugaan kesalahan klasifikasi pada simpul t terbesar.

2. Pemangkasan Pohon Klasifikasi

Semakin banyak pemilahan yang dilakukan maka tingkat akurasi akan semakin tinggi. Sehingga dapat diatasi dengan pemangkasan pada pohon klasifikasi maksimal untuk mendapatkan pohon klasifikasi optimal. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang optimal, yaitu:

$$R(T) = \sum_{t \in \bar{T}} r(t)P(t) = \sum_{t \in \bar{T}} R(t) \quad (4)$$

Algoritma klasifikasi lainnya yaitu *iterative dichotomizer 3*. *Iterative Dichotomizer 3* merupakan algoritma pembelajaran paling dasar pada pohon keputusan yang melakukan pencarian secara menyeluruh data semua kemungkinan pohon keputusan dikembangkan oleh J. Ross Quinlan sejak tahun 1989 (Jollyta, *et al.*, 2020). Algoritma ID3 dapat diimplementasikan menggunakan fungsi rekursif. Algoritma ID3 membangun

pohon keputusan dari atas ke bawah, mulai dengan mengidentifikasi variabel, kemudian mengevaluasi setiap variabel tersebut dengan menggunakan kumpulan data sampel.

Kelebihan dari algoritma *iterative dichotomizer 3* adalah dapat menetapkan aturan yang mudah dipahami, membangun pohon keputusan dengan aman, dan hanya memerlukan beberapa percobaan variabel hingga semua data diklasifikasikan. Untuk membangun pohon keputusan algoritma *iterative dichotomizer 3* menggunakan variabel dengan *informasion gain* tertinggi yang akan dijadikan *root* (Kustiyahningsih, et al., 2021).

Langkah-langkah dalam membangun Algoritma ID3 (Salsabila, et al., 2023), sebagai berikut.

1. Menghitung nilai *entropy* dan *information gain* dari setiap atribut.

Entropy dapat dihitung menggunakan persamaan (Tarigan, et al., 2019) berikut :

$$Entropy(S) = \sum_{i=1}^k -p_i \log_2 p_i \quad (5)$$

dengan, S merupakan himpunan kelas klasifikasi (data set), k merupakan banyaknya kelas klasifikasi dan p_i = proporsi untuk kelas i

Setelah mendapatkan nilai *entropy* untuk suatu kumpulan data, maka dapat mengukur efektivitas suatu atribut dalam mengklasifikasikan data menggunakan *information gain*. Untuk menghitung *information gain* menggunakan persamaan berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} entropy(S_i) \quad (6)$$

dengan, A merupakan variabel pada S , $|S_i|$ merupakan jumlah sampel untuk i , $|S|$ merupakan keseluruhan sampel data dan $Entropy(S_i)$ merupakan *entropy* tiap sampel bernilai i .

2. Memilih variabel yang memiliki nilai *information gain* paling besar yang akan menjadi *root node*.
3. Membentuk *node* yang berisi variabel tersebut.
4. Mengulangi proses perhitungan *information gain* yang akan terus dilakukan sampai setiap data telah termasuk ke dalam kelas yang sama.

Penelitian yang dilakukan oleh Hana, Fida Maisa., et al. (2023) menggunakan *Classification Analysis Regression Tress* (CART) untuk Klasifikasi penyakit diabetes menghasilkan tingkat akurasi pohon klasifikasi dengan pruning dan prepruning menghasilkan akurasi yang cukup besar yaitu 96,15%. Sedangkan klasifikasi dengan algoritma CART tanpa pruning dan prepurning menghasilkan akurasi yang sempurna yaitu sebesar 100%. Dalam hal ini penerapan *Classificaton and regression trees* mampu mampu memprediksi dengan baik. (Hana, et al., 2023)

Penelitian yang menggunakan metode *Iterative Dichotomizer 3* (ID3) dilakukan oleh Irwan, et al. (2022) yang meneliti tentang pohon keputusan dalam memprediksi masa studi mahasiswa UIN Alauddin Makassar menggunakan algoritma ID3 dan C4.5. Tingkat akurasi klasifikasi model pada metode ID3 sebesar 90,91% dan akurasi C4.5 sebesar 78,79%. Hal ini menunjukkan model yang dihasilkan algoritma ID3 lebih baik dibandingkan C4.5 (Irwan, 2022).

Berdasarkan penelitian sebelumnya algoritma CART dan ID3 memiliki akurasi yang baik dalam mengklasifikasikan suatu data sehingga peneliti ingin membandingkan performa kedua metode ini untuk memprediksi pasien hipertensi. Dimana metode klasifikasi digunakan untuk mengetahui faktor yang mempengaruhi pasien hipertensi di Rumah Sakit Umum Dr. Pirngadi Kota Medan.

B. METODE PENELITIAN

Jenis penelitian yang digunakan *Classification Analysis Regression Tress* (CART) dan *Iterative Dichotomizer 3* (ID3) pada penelitian ini adalah pendekatan studi literatur dan kuantitatif. Studi literatur adalah proses mengumpulkan referensi sebagai acuan untuk memecahkan sebuah permasalahan dengan mempelajari dan mengkaji buku-buku, jurnal serta artikel. Penelitian kuantitatif adalah metode penelitian dengan menganalisis populasi dan sampel tertentu dengan menguji hipotesis menggunakan teori yang sudah ada sebelumnya.

Populasi pada penelitian ini adalah pasien hipertensi Rumah Sakit Umum Daerah Dr. Pirngadi Kota Medan dengan jumlah pasien sebesar 429 orang. Sampel yang digunakan pada penelitian ini sebesar 81 orang. Teknik pengambilan sampel pada penelitian ini yaitu menggunakan teknik pengambilan sampel acak atau *random sampling/probability sampling*. Penelitian ini menggunakan variabel respon dan variabel prediktor, sebagai berikut:

1. Variabel respon yang digunakan adalah kejadian hipertensi menggunakan skala nominal.

$$Y(x) = \begin{cases} 1 & , \text{jika } x \text{ hipertensi Primer} \\ 0 & , \text{jika } x \text{ hipertensi sekunder} \end{cases}$$

2. Variabel prediktor yang digunakan yaitu variabel yang berpengaruh dalam kejadian hipertensi, Variabel-variabel tersebut adalah sebagai berikut:

Tabel 1. Variabel Prediktor

| No. | Variabel | Keterangan | Kategori | Skala |
|-----|----------|-------------------------|------------------------------|---------|
| 1. | X_1 | Jenis Kelamin | 1. Laki-laki 2. Perempuan | Nominal |
| 2. | X_2 | Usia | - | Rasio |
| 3. | X_3 | Tekanan Darah Sistolik | - | Rasio |
| 4. | X_4 | Tekanan Darah Diastolik | - | Rasio |
| 5. | X_5 | Berat Badan | - | Rasio |
| 6. | X_6 | Keturunan | 1. Ada 2. Tidak Ada | Nominal |
| 7. | X_7 | Penyakit Lain | 1. Ada 2. Tidak Ada | Nominal |

C. HASIL DAN PEMBAHASAN

Data penelitian menggunakan data pasien hipertensi yang didapatkan dari rekam medis Rumah Sakit Umum DR. Pirngadi Kota Medan tahun 2022 dengan sampel data sebanyak 81 data pasien hipertensi. Berdasarkan hasil survei yang dilakukan terdapat 17 pasien hipertensi sekunder dan 64 pasien hipertensi primer. Dari keseluruhan data yang berjumlah 81 pasien, diambil sebanyak 65 pasien (80%) sebagai data *training* untuk membentuk pohon klasifikasi dan 16 pasien (20%) sebagai data *testing* untuk melihat kemampuan pohon klasifikasi dalam memprediksi pasien hipertensi.

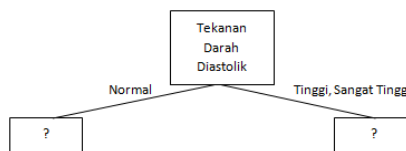
Classification Analysis Regression Trees

Setelah dilakukan pembagian data langkah selanjutnya yaitu pemilahan pemilah dengan menggunakan *goodness of split*. Nilai tertinggi *goodness of split* akan menjadi *parent node*. Nilai *goodness of split* pada masing-masing simpul dapat dilihat pada Tabel 2.

Tabel 2. *Goodness Of Split* masing-masing variabel

| No | Variabel | Pemilah | | <i>Goodness Of Split</i> |
|-----|----------------------------|-------------------------------|--------------------------------|--------------------------|
| | | <i>Candidat Split</i> Kiri | <i>Candidat Split</i> Kanan | |
| 1. | Jenis Kelamin | Laki-laki | Perempuan | 0,1619 |
| 2. | Usia | Muda | Parobaya, Tua | -0,2732 |
| 3. | Usia | Parobaya | Muda, Tua | 0,0999 |
| 4. | Usia | Tua | Muda, Parobaya | 0,1216 |
| 5. | Tekanan Darah Sistolik | Normal | Tinggi, Sangat Tinggi | -0,1149 |
| 6. | Tekanan Darah Sistolik | Tinggi | Normal, Sangat Tinggi | 0,1611 |
| 7. | Tekanan Darah Sistolik | Sangat Tinggi | Normal, Tinggi | 0,1471 |
| 8. | Tekanan Darah Diastolik | Normal | Tinggi, Sangat Tinggi | 0,1819 |
| 9. | Tekanan Darah Diastolik | Tinggi | Normal, Sangat Tinggi | 0,1162 |
| 10. | Tekanan Darah Diastolik | Sangat Tinggi | Normal, Tinggi | -0,029 |
| 11. | Berat Badan | < 65 | ≥ 65 | 0,1471 |
| 12. | Keturunan | Ya | Tidak | 0,1256 |
| 13. | Penyakit Penyerta | Ya | Tidak | 0,1216 |

Berdasarkan tabel diatas dapat dilihat bahwa nilai *goodness of split* tertinggi adalah 0,1819 pada kandidat simpul 8. Sehingga calon simpul 8 yaitu tekanan darah diastolik menjadi *parent node*. Simpul ini mengalami pemilahan menjadi tekanan darah diastolik normal sebagai cabang kiri dan tekanan darah diastolik tinggi dan sangat tinggi sebagai cabang kanan. Simpul-simpul ini akan terus mengalami pemilahan sampai dengan *terminal node*.



Gambar 1. Pemilahan pemilah simpul utama (*Parent Node*)

Kemudian melakukan pelabelan pada setiap *node* berdasarkan aturan jumlah anggota terbanyak dengan menggunakan persamaan (3). Proporsi yang terbesar menjadi label kelas sebuah simpul. Perhitungan pelabelan *node* pada simpul 1 yaitu sebagai berikut.

$$P(\text{Hipertensi Primer}|1) = \frac{51}{65} = 0,7692$$

$$P(\text{Hipertensi Sekunder}|1) = \frac{14}{65} = 0,2308$$

Maka, Simpul 1 diberi label hipertensi primer karena proporsi kelas hipertensi primer lebih besar daripada kelas hipertensi sekunder. Lakukan perhitungan pelabelan *node* hingga simpul 13.

Langkah selanjutnya yaitu pemangkasan pohon klasifikasi dengan mengambil simpul kiri dan simpul kanan pada simpul *t*. Jika simpul kanan dan simpul kiri memenuhi persamaan $R(t) = R(t_L) + R(t_R)$, maka simpul kiri dan simpul kanan dipangkas. Pemangkasan pohon klasifikasi dapat menggunakan persamaan (4). Pada simpul 8 sebagai nilai $R(t)$ diperoleh,

$$r(\text{simpul 8}) = 1 - \max P(j|\text{simpul 8}) = 1 - 0,8182 = 0,1818$$

$$P(\text{simpul 8}) = \frac{11}{65} = 0,1692$$

$$R(\text{simpul 8}) = r(\text{simpul 8}) * P(\text{simpul 8}) = 0,1818 * 0,1692 = 0,0308$$

Kemudian menghitung nilai $R(t_L)$ dan $R(t_R)$ pada simpul 14 dan simpul 15. Pada simpul 14 diperoleh,

$$r(\text{simpul 14}) = 1 - \max P(j|\text{simpul 14}) = 1 - 0,75 = 0,25$$

$$P(\text{simpul 14}) = \frac{4}{65} = 0,0615$$

$$R(\text{simpul 14}) = r(\text{simpul 14}) * P(\text{simpul 14}) = 0,25 * 0,0615 = 0,0154$$

Pada simpul 15 diperoleh,

$$r(\text{simpul 15}) = 1 - \max P(j|\text{simpul 15}) = 1 - 0,8571 = 0,1429$$

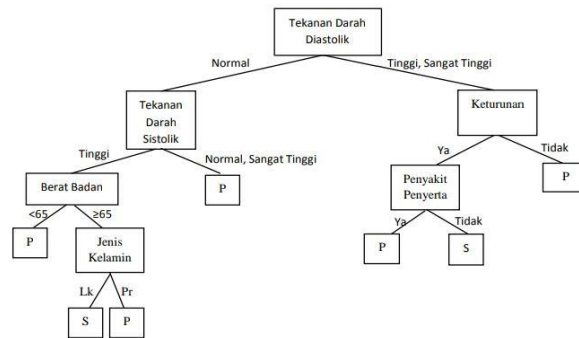
$$P(\text{simpul 15}) = \frac{7}{65} = 0,1077$$

$$R(\text{simpul 15}) = r(\text{simpul 15}) * P(\text{simpul 15}) = 0,1429 * 0,1077 = 0,0154$$

Maka,

$$R(\text{simpul 14}) + R(\text{simpul 15}) = 0,0154 + 0,0154 = 0,0308 = R(\text{simpul 8})$$

Proses pemangkasan pohon dilakukan sampai tidak ada lagi kemungkinan adanya pemangkasan. Sehingga pohon klasifikasi setelah dilakukan proses pemangkasan dapat dilihat pada Gambar 2.



Gambar 2. Pohon Klasifikasi setelah dilakukan pemangkasan

Selanjutnya, dilakukan pengujian akurasi untuk mengetahui tingkat keakuratan model yang diperoleh. Berdasarkan pohon pengklasifikasian CART pada Gambar 2, akurasi model dihitung menggunakan data *testing*. Berikut ini perhitungan tingkat akurasi ketepatan klasifikasinya dapat dilihat pada Tabel 3.

Tabel 3. Hasil Ketetapan Akurasi Klasifikasi Pohon Klasifikasi

| Observasi | Prediksi | | Total |
|-----------|----------|----------|-------|
| | Primer | Sekunder | |
| Primer | 13 | 0 | 13 |
| Sekunder | 1 | 2 | 3 |
| Total | 14 | 2 | 16 |

$$1 - APPER = \frac{13 + 2}{16} \times 100\% = 93,75 \%$$

Berdasarkan perhitungan nilai akurasi pada pohon klasifikasi CART diperoleh sebesar 93,75%. Selanjutnya nilai akurasi ini akan dibandingkan dengan metode ID3 untuk melihat algoritma yang paling baik dalam mengklasifikasikan pasien hipertensi

Iterative Dichotomizer 3 (ID3)

Pembentukan aturan pada algoritma ID3 dilakukan dengan membentuk pohon keputusan berdasarkan nilai *gain* tertinggi dari setiap variabel. Perhitungan nilai *entropy* dan *gain* menggunakan data *training*. Langkah pertama menghitung nilai *entropy* semua data dengan menggunakan persamaan (5) sebagai berikut.

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^k -p_i \log_2 p_i \\
 &= -\left(\frac{51}{65} \log_2 \frac{51}{65}\right) - \left(\frac{14}{65} \log_2 \frac{14}{65}\right) \\
 &= 0,7517
 \end{aligned}$$

Kemudian menghitung *entropy* total dengan variabel masing-masing. Sebagai contoh nilai *entropy* variabel usia sebagai berikut:

$$\begin{aligned}
 Entropy(Usia, Muda) &= \sum_{i=1}^k -p_i \log_2 p_i \\
 &= -\left(\frac{1}{2} \log_2 \frac{1}{2}\right) - \left(\frac{1}{2} \log_2 \frac{1}{2}\right) \\
 &= 1
 \end{aligned}$$

$$Entropy(Usia, Parobaya) = -\left(\frac{16}{21} \log_2 \frac{16}{21}\right) - \left(\frac{5}{21} \log_2 \frac{5}{21}\right)$$

$$\begin{aligned}
 &= 0,7919 \\
 Entropy(Usia, Tua) &= -\left(\frac{34}{42} \log_2 \frac{34}{42}\right) - \left(\frac{8}{42} \log_2 \frac{8}{42}\right) \\
 &= 0,7025
 \end{aligned}$$

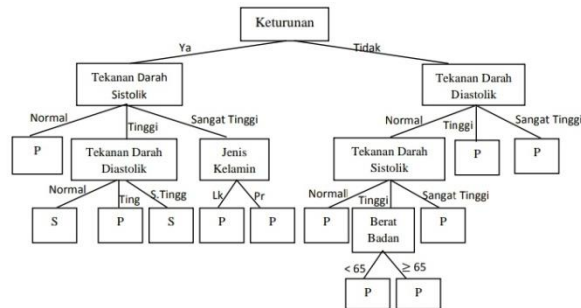
Lakukan perhitungan *entropy* total pada semua variabel. Selanjutnya untuk mendapatkan *root node* (simpul utama) maka dihitung nilai *information gain* semua variabel menggunakan persamaan (6) sebagai berikut.

$$\begin{aligned}
 Gain(S, Usia) &= 0,7517 - \left(\frac{2}{65}(1) + \frac{21}{65}(0,7919) + \frac{42}{65}(0,7024)\right) \\
 &= 0,1111
 \end{aligned}$$

Tabel 4. Hasil Perhitungan *Information Gain*

| Node | Variabel | Values | Jumlah Kasus | Primer | Sekunder | Entropy | Gain | |
|------|----------|------------------------|---------------|--------|----------|---------|--------|---------------|
| 1 | Total | | 65 | 51 | 14 | 0,7517 | | |
| | Usia | Muda | | 2 | 1 | 1 | 1 | 0,0111 |
| | | Parobaya | | 20 | 16 | 5 | 0,7919 | |
| | | Tua | | 42 | 34 | 8 | 0,7025 | |
| | | Tekanan Darah Sistolik | Normal | 8 | 5 | 3 | 0,9544 | |
| | | Tinggi | 31 | 24 | 7 | 0,7706 | | |
| | | Tekanan Darah Sistolik | Sangat Tinggi | 26 | 22 | 4 | 0,6194 | |
| | | Tekanan Darah Sistolik | Normal | 31 | 21 | 10 | 0,9072 | 0,0481 |
| | | | Tinggi | 22 | 19 | 3 | 0,5746 | |
| | | | Sangat Tinggi | 12 | 11 | 1 | 0,4138 | |
| | | Berat Badan | <65 | 39 | 29 | 10 | 0,8213 | 0,0111 |
| | | | >65 | 26 | 22 | 4 | 0,6194 | |
| | | Keturunan | Ya | 21 | 13 | 8 | 0,9587 | 0,0529 |
| | | | Tidak | 44 | 38 | 6 | 0,5746 | |
| | | Penyakit Penyerta | Ya | 42 | 34 | 8 | 0,7025 | 0,0047 |
| | | | Tidak | 23 | 17 | 6 | 0,8281 | |
| | | Jenis Kelamin | Laki-Laki | 32 | 25 | 7 | 0,7579 | 0 |
| | | | Perempuan | 33 | 26 | 7 | 0,7455 | |

Berdasarkan hasil perhitungan nilai gain pada Tabel 4 maka diperoleh nilai tertinggi yaitu variabel keturunan sebesar 0,0529, sehingga variabel keturunan akan menjadi *root node* pada pohon klasifikasi. Variabel keturunan memiliki 2 kategori yaitu ya dan tidak yang akan dijadikan cabang dari *root node* keturunan. Setelah dilakukan rekursif terhadap node-node yang lain dan pada terminal node maka diperoleh pohon keputusan dengan 12 terminal node dapat dilihat pada Gambar 3.



Gambar 3. Pohon Klasifikasi ID3

Selanjutnya, dilakukan pengujian akurasi untuk mengetahui tingkat keakuratan model yang diperoleh. Berdasarkan pohon pengklasifikasian ID3 pada Gambar 3, akurasi model dihitung menggunakan data testing. Berikut ini perhitungan tingkat akurasi ketepatan klasifikasinya dapat dilihat pada Tabel 5.

Tabel 5. Hasil Ketetapan Akurasi Klasifikasi Pohon Klasifikasi

| Observasi | Prediksi | | Total |
|-----------|----------|----------|-------|
| | Primer | Sekunder | |
| Primer | 13 | 0 | 13 |
| Sekunder | 2 | 1 | 3 |
| Total | 15 | 1 | 16 |

$$1 - APPER = \frac{13 + 2}{16} \times 100\% = 93,75 \%$$

Berdasarkan perhitungan nilai akurasi pada pohon klasifikasi CART diperoleh sebesar 93,75%.

D. SIMPULAN DAN SARAN

Hipertensi dijuluki sebagai *silent killer* atau pembunuh diam-diam karena penyakit ini tidak memiliki gejala yang spesifik, dapat menyerang siapa saja, kapan saja, bahkan berujung kematian. Sehingga dibutuhkan suatu sistem cerdas yang dapat mengklasifikasikan pasien hipertensi. Dengan menggunakan algoritma *Classification Analysis Regression Trees* dan *Iterative Dichotomizer 3* menghasilkan klasifikasi pasien hipertensi di rumah sakit umum daerah Dr. Pirngadi kota Medan bahwa tingkat akurasi ketetapan klasifikasi menggunakan *Classification Analysis Regression Trees* (CART) sebesar 93,75%. Sedangkan nilai akurasi pada *Iterative Dechotomizer 3* (ID3) sebesar 87,5%. Sehingga dapat disimpulkan metode *Classification Analysis Regression Trees* (CART) lebih baik dibandingkan dengan metode *Iterative Dechotomizer 3* (ID3) dalam mengklasifikasikan pasien hipertensi.

Variabel yang sangat mempengaruhi pada metode *Classification Analysis Regression Trees* yaitu tekanan darah diastolik. Sedangkan pada metode *Iterative Dichotomizer 3* (ID3) variabel yang sangat mempengaruhi yaitu variabel keturunan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Islam Negeri Sumatera Utara dan RSUD Pirngadi Medan yang telah membantu dalam penelitian ini.

REFERENSI

- Arum, Y. T. G. (2019). Hipertensi pada Penduduk Usia Produktif (15-64 Tahun). *Higeia*, **3(3)**, 345-356
- Distante, A., & Distante, C. (2020). *Handbook of Image Processing and Computer Vision*. Italy : Springer Internasional Publishing.
- Hana, F. M., Wahyudin. w. C., Ulya, S., dan Negara, D. S. (2023). Implementasi Algoritma CART Dalam Klasifikasi Penyakit Diabetes. *Jurnal Ilmu Komputer dan Matematika*. 1-8
- Hastuti, A. P. (2022). *Hipertensi*. Jawa Tengah: Lakeisha.
- Ikhsan, M., & Rosmaini, E. 2022. Use of Classification and Regression Tree (CART) Method for Classification of Labor Force Participation Levels in Medan City in 2019. *Jurnal Matematika dan Pendidikan Matematika*, **V(2)**, 95-103
- Insan, N., Hadijati, M., & Irwansyah. (2020). Perbandingan Metode *Classification and Regression Trees* (CART) dengan *Naive Bayes Classification* (NBC) dalam klasifikasi Status Gizi Balita di Kelurahan Pagesangan Barat. *Eigen Mathematics Journal*, **03(01)**. 9-22
- Irwan, Sauddin, A., Ida, N. (2022). Penerapan Pohon Keputusan Dalam Memprediksi Masa Studi Mahasiswa UIN Alauddin Makassar. *Jurnal Informatika Sains dan Teknologi (INSTEK)*. **7(2)**, 201-210
- Jatmiko, Y. A., Padmadisastra. S., & Chadidjah. A. (2019). Analisis Perbandingan Kinerja CART Konvensional, *Bagging* dan *Random Forest* Pada Klasifikasi Objek: Hasil Dari Dua Simulasi. *Media Statistika*. **12(2)**, 1-12
- Jollyta, D., Ramdhan, W., & Zarlis, M. (2020). *Konsep Data Mining dan Penerapan*. Yogyakarta : Deepublish.
- Kurnia, A., & Ners. (2020). *Self Management Hipertensi*. Surabaya : CV. Jakad Media Publishing.
- Kustiyahningsih, Y., Rahmanita. E., Purbandini., Ikromul, I., & Sasmeka, V. (2021). Metode Multi Criteria Decision Making untuk Pendukung Keputusan. Malang: Media Nusa Creative.
- Kemendes RI. (2019). Hipertensi Penyakit Paling Banyak Didap Masyarakat. Available: <https://www.kemkes.go.id/article/view/19051700002/hipertensi-penyakit-paling-banyak-diidap-masyarakat.html> [accessed Maret. 3, 2023]
- Mangunsong, E. Y., Prasetya, N. H., Rakhmawati, Fibri. 2021. Classification of Pneumonia Risk Factor Characteristics in Toddlers Using Classification and Regression Trees (CART). *Journal of Innovation and Technology In Mathematics and Mathematics Education*, **1(2)**, 88-93
- Misshuari, I. W., & Chairunnisah. (2020). Penerapan Metode *Classification And Regression Trees* (CART) untuk Menentukan Faktor-Faktor yang Mempengaruhi Pembayaran Kredit oleh Nasabah di PT BPRS Gebu Prima Medan. *Karismatika*, **6(3)**, 10-20
- Sauddin, A., Alwi, W., & An, A.N. 2019. Klasifikasi Tingkat Partisipasi Angkatan Kerja Kota Makassar Menggunakan Metode CART. *Jurnal Matematika dan Statistika serta Aplikasinya*, **7(2)**, 20-36
- Yudha, B. L., Muflikhah, L., Wihandika, R. C. Klasifikasi Risiko Hipertensi Menggunakan Metode *Neighbor Weighted K-Nearest Neighbor* (NWKNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, **2(2)**, 897-904