

## IMPLEMENTASI DATA MINING DALAM KLASIFIKASI HASIL DIAGNOSA PASIEN BPJS MENGGUNAKAN ALGORITMA CART

Nurhaeka Tou<sup>1)</sup>, Putri Mentari Endraswari<sup>2)</sup>

<sup>12</sup> Teknik Elektro Fakultas Teknik, Universitas Bangka Belitung, Gang IV No.1, Balun ijuk, Merawang,  
Bangka Belitung  
Co Responden Email: nurhaeka@ubb.ac.id

### Article history

Received 15 April 2022

Revised 05 June 2022

Accepted 19 June 2022

Available online 30 June 2022

### Keywords

Data Mining, Classification,  
Patient, BPJS, CART

### Abstract

*Puskesmas is one of the BPJS patient referral health facilities that can generate and collect a lot of medical record data every day. The pile of medical record data is generated from routine operational activities and is also used for operational needs. However, sometimes the heap of data is left unattended and unused. So far, Puskesmas Umbulharjo 1 has used this data only to make policies in the form of health education to the public, by providing information on the number of patients seeking treatment, types of illness, and reports of patient discharge. However, the pattern of disease tendencies suffered by the community has not been explored further, so that it can be used as a reference in health education so that it is more targeted. Therefore, this study aims to classify the diagnosis results of BPJS patients based on the relationship between symptoms and types of disease using the CART algorithm. The data used is secondary data obtained from medical records of the results of the examination of BPJS patients at the Umbulharjo 1 Health Center as much as 200 data. Based on the results of the analysis carried out, the CART classification shows that the types of TB, pneumonia, gastritis, and hypertension are influenced by symptoms of headache, abdominal pain, nausea, and vomiting. The results of the accuracy of the classification accuracy using the CART method of 71.5%.*

### Riwayat

Diterima 15 April 2022

Revisi 05 Juni 2022

Disetujui 19 Juni 2022

Terbit 30 Juni 2022

### Kata Kunci

Data Mining, Klasifikasi,  
Pasien, BPJS, CART

### Abstrak

*Puskesmas atau Pusat Kesehatan Masyarakat merupakan salah satu fasilitas kesehatan rujukan pasien BPJS yang dapat menghasilkan dan mengumpulkan banyak data rekam medis setiap harinya. Tumpukan data rekam medis tersebut dihasilkan dari kegiatan rutin operasional dan digunakan juga untuk kebutuhan operasional. Namun, terkadang tumpukan data tersebut dibiarkan begitu saja dan tidak digunakan. Selama ini pihak Puskesmas Umbulharjo 1 memanfaatkan data tersebut hanya untuk membuat kebijakan berupa penyuluhan kesehatan kepada masyarakat, dengan memberikan informasi jumlah pasien yang berobat, jenis penyakit yang diderita, serta laporan keluarnya pasien. Akan tetapi, terkait pola kecenderungan penyakit yang diderita oleh masyarakat belum digali lebih dalam, agar dapat digunakan sebagai acuan dalam penyuluhan kesehatan, sehingga lebih tepat sasaran. Oleh karena itu, penelitian ini bertujuan mengklasifikasikan hasil diagnosa pasien BPJS berdasarkan hubungan gejala dan jenis penyakit menggunakan algoritma CART. Data yang digunakan merupakan data sekunder yang diperoleh dari rekam medis hasil pemeriksaan pasien BPJS di Puskesmas Umbulharjo 1 sebanyak 200 data. Berdasarkan hasil analisis yang dilakukan, klasifikasi CART menunjukkan bahwa jenis penyakit TBC, Penunomia, Gastritis, dan Hipertensi dipengaruhi oleh gejala sakit kepala, sakit perut, serta mual dan muntah. Dengan hasil akurasi ketepatan klasifikas menggunakan metode CART sebesar 69 %.*

## PENDAHULUAN

Masalah kesehatan merupakan bentuk tanggung jawab bersama antar masyarakat dan pemerintah. Kesehatan memiliki peran yang sangat penting dalam kehidupan masyarakat, karena kesehatan adalah aset kesejahteraan jiwa, badan dan sosial untuk setiap individu. Sehingga, Pemerintah dituntut untuk dapat membuat sebuah sistem pelayanan kesehatan yang bermutu dan berkualitas.

Salah satu upaya pemerintah dalam meningkatkan kualitas kesehatan pada masyarakat adalah dengan membuat program jaminan sosial. Jaminan sosial adalah salah satu program yang diselenggarakan untuk menjamin masyarakat Indonesia mendapatkan kebutuhan hidup dasar yang layak (Suwardika, 2019) . Program jaminan sosial yang diselenggarakan oleh pemerintah diawali dengan Jamkesmas, Jamkesda, ASKES, dan Badan Penyelenggara Jaminan Sosial (BPJS) yang menjadi program baru dari pemerintah (Yunus, Husni, & Mufadhhal, 2021).

Program BPJS Kesehatan adalah Badan Usaha milik Negara yang kemudian berubah menjadi Badan Hukum Publik yang ditugaskan oleh pemerintah untuk menyelenggarakan jaminan kesehatan bagi masyarakat Indonesia. BPJS kesehatan diharapkan dapat memberikan proteksi, sehingga semua masyarakat mendapatkan akses kesehatan secara merata (Eka & Zain, 2014). Adanya program BPJS kesehatan sangat membantu masyarakat dalam meringankan biaya pengobatan, dan saat ini banyak ditemui pasien yang menggunakan layanan BPJS di Pusat Kesehatan Masyarakat (Puskesmas).

Puskesmas Umbulharjo 1 Yogyakarta merupakan salah satu fasilitas kesehatan rujukan BPJS yang menerima banyak pasien dan dapat mengumpulkan banyak data rekam medis setiap harinya. Tumpukan data rekam medis ini dihasilkan dari kegiatan rutin operasional setiap hari, dan digunakan pula untuk kebutuhan operasional. Namun, terkadang tumpukan data tersebut dibiarkan begitu saja dan tidak dipergunakan, padahal dari data yang tertumpuk dan berlimpah ini, muncul sebuah pertanyaan mengenai hal apa yang bisa dilakukan dan dipelajari dari keseluruhan data dan Informasi tersebut (Amriana, Joeffie, & Meidji, 2019)

Sejauh ini puskesmas memanfaatkan data rekam medis untuk membuat kebijakan berupa penyuluhan kesehatan kepada masyarakat, dengan memberikan informasi jumlah pasien yang berobat, jenis penyakit yang diderita, serta laporan keluarnya pasien. Namun, terkait pola kecenderungan penyakit yang diderita oleh masyarakat belum digali lebih dalam, agar dapat digunakan sebagai acuan dalam penyuluhan kesehatan, sehingga lebih tepat sasaran.

Oleh karena itu, dibutuhkan mekanisme atau metode yang sesuai dan otomatis untuk menemukan pola, pengetahuan, dan informasi yang diperlukan bagi Puskesmas dan dinas terkait. Sehingga, dalam penelitian ini penulis tertarik melakukan penelitian mengenai "Implementasi Data Mining dalam Klasifikasi Hasil Diagnosa Pasien BPJS Menggunakan Algoritma CART".

*Data mining* merupakan sebuah proses untuk mendapatkan informasi dan pengetahuan baru dari *Big Data* menggunakan metode tertentu (Asuddin, Alwi, & Ningsih, 2019) . Salah satu metode data mining adalah *Classification and Regression Tree* (CART). CART merupakan metode statistik nonparametrik yang dapat menggambarkan hubungan antar variabel respon (dependen) dengan satu atau bahkan lebih variabel prediktor (independen) (Safitri, Wiguna, Kusumawardani, & Wibowo, 2021). Algoritma CART memiliki beberapa keunggulan dibandingkan dengan algoritma yang lain seperti, variabel dependen ataupun independen dalam CART tidak menggunakan asumsi distribusi tertentu, variabel independennya dapat berupa kategorik (nominal atau ordinal) ataupun kontinu, dalam CART tidak berlaku adanya transformasi data, dan intepretasinya mudah dipahami (Chairunisa, Adiwijaya, & Astuti, 2021).

Beberapa penelitian menggunakan metode CART sudah pernah dilakukan sebelumnya, seperti dalam Penelitian yang dilakukan oleh (Safitri et al., 2021) dengan judul "Klasifikasi Kepuasan Pelanggan Menggunakan Metode Classification and Regression Tree (CART)". Penelitian tersebut menghasilkan klasifikasi CART yang dilakukan menunjukkan bahwa untuk menjadi *Promoters*, maka aplikasi e-boarding harus memiliki nilai *ease of learnig dan satisfaction* yang tinggi.

Penelitian kedua dilakukan oleh (Agwil, Fransiska, & Hidayati, 2020) dengan judul “Analisis Ketepatan Waktu Lulus Mahasiswa dengan Menggunakan Bagging CART”. Hasil penelitian tersebut menunjukkan hasil klasifikasi CART tunggal dan Bagging CART menunjukkan hasil akurasi yang baik. Dimana, hasil klasifikasi CART ketepatan waktu lulus mahasiswa dipengaruhi oleh IPK, biaya kuliah, dan daerah asal.

Berbeda dengan dua penelitian sebelumnya, hasil dari penelitian ini akan berupa pohon keputusan klasifikasi hasil diagnosa pasien BPJS berdasarkan hubungan gejala dan jenis penyakit, yang kemudian diimplementasikan ke dalam sistem. Penelitian ini diharapkan dapat dijadikan acuan oleh pihak Puskesmas Umbulharjo 1 sebagai pendukung dalam mengambil keputusan, melakukan penyuluhan kesehatan secara optimal sesuai kebutuhan, melakukan pencegahan penyakit yang tepat sasaran, dan dapat meningkatkan mutu Puskesmas Umbulharjo 1 dalam memberikan pelayanan kesehatan pada pasien BPJS Kesehatan.

## TINJAUAN PUSTAKA

### Data Mining

*Data mining* merupakan sebuah proses yang mengadopsi teknik statistik, matematika, kecerdasan buatan dan *machine learning* agar dapat mengekstraksi informasi dan pengetahuan dari berbagai *database* yang berukuran besar (Anggraeni, Saputra, & Noranita, 2013). Selain itu, *data mining* juga didefinisikan sebagai “menambang” ataupun menemukan informasi dan pengetahuan dari sekumpulan data yang berukuran besar (Asuddin et al., 2019). *Data mining* ini merupakan bagian dari *Knowledge Discovery in Database* (KDD) yang melakukan perubahan terhadap data mentah menjadi sebuah informasi, pengetahuan baru, ataupun keterkaitan pola atau hubungan yang terdapat dalam data yang berukuran besar (Anggraeni et al., 2013).

### Classification and Regression Tree (CART)

Metode CART adalah sebuah metode yang menggunakan teknik eksplorasi dari pohon keputusan, dimana nilai setiap binernya memiliki dua cabang dalam setiap node keputusan (Eka & Zain, 2014). Selain itu, metode CART juga didefinisikan sebagai yang

mampu menggambarkan hubungan antar variabel dependen (terikat) dengan variabel independen (bebas) (Margasari, 2014). Metode CART ini bertujuan untuk menemukan kelompok dari sebuah data yang akurat yang menjadi ciri dari suatu teknik pengklasifikasian (Sari, 2020). Metode CART memiliki beberapa aturan dalam menyelesaikan masalah sebagai berikut:

#### 1. Menentukan Kandidat *Split* (pemisah)

Penentuan masing-masing parameter dalam metode CART dilakukan dengan cara membagi kandidat pemisah menjadi dua bagian yaitu kandidat kanan dan kandidat kiri berdasarkan pada *class variable* (Subarkah, 2020). Kandidat pemisah yang dipilih bergantung pada jenis variabel terikat dan kriteria *goodness of split*. Namun, sebelum melakukan pemilihan kandidat pemisah, terlebih dahulu melakukan pencarian *gain information* pada setiap simpul menggunakan persamaan 1.

$$GI(t) = - \sum_{j=1}^n P(j|t) \log_2 P(j|t) \quad (1)$$

Keterangan :

$GI(t)$  = *Gain information* pada simpul  $t$   
 $P(j|t)$  = Kelas  $j$  terhadap simpul  $t$  dimana  $j = 1, 2, 3, \dots, n$  dengan persamaan  $P(j|t) = \frac{n_j(t)}{n(t)}$   
 $n_j(t)$  = Jumlah data kelas  $j$  pada simpul  $t$   
 $n(t)$  = Jumlah data pada simpul  $t$

Teknik pemilihan kandidat pemisah pada metode CART menggunakan *indeks gini*  $i(t)$ . *Indeks gini* adalah suatu pengukuran tingkat keragaman kelas antara nilai-nilai atribut pada sebuah simpul di dalam klasifikasi (Asuddin et al., 2019). Tujuan dari *indeks gini* untuk dapat menemukan kandidat pemisah yang paling optimal dengan menggunakan persamaan 2.

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

Keterangan:

$i(t)$  = *Indeks Gini*  
 $P(j|t)$  = Bagian kelas  $j$  pada simpul  $t$ , dimana  $j = 1, 2, 3, \dots, n$  dengan  $P(j|t) = \frac{n_j(t)}{n(t)}$   
 $n_j(t)$  = Jumlah data kelas  $j$  pada simpul  $t$   
 $n(t)$  = Jumlah data pada simpul  $t$

Parameter yang dipilih akan membentuk menjadi sebuah himpunan kelas yang diebut dengan simpul atau *node*. *Node* tersebut akan melakukan pemilihan secara iterasi sampai pada iterasi terakhir yang memperoleh terminal *nodes* (Sumartini, 2015). Tahap selanjutnya yang dilakukan adalah memilih kriteria *goodness-of-split* yang akan menjadi tahap evaluasi pemilihan kandidat pemisah  $s$  pada  $t$  menggunakan persamaan 3.

$$\phi(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (3)$$

Keterangan:

$t_R$  = *Right node* (simpul kanan)  
 $t_L$  = *Left node* (simpul kiri)  
 $P_R$  = Jumlah objek pada parameter  $t_R$   
 $P_L$  = Jumlah objek pada parameter  $t_L$

Probabilitas jumlah objek yang terdapat pada paramter  $t_R$  dan  $t_L$  dapat dihitung menggunakan persamaan 4 dan 5.

$$P_L = \frac{\text{Jumlah kasus simpul kiri}}{\text{Jumlah data training}} \quad (4)$$

$$P_R = \frac{\text{Jumlah kasus simpul kanan}}{\text{Jumlah data training}} \quad (5)$$

Selanjutnya, dalam proses pengembangan pohon keputusan dilakukan pencarian pada semua kemungkinan kandidat pemisah pada simpul  $t$ , sehingga akan ditemukan kandidat pemisah  $s^*$  yang dapat memberikan penurunan nilai keheterogenan tertinggi dengan menggunakan persamaan 6.

$$\phi(s^*, t) = \max_{s \in S} \phi(s, t) \quad (6)$$

## 2. Pemilihan Simpul Terminal

Simpul  $t$  akan ditetapkan menjadi simpul terminal, jika pada simpul  $t$  tidak terjadi penurunan nilai keheterogenan berdasarkan kriteria yang apada pada *goodness-of-split*. Selanjutnya, proses pencarian pohon keputusan akan berhenti, apabila iterasi yang dilakukan sudah mencapai batas maksimum jumlah level yang telah ditentukan (Suwardika, 2019). Sehingga, tahap pemecahan pohon keputusan terhadap *node*  $t$  menjadi  $t_R$  dan  $t_L$  menggunakan persamaan 7.

$$R(t) > R(t_R) + R(t_L) \quad (7)$$

Keterangan :

$R(t)$  = Nilai kesalahan terhadap node  $t$   
 $R(t_L)$  = Nilai kesalahan terhadap simpul kiri pada node  $t$ .

$R(t_R)$  = Nilai kesalahan terhadap simpul kanan pada node  $t$ .

## 3. Penentuan Label Kelas

Penentuan label kelas merupakan sebuah proses untuk mengidentifikasi setiap *nodes* yang ada pada kelas tertentu (Putu & Rusli, 2017). Proses penentuan label kelas dilakukan pada *terminal nodes*, *nonterminal nodes*, dan *root nodes*. Namun, Penentuan label kelas yang paling penting dilakukan pada *terminal nodes*, yang digunakan untuk memprediksi objek pada kelas yang terdapat dalam simpul terminal (Asuddin et al., 2019). Menentukan label kelas menggunakan persamaan 8.

$$P(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (8)$$

Keterangan:

$P(j|t)$  =Jumlah kelas  $j$  pada simpul  $t$   
 $N_j(t)$  = Jumlah kelas pada simpul  $t$   
 $N(t)$  = Jumlah pengamatan disimpul  $t$

## Pengujian Akurasi

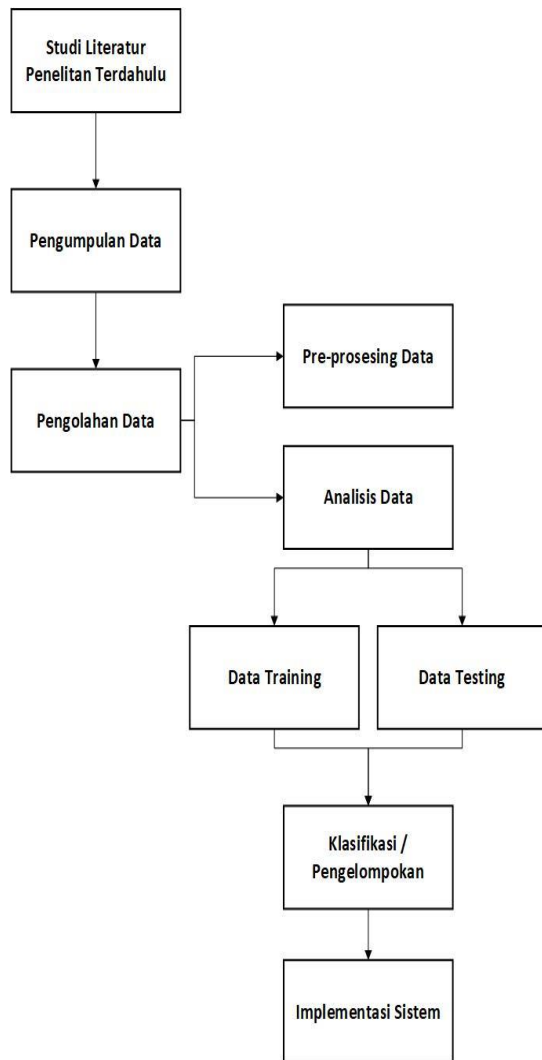
Pengujian akurasi digunakan untuk melihat persentase ketepatan suatu algoritma dalam melakukan klasifikasi. Perhitungan akurasi menggunakan persamaan 9.

$$\text{Accuracy} = \frac{TP+TN}{n} \quad (9)$$

Dimana, TP adalah jumlah *record* positif yang diprediksi secara benar oleh algoritma klasifikasi, TN adalah jumlah *record* negatif yang diprediksi salah oleh algoritma klasifikasi, dan  $n$  adalah jumlah keseluruhan data yang diklasifikasi (Putri, Nurwindiana, & Khoiriyah, 2020).

## METODE PENELITIAN

Penelitian yang dilakukan memiliki beberapa tahapan dalam mengklasifikasi hasil diagnosa pasien BPJS yaitu studi literatur, pengumpulan data, pengolahan data, analisis data, dan implementasi sistem. Diagram alir penelitian disajikan pada Gambar 1.



Gambar 1. Diagram alir penelitian

## 1. Studi Literatur

Tahapan studi literatur dilakukan untuk mencari acuan-acuan untuk dijadikan dasar dalam penelitian ini, terutama penelitian-penelitian yang menggunakan metode *Classification and Regression Tree* (CART) dalam pengelompokan ataupun pengklasifikasian.

## 2. Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian adalah teknik wawancara bersama Pak Jarwo dan Pak Arnawan selaku petugas rekam medis untuk mendapatkan data rekam medis pasien BPJS. Data rekam medis yang didapatkan berupa file excell sebanyak 250 data.

## 3. Pengolahan Data

### a. Pre-prosesing Data

Tahap pre-prosinging data dilakukan untuk melakukan pengecekan terhadap kelengkapan data sebelum masuk ke tahap analisis data. Pengecekan data meliputi, pengecekan *missing value*, redundansi data, ketidak relevan data dan cleaning data yang tidak lengkap. Hasil dari tahapan ini adalah sebuah data yang siap untuk dikomputasi.

### b. Data Selection

Tahap *data selection* digunakan untuk memilih data yang akan digunakan dan data - data yang tidak digunakan dalam proses mining. Dari keseluruhan data yang didapatkan, data yang digunakan dalam proses mining yaitu data umur, jenis kelamin, antropi, batuk, sesak nafas, sakit kepala, mual dan muntah, menggigil, demam dan jenis penyakitnya (Pneumonia, TBC, Hipertensi dan Gastritis).

### c. Transformasi Data

Tahap ini dilakukan untuk mentransformasikan atau penggabungan data ke dalam format yang sesuai dengan format proses data mining. Dalam penelitian ini data yang perlu dilakukan transformasi adalah data umur. Data umur dalam data yang diperoleh berkisar 1 tahun sampai 70 tahun. kemudian dilakukan transformasi data umur yang disajikan pada Tabel 1. Hasil transformasi data umur.

Tabel 1. Hasil transformasi data umur

Data Umur	Transformasi Umur
1 - 9	1 – 10 Tahun
10 - 19	10 – 19 Tahun
20 – 49	20 – 49 Tahun
50 – 69	50 – 69 Tahun
70 – 80	>= 70 Tahun

## 4. Analisa Data

Tahap analisis data dilakukan untuk mengklasifikasi hasil diagnosa dari pasien BPJS. Klasifikasi data dilakukan berdasarkan jenis penyakit yaitu Pneumonia, TBC,

Hipertensi dan Gastritis. Proses klasifikasi dilakukan dengan menentukan data set yang akan dianalisis dan dilakukan perhitungan secara komputasi dengan memasukkan algoritma CART dalam coding sistem menggunakan Python.

Data yang digunakan dibagi menjadi dua yaitu *data training* dan *data testing*. Data training dalam penelitian ini digunakan untuk menguji dan melatih agar data yang dipakai dapat melakukan klasifikasi sesuai aturan yang ada dalam algoritma CART. Kemudian, *data testing* merupakan data baru yang digunakan untuk menguji keakuratan atau akurasi dalam sebuah sistem.

Setelah melakukan klasifikasi, proses selanjutnya adalah menghitung nilai uji accuracy, untuk mengetahui keandalan dari metode CART yang digunakan.

### 5. Implementasi Sistem

Tahap ini digunakan untuk mengimplementasikan proses klasifikasi pasien BPJS ke dalam sebuah sistem berbasis web, agar bisa digunakan oleh pihak puskesmas untuk mengklasifikasikan data baru dari pasien BPJS.

### HASIL DAN PEMBAHASAN

Penelitian ini dilakukan untuk mengklasifikasikan hasil diagnosa pasien BPJS menggunakan algoritma *Classification and Regression Tree* (CART) berdasarkan gejala dan jenis penyakit yang diderita. Data yang digunakan dalam penelitian merupakan data sekunder hasil pemeriksaan pasien BPJS pada Puskesmas Umbulharjo 1 Yogyakarta. Setelah dilakukan tahap *data selection*, *data preprocessing*, dan *data transformation*, didapatkan sebanyak 200 data dengan 12 atribut yang digunakan dalam penelitian ini.

Selanjutnya, data tersebut dilakukan proses klasifikasi menggunakan algoritma *Classification and Regression Tree* (CART). Tahap pertama proses klasifikasi menggunakan CART adalah membagi data menjadi data training dan data testing. Dengan data training tersebut, kemudian dilakukan pemilihan *candidate split* yang dapat dilihat pada Tabel 2.

Tabel 2. *Candidate split* (CS)

CS	<i>Left Child Node</i> , tR	<i>Right Child Node</i> , tL
----	--------------------------------	---------------------------------

1	Usia = 1-9 Thn	Usia ∈ {10 – 19, 20 – 49, 50 - 69 > 70 tahun}
2	Usia = 10-19 Thn	Usia ∈ {1 – 9, 20 – 49, 50 - 69 > 70 tahun}
3	Usia = 20-49 Thn	Usia ∈ {1 – 9, 10 - 19, 50 - 69 > 70 tahun}
4	Usia = 50-69 Thn	Usia ∈ {1 – 9, 10 - 19, 20 - 49, > 70 tahun}
5	Usia = > 70 Thn	Usia ∈ {1 – 9, 10 - 19, 20 - 49, 50 - 69 tahun}
6	JK = Laki-laki	JK ∈ {Perempuan}
7	JK = Perempuan	JK ∈ {Laki - Laki}
8	Batuk = Ya	Batuk ∈ {Tidak}
9	Batuk = Tidak	Batuk ∈ {Ya}
10	Sesak Nafas = Ya	Sesak Nafas ∈ {Tidak}
11	Sesak Nafas = Tidak	Sesak Nafas ∈ {Ya}
12	Mual & Muntah = Ya	Mual & Muntah ∈ {Tidak}
13	Mual & Muntah = Tidak	Mual & Muntah ∈ {Ya}
14	Demam = Ya	Demam ∈ {Tidak}
15	Demam = Tidak	Demam ∈ {Ya}
16	Sakit Kepala = Ya	Sakit Kepala ∈ {Tidak}
17	Sakit Kepala = Tidak	Sakit Kepala ∈ {Ya}
18	Menggigil = Ya	Menggigil ∈ {Tidak}
19	Menggigil = Tidak	Menggigil ∈ {Ya}
20	Sakit Perut = Ya	Sakit Perut ∈ {Tidak}
21	Sakit Perut = Tidak	Sakit Perut ∈ {Ya}
22	Lemas = Ya	Lemas ∈ {Tidak}
23	Lemas = Tidak	Lemas ∈ {Ya}
24	Keringat Dingin = Ya	Keringat Dingin ∈ {Tidak}
25	Keringat Dingin = Tidak	Keringat Dingin ∈ {Ya}
26	Nafsu Makan Turun = Ya	Nafsu Makan Turun ∈ {Tidak}
27	Nafsu Makan Turun = Tidak	Nafsu Makan Turun ∈ {Ya}
28	Berat Badan Turun = Ya	Berat Badan Turun ∈ {Tidak}
20	Berat Badan Turun = Tidak	Berat Badan Turun ∈ {Ya}

Tahap selanjutnya setelah menentukan *Candidate Split* adalah menentukan indeks gini menggunakan persamaan 2 dan 3. Tabel 3 menunjukkan hasil perhitungan indeks gini pada pencarian node pertama, Tabel 4 hasil perhitungan indeks gini node kedua, dan Tabel 5 hasil perhitungan indeks gini node ketiga.

Tabel 3. Hasil perhitungan indeks gini pada node pertama

Split	PL	PR	2PLPR	Q(s t)	$\phi(s t)$
1	0,160	0,840	0,269	1,969	<b>0,529</b>
2	0,050	0,950	0,095	0,874	<b>0,083</b>
3	0,330	0,670	0,442	0,559	<b>0,247</b>
4	0,335	0,665	0,446	0,542	<b>0,241</b>
5	0,125	0,875	0,219	0,537	<b>0,117</b>
6	0,400	0,605	0,484	0,418	<b>0,202</b>
7	0,605	0,395	0,478	0,418	<b>0,200</b>
8	0,380	0,620	0,471	2,013	<b>0,944</b>
9	0,620	0,380	0,471	2,013	<b>0,944</b>
10	0,155	0,845	0,262	2,000	<b>0,524</b>
11	0,845	0,155	0,262	2,000	<b>0,524</b>
12	0,775	0,225	0,349	2,016	<b>0,703</b>
13	0,225	0,775	0,349	2,016	<b>0,703</b>
14	0,155	0,845	0,262	2,000	<b>0,524</b>
15	0,845	0,155	0,262	2,000	<b>0,524</b>
<b>16</b>	<b>0,620</b>	<b>0,380</b>	<b>0,471</b>	<b>2,005</b>	<b>0,949</b>
17	0,380	0,620	0,471	2,005	<b>0,949</b>
18	0,225	0,775	0,349	2,016	<b>0,703</b>
19	0,775	0,225	0,349	2,016	<b>0,703</b>
20	0,330	0,670	0,442	2,000	<b>0,884</b>
21	0,670	0,330	0,442	2,000	<b>0,884</b>
22	0,290	0,710	0,412	1,990	<b>0,819</b>
23	0,710	0,290	0,412	1,990	<b>0,819</b>
24	0,230	0,770	0,354	2,000	<b>0,708</b>
25	0,770	0,230	0,354	2,000	<b>0,708</b>
26	0,230	0,770	0,354	2,000	<b>0,708</b>
27	0,770	0,230	0,354	2,000	<b>0,708</b>
28	0,230	0,770	0,354	2,000	<b>0,708</b>
29	0,770	0,230	0,354	2,000	<b>0,708</b>

Hasil perhitungan indeks gini pada node pertama didapatkan nilai  $\phi(s|t)$  terbesar 0.949 yaitu pada *split* ke 16 dengan kategori Sakit Kepala yang akan menjadi *root node*. Kemudian, berdasarkan Tabel 2 *split* ke 16 juga diperoleh calon cabang kiri (node 2) Sakit Kepala = YA dengan notkah keputusan Penunomia dan TBC, dan calon cabang kanan (node 3) Sakit Kepala = Tidak dengan notkah keputusan Gastritis dan Hipertensi.

Tahap selanjutnya menggunakan node 2 untuk menghitung seluruh nilai indeks gini pada notkah keputusan Pneunomia dan TBC, serta menggunakan node 3 untuk menghitung

seluruh nilai indeks gini pada notkah keputusan Gastritis dan Hipertensi. Tabel 4 menunjukkan hasil perhitungan indeks gini pada node 2.

Tabel 4. Hasil perhitungan indeks gini pada node kedua

Split	PL	PR	2PLPR	Q(s t)	$\phi(s t)$
1	0,408	0,592	0,483	1,674	0,809
2	0,053	0,947	0,100	0,722	0,072
3	0,289	0,711	0,411	1,148	0,472
4	0,224	0,776	0,347	1,050	0,365
5	0,026	0,974	0,051	0,838	0,043
6	0,500	0,500	0,500	0,000	0,000
7	0,500	0,500	0,500	0,000	0,000
8	0,000	1,000	0,000	1,000	0,000
9	0,000	1,000	0,000	1,000	0,000
10	0,408	0,592	0,483	1,996	0,964
11	0,592	0,408	0,483	1,996	0,964
<b>12</b>	<b>0,408</b>	<b>0,592</b>	<b>0,483</b>	<b>2,000</b>	<b>0,966</b>
13	0,592	0,408	0,483	2,000	0,966
14	0,408	0,592	0,483	2,000	0,966
15	0,592	0,408	0,483	2,000	0,966
16	Data sudah digunakan				
17	1,000	0,000	0,000	1,000	0,000
18	0,592	0,408	0,483	2,000	0,966
19	0,408	0,592	0,483	2,000	0,966
20	0,000	1,000	0,000	1,000	0,000
21	1,000	0,000	0,000	1,000	0,000
22	0,000	1,000	0,000	1,000	0,000
23	1,000	0,000	0,000	1,013	0,000
24	0,592	0,408	0,483	2,000	0,966
25	0,408	0,592	0,483	2,000	0,966
26	0,592	0,408	0,483	2,000	0,966
27	0,408	0,592	0,483	2,000	0,966
28	0,592	0,408	0,483	2,000	0,966
29	0,408	0,592	0,483	2,000	0,966

Hasil perhitungan indeks gini pada node 2 didapatkan nilai  $\phi(s|t)$  sebesar 0.996. Namun, pada hasil perhitungan tersebut nilai 0.996 terdapat pada lebih dari satu *split*, sehingga nilai yang digunakan adalah nilai pada *split* yang terlebih dahulu muncul yaitu *split* 12. Selanjutnya, berdasarkan Tabel 4 diperoleh calon cabang kiri (node 4) Mual dan Muntah = YA dengan notkah keputusan TBC, dan calon

cabang kanan (node 5) Mual dan Muntah = Tidak, dengan notkah keputusan Penunomia.

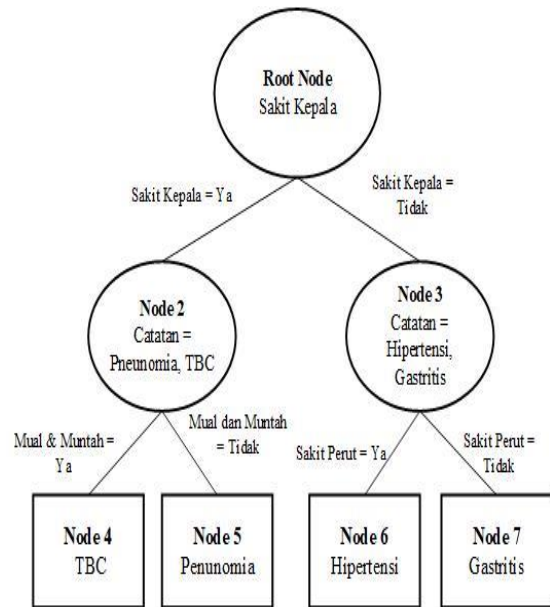
Tabel 5. Hasil perhitungan indeks gini pada node ketiga

Split	PL	PR	2PLPR	Q(s t)	$\emptyset(s t)$
1	0,008	0,992	0,016	1,000	0,016
2	0,048	0,952	0,092	1,000	0,092
3	0,355	0,258	0,183	1,500	0,275
4	0,403	0,589	0,475	0,276	0,131
5	0,185	0,815	0,302	0,240	0,073
6	0,339	0,661	0,448	0,314	0,141
7	0,661	0,339	0,448	0,314	0,141
8	1,000	0,000	0,000	1,000	0,000
9	1,000	0,000	0,000	1,000	0,000
10	1,000	0,000	0,000	1,000	0,000
11	1,000	0,000	0,000	1,000	0,000
12	Sudah Digunakan				
13	0,000	1,000	0,000	1,000	0,000
14	0,000	1,000	0,000	1,000	0,000
15	1,000	0,000	0,000	1,000	0,000
16	Sudah Digunakan				
17	0,000	1,000	0,000	1,000	0,000
18	0,000	1,000	0,000	1,000	0,000
19	1,000	0,000	0,000	1,000	0,000
<b>20</b>	<b>0,532</b>	<b>0,468</b>	<b>0,498</b>	<b>2,000</b>	<b>0,996</b>
21	0,468	0,532	0,498	2,000	0,996
22	0,468	0,532	0,498	2,000	0,996
23	0,532	0,468	0,498	2,000	0,996
24	0,000	1,000	0,000	1,000	0,000
25	1,000	0,000	0,000	1,000	0,000
26	0,000	1,000	0,000	1,000	0,000
27	1,000	0,000	0,000	1,000	0,000
28	0,000	1,000	0,000	1,000	0,000
29	1,000	0,000	0,000	1,000	0,000

Selanjutnya, hasil perhitungan indeks gini pada node 3 didapatkan nilai  $\phi(s|t)$  sebesar 0.966. Namun, pada hasil perhitungan tersebut nilai 0.966 terdapat pada lebih dari satu *split*, sehingga nilai yang digunakan adalah nilai pada *split* yang terlebih dahulu muncul yaitu *split* 20. Berdasarkan Tabel 5 diperoleh calon cabang kiri (node 6) Sakit Perut = YA dengan notkah keputusan Hipertensi, dan calon cabang kanan

(node 7) Sakit Perut = Tidak dengan notkah keputusan Gastritis.

Proses iterasi perhitungan indeks gini pada node 2 dan node 3 akan berhenti, apabila sebaran sampel telah berada pada satu kelas atau nilai sampel  $< 5$ . Sehingga, node selanjutnya akan dijadikan sebagai node terminal. Adapun hasil akhir pohon keputusan menggunakan algoritma CART disajikan pada Gambar 2.



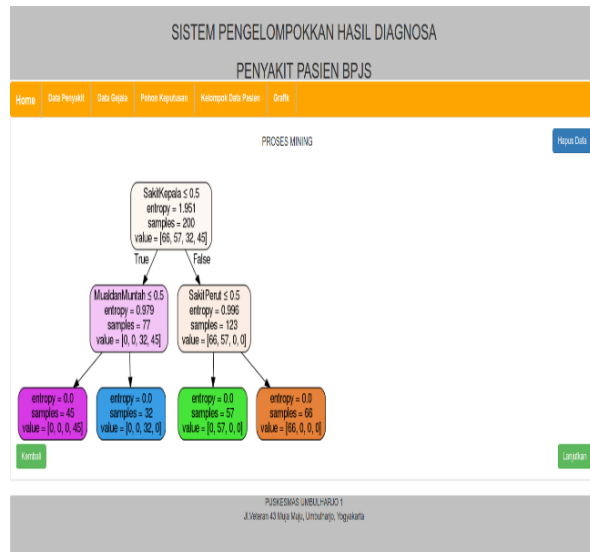
Gambar 2. Pohon keputusan

Gambar 2 menunjukkan hasil akhir klasifikasi menggunakan algoritma CART berupa pohon keputusan. Pada Gambar tersebut node 4, node 5, node 6, dan node 7 dijadikan sebagai terminal, karena pada node tersebut jumlah sampel berada pada satu kelas saja. Gambar pohon keputusan tersebut menunjukkan bahwa:

1. Ketika Sakit Kepala = “YA” maka variabel yang muncul adalah Mual dan Muntah
2. Ketika Mual dan Muntah = “YA” maka menghasilkan keputusan TBC.
3. Ketika Mual dan Muntah = “Tidak” maka menghasilkan keputusan Penunomia.
4. Selanjutnya, ketika Sakit Kepala = “Tidak” maka variabel yang muncul adalah Sakit Perut.
5. Ketika Sakit Perut = “YA” maka menghasilkan keputusan Hipertensi
6. Ketika Sakit Perut = “Tidak” maka menghasilkan keputusan Gastritis.



Selanjutnya hasil klasifikasi tersebut diimplementasikan kedalam sistem berbasis web. Adapun tampilan sistemnya dapat dilihat pada Gambar 3 dan Gambar 4.



Gambar 3. Pohon keputusan dalam sistem



Gambar 4. Grafik pengelompokan data pasien

Kemudian, dilakukan perhitungan akurasi untuk melihat ketepatan algoritma CART dalam melakukan pengujian klasifikasi sebagai berikut :

$$Accuracy = \frac{TP + TN}{n}$$

$$Accuracy = \frac{32 + 30 + 29 + 36 + 11}{200}$$

$$Accuracy = 0.69 \times 100 \%$$

$$Accuracy = 69 \%$$

Hasil perhitungan menunjukkan bahwa accuracy ketepatan klasifikasi algoritma CART sebesar 69 %.

### KESIMPULAN

Berdasarkan hasil penelitian implementasi data mining dalam klasifikasi hasil diagnosa pasien BPJS menggunakan algoritma CART, bahwa metode CART dapat mengklasifikasikan hasil diagnosa pasien BPJS berdasarkan gejala yang dialami pasien. Hasil klasifikasi menunjukkan jenis penyakit TBC, Penunomia, Hipertensi dan Gastritis yang dipengaruhi oleh gejala sakit kepala, mual dan muntah, dan sakit perut. Dengan hasil Accuracy ketepatan klasifikasi algoritma CART sebesar 69 %. Sistem web ini cukup membantu dalam mengklasifikasikan data diagnosa pasien BPJS berdasarkan gejala yang dialami pasien. Penelitian ini masih bisa untuk dikembangkan lagi dengan menambahkan jumlah data uji yang lebih banyak lagi dan sistem bisa dikembangkan lagi dengan menambahkan grafik tren penyakit yang dideria setiap bulan.

### REFERENSI

Agwil, W., Fransiska, H., & Hidayati, N. (2020). Analisis Ketepatan Waktu Lulus Mahasiswa dengan Menggunakan Bagging CART. *Jurnal Pendidikan Matematika Dan Matematika*, 6(2), 155–166.

Amriana, A., Joeфри, Y. Y., & Meidji, F. N. (2019). Penerapan Data Mining Untuk Pengelompokan Hasil Diagnosa Penyakit Pasien Pengguna BPJS Kesehatan (Studi Kasus Pada Rsud Undata Palu). *ScientiCO: Computer Science and Informatics Journal*, 1(1), 51. <https://doi.org/10.22487/j26204118.2018.v1.i1.11901>

Anggraeni, H. D., Saputra, R., & Noranita, B. (2013). Aplikasi Data Mining Analisis Data Transaksi Penjualan Obat Menggunakan Algoritma Apriori. *Journal of Informatics and Technology*, 2(2), 5–20. Retrieved from [https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book\\_part](https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part)

Asuddin, A., Alwi, W., & Ningsih. (2019). Klasifikasi Tingkat Partisipasi Angkatan

- Kerja Kota Makassar Menggunakan Metode CART. *Jurnal Matematika Dan Statistika Serta Aplikasinya*, 7(2).
- Chairunisa, R., Adiwijaya, & Astuti, W. (2021). Perbandingan CART dan Random Forest untuk Deteksi Kanker Berbasis Klasifikasi Data Microarray. *Jurnal Rekayasa Sistem Dan Teknologi*, 1(10), 805–812.
- Eka, F., & Zain, I. (2014). Klasifikasi Pengangguran Terbuka Menggunakan CART ( Classification and Regression Tree ) di Provinsi Sulawesi Utara. *Jurnal Sains Dan Seni Pomits*, 3(1), 1–6.
- Margasari, A. (2014). *Penerapan Metode CART (Classification and Regression Trees) dan Analisis Regresi Logistik Biner pada Klasifikasi Profil Mahasiswa FMIPA Universitas Brawijaya*.
- Putri, M. A., Nurwindiana, & Khoiriyah, N. (2020). *Klasifikasi Faktor yang Berpengaruh Terhadap Keparahan Tingkat Luka Pekerja Menggunakan Metode CART ( Classification and Regression Trees )*. 59–79.
- Putu, S. I. K., & Rusli, M. (2017). Ketepatan Klasifikasi Bagging CART Pada Klasifikasi Ketidaktepatan Waktu Kelulusan Mahasiswa STIKOM Bali. *Konferensi Nasional Sistem Dan Informatika*, 237–240.
- Safitri, S. T., Wiguna, C., Kusumawardani, D. M., & Wibowo, I. Y. (2021). Klasifikasi Kepuasan Pelanggan Menggunakan Metode Classification and Regression Tree. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(1), 337–349.
- Sari, A. Y. (2020). *Penerapan Metode Classification And Regression Trees Pada Klasifikasi Kelayakan Peserta BPJS-PBI Di Kelurahan Meranti Pandak Kota Pekanbaru*.
- Subarkah, P. (2020). Penerapan Algoritme Klasifikasi Classification And Regression Trees (Cart) Untuk Diagnosis Penyakit Diabetes Retinopathy. *Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(2), 294–301. <https://doi.org/10.30812/matrik.v19i2.676>
- Sumartini, S. H. (2015). Penggunaan Metode Classification and Regression Trees (CART) untuk Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya (Vol. 4). Retrieved from <https://www.neliti.com/publications/15687/penggunaan-metode-classification-and-regression-trees-cart-untuk-klasifikasi-re>
- Suwardika, G. (2019). Analisis Klasifikasi untuk Memprediksi Penyakit Hepatitis Menggunakan Algoritma Neural Network. *Journal of Education and Learning*, 1(3), 183–191.
- Yunus, M., Husni, M., & Mufadhdhal, M. M. (2021). Klasifikasi Sentimen Terhadap Badan Penyelenggara Jaminan Sosial (BPJS) Pada Media Sosial Twitter Menggunakan Naive Bayes. *Smatika Jurnal*, 11(02), 81–91. <https://doi.org/10.32664/smatika.v11i02.577>