

# PERBANDINGAN ALGORITMA C4.5 DAN NAIVE BAYES DALAM MENDETEKSI HIPERTENSI DI PUSKESMAS BANYUBIRU

<sup>1</sup>Kurniati Pratami, <sup>2</sup>Iwan Setiawan Wibisono

<sup>1,2</sup>Universitas Ngudi Waluyo, Semarang

Email : <sup>1</sup>kurniatipratami72@gmail.com, <sup>2</sup>loyal.wb99@gmail.com

## ABSTRACT

Hypertension is the number 1 cause of death in the world every year because it is the entrance to other diseases, such as: heart, kidney failure, diabetes, and stroke (Director of Prevention and Control of Non-Communicable Diseases (PPTM) of the Indonesian Ministry of Health, dr. Cut Putri Ariane, M.H.Kes.). The purpose of this study was to find the best method for detecting hypertension. In this study, the authors used the C4.5 and Naive Bayes methods. The results showed that the C4.5 model has a better accuracy of 74.00% than the Naive Bayes model with an accuracy of 67.00%. The conclusion of this study is the application of the C4.5 algorithm as the best method for detecting controlled and uncontrolled hypertension at the Banyubiru Health Center.

Keywords: Hypertension, C4.5, and Naive Bayes, Banyubiru Health Center.

## ABSTRAK

Hipertensi menjadi penyebab kematian nomor 1 di dunia setiap tahunnya karena merupakan pintu masuk penyakit lain, seperti : jantung, gagal ginjal, diabetes, dan stroke (Direktur Pencegahan dan Pengendalian Penyakit Tidak Menular (PPTM) Kemenkes RI, dr. Cut Putri Ariane, M.H.Kes.). Tujuan dari penelitian ini adalah mencari metode terbaik dalam mendeteksi penyakit hipertensi. Dalam penelitian ini, penulis menggunakan metode C4.5 dan Naive Bayes. Hasil penelitian menunjukkan bahwa model C4.5 memiliki akurasi yang lebih baik dengan 74,00 % dibandingkan model Naive Bayes dengan akurasi 67,00 %. Kesimpulan penelitian ini adalah penerapan algoritma C4.5 sebagai metode terbaik dalam mendeteksi hipertensi terkendali dan tak terkendali di Puskesmas Banyubiru.

Kata Kunci : Hipertensi, C4.5, dan Naive Bayes, Puskesmas Banyubiru.

## PENDAHULUAN

Hipertensi atau tekanan darah tinggi merupakan terjadinya peningkatan tekanan darah yang tidak normal dalam pembuluh darah arteri dan terjadi secara terus menerus (Muriyarti and Yahya, 2018). Hipertensi merupakan salah satu Penyakit Tidak Menular (PTM) yang distribusinya tinggi dan terus meningkat. Hipertensi merupakan salah satu penyebab utama kematian dini diseluruh dunia. Syer Ree Tee dkk dalam penelitiannya menunjukkan usia, tingkat pendidikan, konsumsi alkohol dan Indeks Masa Tubuh (IMT) merupakan faktor

resiko yang berhubungan dengan distribusi hipertensi pada masyarakat pedesaan.

Syer Ree Tee dkk dalam penelitiannya menunjukkan usia, tingkat pendidikan, konsumsi alkohol dan Indeks Masa Tubuh (IMT) merupakan faktor risiko yang berhubungan dengan distribusi hipertensi pada masyarakat pegunungan (Santosa, Chasani and Pramudo, 2016). Berdasarkan uraian diatas, puskesmas memiliki peran penting sebagai primary health care dalam penemuan kasus penyakit hipertensi. Metode yang paling populer digunakan untuk teknik klasifikasi adalah

Decision Trees, Naive Bayes Classifiers (NBC), Statistical analysis, dan lain lain.

Decision trees adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Naive Bayes adalah sebuah metode klasifikasi yang berakar pada teorema bayes. Ciri utama dari naive bayes adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi. C4.5 adalah salah satu algoritma yang digunakan untuk klasifikasi dataset. Dasar dari algoritma C4.5 adalah pembentukan pohon keputusan (Decision Tree). Neural network adalah klasifikasi data yang mempunyai relasi dengan synapse yang mengelilingi neuron-neuron lainnya (Wibawa et al. 2018).

Beberapa peneliti telah melakukan penelitian pada topik yang sama menggunakan aturan asosiasi SVM dan BPNN untuk deteksi kanker payudara, SVM, dan Naive Bayes untuk deteksi penyakit hati, SVM dan JST untuk deteksi penyakit ginjal. Untuk meningkatkan kinerja model, diperlukan pemilihan fitur yang lebih besar daripada yang lain. Beberapa metode yang digunakan adalah Particle Swarm Optimization (PSO) dan Genetika Algoritma (GA). PSO memberikan hasil yang lebih baik daripada GA dalam hal fungsi kebugaran setiap kriteria nilai. (Chung-Jui Tu dkk. 2007). Hasil membuktikan bahwa PSO-SVM lebih unggul dengan waktu komputasi yang lebih cepat dibandingkan GA-SVM. Sehingga PSO dapat memaksimalkan pembobotan atribut atau pemilihan atribut untuk meningkatkan klasifikasi (Novichasari and Wibisono 2020).

Beberapa peneliti lain juga telah membuktikan bahwa PSO telah meningkatkan klasifikasi akurasi metode dalam beberapa kasus yang berbeda. Prisma Handayanna menerapkan SVM-PSO untuk prediksi diabetes dan Junta Zeniarja mengimplementasikan SVM-PSO untuk opinion mining. Farid Melgani dan Yakoub Bazi membuktikan bahwa PSO-SVM lebih unggul dari SVM, KNN, dan klasifikasi RBF untuk klasifikasi sinyal

elektrokardiogram (Novichasari and Wibisono 2020). Analisis retrospektif dilakukan pada 694 data. Tiga Pohon Keputusan, Empat Algoritma Statistik dan Dua Jaringan Syaraf Tiruan (JST) dibandingkan. Hasilnya, sebuah Multilayer Perceptron (MLP) yang memiliki Jaringan Syaraf Tiruan menunjukkan kinerja yang lebih baik pada hipertensi estimasi dibandingkan metode lainnya.

Aljumah dan Siddiqui menghitung probabilitas dan prediksi hipertensi menggunakan teknik data mining dan menyimpulkan bahwa berhenti merokok adalah intervensi terbaik diikuti dengan olahraga, mengatur pola makan, dan konsumsi obat untuk hipertensi. Kokyer dalam karyanya membuat database hipertensi milik pasien rumah sakit dalam waktu yang berbeda yang meliputi : usia, jenis kelamin, indeks massa tubuh, HDL, LDL, trigliserida, asam urat, merokok dan data dianalisis melalui Tabel Keputusan dan Hutan Acak dan data mining (Afeni, Aruleba, and Oloyede 2017).

## **METODE PENELITIAN**

### **1. Pengumpulan data**

#### **a. Observasi**

Penulis melakukan observasi dengan salah satu staff di puskesmas bertujuan untuk mengumpulkan data-data pasien yang akan dijadikan populasi dalam penelitian ini. Metode pengumpulan data ini dilakukan dengan melakukan pengamatan secara langsung terhadap pasien yang menderita hipertensi dan mengambil 100 data pasien hipertensi untuk dijadikan sample.

#### **b. Wawancara**

Penulis juga melakukan wawancara secara langsung dengan ibu Amronah, AMK yang ahli atau mengerti tentang penyakit hipertensi. Tujuan penulis melakukan wawancara adalah untuk mendapatkan kelengkapan data yang dibutuhkan dan memperoleh

informasi mengenai pasien hipertensi di puskesmas Banyubiru.

c. Studi Pustaka

Penulis mengkaji beberapa informasi melalui buku-buku, jurnal dan media internet yang ada kaitannya dengan penelitian, dibahas untuk dapat dipergunakan sebagai bahan perbandingan atau pelengkap.

Metode ini digunakan untuk mengumpulkan dan mencari data tentang informasi yang dibutuhkan untuk melakukan penelitian.

2. Analisis Data

Penulis menggunakan spesifikasi software dan hardware sebagai alat bantu dalam penelitian yang tercantum pada table dibawah ini.

Tabel 1. Teknik Analisis Data

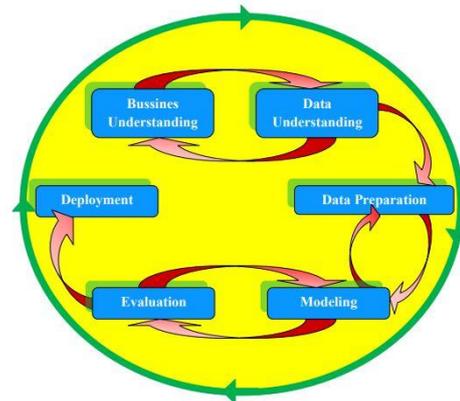
Software	Hardware
Sistem Operasi : Windows 2010	Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz 1.99 GHz
Aplikasi : Rapid Miner versi 5.3	RAM : 4,00 GB

1. Data sampel sama dengan populasi akan dilakukan pengolahan kedalam Microsoft Excel.
2. Pengolahan data menggunakan Microsoft Excel untuk mendapatkan hasil data testing.
3. Pengolahan data menggunakan aplikasi Rapid Miner untuk memasukan data testing yang akan diuji.
4. Data testing akan dilakukan pengolahan menggunakan metode Naive Bayes dan C4.5 kemudian hasil dari data tersebut akan menghasilkan validasi data dengan hasil akurasi yang sama dengan data yang sudah diuji di dalam aplikasi Rapid Miner.
5. Data yang sudah dilakukan pengolahan menggunakan aplikasi Rapid Miner akan menghasilkan keakuratan data hasil dari pengolahan data siswa

tersebut dengan hasil akurasi yang akurat.

3. Metode yang diusulkan

Penulis menggunakan metode Cross-Industry Standart Proses for Data Mining (CRIPS-DM) dikembangkan pada tahun 1996 oleh analis dari beberapa industri. CRIPS-DM menyediakan standart proses data mining sebagai pemecahan masalah secara umum dari bisnis atau unit penelitian. CRIPS-DM memiliki siklus hidup yang terbagi dalam enam fase, yaitu :



Gambar 1. Siklus CRIPS-DM

- Pemahaman Bisnis (Business Understanding) merupakan tahap awal yaitu pemahaman penelitian, penentuan tujuan dan rumusan masalah data mining.
- Pemahaman Data (Data Understanding) dalam tahap ini dilakukan pengumpulan data, mengenali lebih lanjut data yang akan digunakan.
- Pengolahan Data (Data Preparation) adalah pekerjaan berat yang perlu dilaksanakan secara intensif.
- Pemodelan (Modeling) adalah memilih teknik pemodelan yang disesuaikan aturan model untuk hasil yang maksimal. Dapat kembali ke tahap pengolahan untuk menjadikan data ke dalam bentuk yang sesuai dengan model tertentu.
- Evaluasi (Evaluation) adalah mengevaluasi satu atau model yang digunakan dan menetapkan apakah

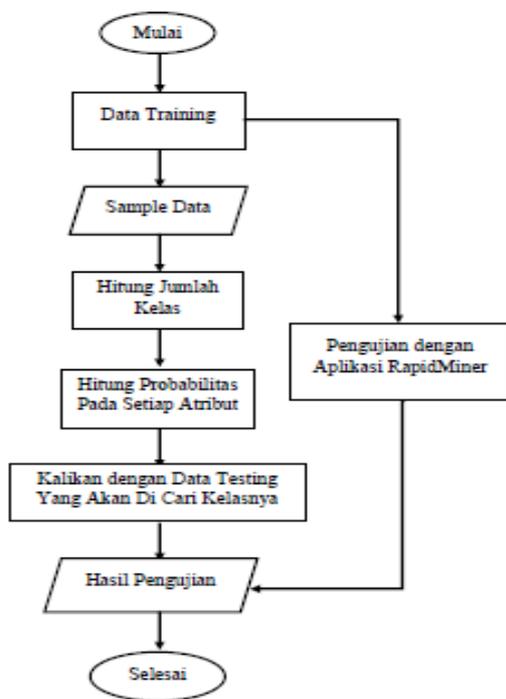
terdapat model yang memenuhi tujuan pada tahap awal. Kemudian menentukan apakah ada permasalahan yang tidak dapat tertangani dengan baik serta mengambil keputusan hasil penelitian.

- Penyebaran (Deployment) adalah menggunakan model yang dihasilkan seperti pembuatan laporan atau penerapan proses data mining pada departemen lain.

3. Hitung jumlah kasus yang sama dengan kelas yang sama sehingga di peroleh nilai pada masing – masing atribut.
4. Kalikan semua hasil sesuai dengan data testing yang akan dicari kelasnya
5. Bandingkan Hasil  
Bandingkan hasil berdasarkan dari perhitungan probabilitas terkendali dan tidak terkendali pada langkah sebelumnya.

a. Metode Naive Bayes

1) Flowchart



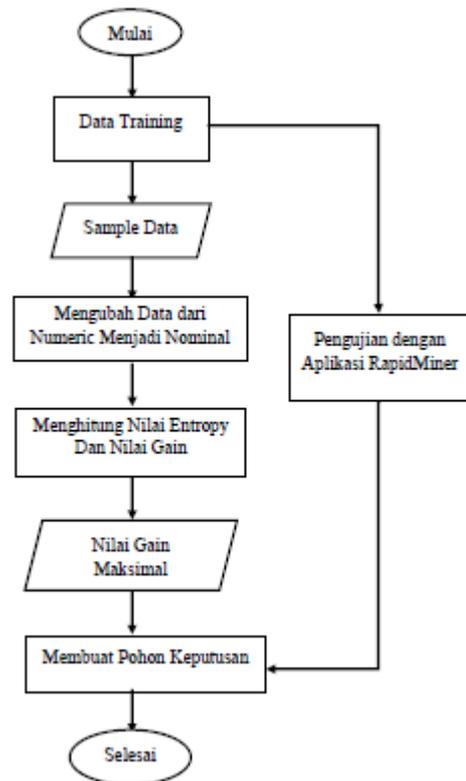
Gambar 2. Flowchart Naive Bayes

2) Langkah – langkah Algoritma Naive Bayes :

1. Siapkan Dataset. Dataset yang digunakan untuk perhitungan manual ini menggunakan dataset pasien hipertensi.
2. Hitung jumlah kelas pada Data Training. Jumlah kelas pada data training terdiri dari dua kategori yaitu terkendali dan tidak terkendali.

b. Metode C4.5 (Decision Tree)

• Flowchart C4.5



Gambar 3 Flowchart C4.5

- Menentukan nilai entropy
- Hitung nilai gain
- Pilih gain maksimal sebagai node akar
- Ulangi langkah-langkah tersebut hingga membentuk pohon keputusan.

**HASIL DAN PEMBAHASAN**

1. Hasil Penelitian

Hasil penelitian berdasarkan tahapan-tahapan proses dalam KDD (Knowledge Data Discovery) sebagai berikut:

- Data Selection, tahapan ini dilakukan untuk memilih data yang sesuai dengan variabel yang dibutuhkan dalam penelitian. Caranya adalah dengan memilih atau menentukan atribut-atribut data mana yang akan digunakan dalam penelitian dari sekelompok data operasional yang ada.
- Pre-processing/Cleaning, proses cleaning tersebut dilakukan terhadap keseluruhan data yang diteliti yang berjumlah 100 pasien. Setelah dilakukan proses cleaning data sejumlah 100, dihasilkan sample data sebanyak 10 record data yang digunakan untuk proses analisis berikutnya.
- Transformation, tahap ini menghasilkan satu recordset data yang siap untuk analisis data.
- Analisis data. Tahap ini menggunakan 2 cara yaitu dengan teknik perhitungan manual dan pengujian pada rapidminer.

Berdasarkan hasil penelitian, dapat diketahui bahwa metode C4.5 sebagai metode terbaik dalam mendeteksi hipertensi terkendali dan tidak terkendali. Hal ini dapat dilihat berdasarkan nilai akurasi C4.5 (74%) yang lebih besar daripada nilai akurasi Naive Bayes (67%).

## 2. Pembahasan dan Pengujian Metode

### - Metode Naive Bayes

#### 1) Perhitungan manual

Langkah – langkah Algoritma Naive Bayes :

1. Siapkan Dataset. Dataset yang digunakan untuk perhitungan manual ini menggunakan dataset pasien hipertensi.
2. Hitung jumlah kelas pada Data Training. Jumlah kelas pada data training terdiri dari dua kategori yaitu terkendali dan tidak terkendali,

sehingga probabilitasnya sebagai berikut:

Jumlah pasien terkendali = 7

Jumlah pasien tidak terkendali = 3

Maka,  $P(C = \text{“terkendali”}) = \frac{7}{10} = 0,7$

$P(C = \text{“tidak terkendali”}) = \frac{3}{10} = 0,3$

3. Hitung jumlah kasus yang sama dengan kelas yang sama sehingga di peroleh nilai pada masing – masing atribut.

Tabel 2. Perhitungan Naive Bayes

Atribut	Ket = Terkendali)	Ket = Tidak Terkendali)
P(Jenis Kelamin = Perempuan)	$\frac{4}{10} = 0,4$	$\frac{3}{10} = 0,3$
P(Umur = 50)	$\frac{1}{10} = 0,1$	$\frac{1}{10} = 0,1$
P(Desa = Banyubiru)	$\frac{1}{10} = 0,1$	$\frac{1}{10} = 0,1$
P(Status = Hipertensi)	$\frac{5}{10} = 0,5$	$\frac{3}{10} = 0,3$

4. Kalikan semua hasil sesuai dengan data testing yang akan dicari kelasnya

$P(X=C = \text{“terkendali”}) = 0,4 \times 0,1 \times 0,1 \times 0,5 = 0,002$

$P(X=C = \text{“tidak terkendali”}) = 0,3 \times 0,1 \times 0,1 \times 0,3 = 0,0009$

$P(C = \text{“terkendali”}|X) = 0,002 \times 0,7 = 0,0014$

$P(C = \text{“tidak terkendali”}|X) = 0,0009 \times 0,3 = 0,00027$

5. Bandingkan Hasil

Dari perhitungan probabilitas terkendali dan tidak terkendali pada langkah sebelumnya, dapat disimpulkan bahwa data pasien atas nama Siyamti, jenis kelamin perempuan, umur 50 th, desa Banyubiru, status hipertensi, masuk dalam kelas terkendali karena probabilitas terkendali (0,0014) lebih tinggi dibandingkan probabilitas tidak terkendali (0,00027).

Tabel 3. Confusion Matrix Naive Bayes

	True Terkendali	True Tak Terkendali
Pred. Terkendali	57	10
Pred. Tak Terkendali	23	10

Dari tabel tersebut dapat dilakukan perhitungan :

- Accuracy

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{57+10}{57+10+23+10} = \frac{67}{100} = 0,67 = 67\%$$

- Sensitivity

$$\text{Sensitivity} = \frac{TP}{TP+FN} = \frac{57}{57+10} = \frac{57}{67} = 0,8507 = 85,07\%$$

- Specificity

$$\text{Specificity} = \frac{TN}{TN+FP} = \frac{10}{10+23} = \frac{10}{33} = 0,3030 = 30,30\%$$

- PPV

$$\text{PPV} = \frac{TP}{TP+FP} = \frac{57}{57+23} = \frac{57}{80} = 0,7125 = 71,25\%$$

- NPV

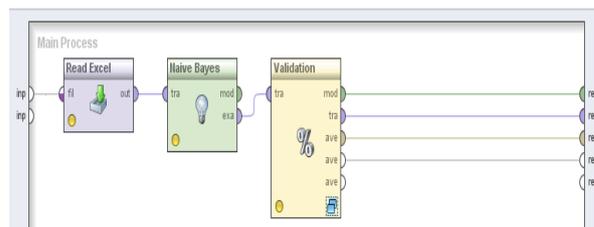
$$\text{NPV} = \frac{TN}{TN+FN} = \frac{10}{10+10} = \frac{10}{20} = 0,5 = 50\%$$

## 2) Pengujian dengan RapidMiner

Dataset yang disimpan dalam format excel dibuka menggunakan operator Read Excel.

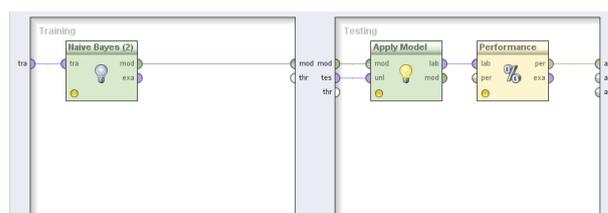
Kemudian keluarannya dihubungkan ke

Set Role terlebih dahulu untuk menentukan atribut label kemudian baru di hubungkan ke Validation (X-Validation).



Gambar 4. Process Naive Bayes

Untuk melatih dan menguji model di bagian training diisi operator Naive Bayes, pada testing diisi operator Apply Model dan Performance.



Gambar 5. Validasi Naive Bayes

Kemudian model dijalankan sehingga didapat hasil kinerja model seperti pada Gambar didapat akurasi 67,00%.

Multiclass Classification Performance			
Table View			
accuracy: 67.00% +/- 14.87% (mikro: 67.00%)			
	true Terkendali	true Tak Terkendali	class precision
pred. Terkendali	57	10	85.07%
pred. Tak Terkendali	23	10	30.30%
class recall	71.25%	50.00%	

Gambar 6. Akurasi Naive Bayes- Metode C4.5

## 1) Perhitungan Manual

- Menentukan nilai entropy

Entropy (keterangan)

$$\left(-\frac{7}{10} \log_2 \frac{7}{10}\right) + \left(-\frac{3}{10} \log_2 \frac{3}{10}\right) = 0,23$$

Entropy (jenis kelamin perempuan)

$$\left(-\frac{4}{7} \log_2 \frac{4}{7}\right) + \left(-\frac{3}{7} \log_2 \frac{3}{7}\right) = 0,25$$

Entropy (jenis kelamin laki-laki)

$$\left(-\frac{3}{3} \log_2 \frac{3}{3}\right) + \left(-\frac{0}{3} \log_2 \frac{0}{3}\right) = 0,52$$

Entropy (usia muda)

$$\left(-\frac{1}{1} \log_2 \frac{1}{1}\right) + \left(-\frac{0}{1} \log_2 \frac{0}{1}\right) = 0$$

Entropy (usia paruh baya)

$$\left(-\frac{2}{3}\log_2\frac{2}{3}\right) + \left(-\frac{1}{3}\log_2\frac{1}{3}\right) = 0,22$$

Entropy (usia tua)

$$\left(-\frac{4}{6}\log_2\frac{4}{6}\right) + \left(-\frac{2}{6}\log_2\frac{2}{6}\right) = 0,27$$

Entropy (status hipertensi)

$$\left(-\frac{5}{8}\log_2\frac{5}{8}\right) + \left(-\frac{3}{8}\log_2\frac{3}{8}\right) = 0,25$$

Entropy (normal)

$$\left(-\frac{2}{2}\log_2\frac{2}{2}\right) + \left(-\frac{0}{2}\log_2\frac{0}{2}\right) = 0,5$$

Entropy (Desa Banyubiru)

$$\left(-\frac{1}{2}\log_2\frac{1}{2}\right) + \left(-\frac{1}{2}\log_2\frac{1}{2}\right) = 0$$

Entropy (Desa Kebondowo)

$$\left(-\frac{1}{1}\log_2\frac{1}{1}\right) + \left(-\frac{0}{1}\log_2\frac{0}{1}\right) = 0$$

Entropy (Desa Sepakung)

$$\left(-\frac{2}{2}\log_2\frac{2}{2}\right) + \left(-\frac{0}{2}\log_2\frac{0}{2}\right) = 0,5$$

Entropy (Desa Tegaron)

$$\left(-\frac{1}{1}\log_2\frac{1}{1}\right) + \left(-\frac{0}{1}\log_2\frac{0}{1}\right) = 0$$

Entropy (Desa Kemambang)

$$\left(-\frac{1}{1}\log_2\frac{1}{1}\right) + \left(-\frac{0}{1}\log_2\frac{0}{1}\right) = 0$$

Entropy (Desa Wirogomo)

$$\left(-\frac{1}{1}\log_2\frac{1}{1}\right) + \left(-\frac{0}{1}\log_2\frac{0}{1}\right) = 0$$

Entropy (Desa Gedong)

$$\left(-\frac{0}{1}\log_2\frac{0}{1}\right) + \left(-\frac{1}{1}\log_2\frac{1}{1}\right) = 0$$

Entropy (Desa Rowoboni)

$$\left(-\frac{0}{1}\log_2\frac{0}{1}\right) + \left(-\frac{1}{1}\log_2\frac{1}{1}\right) = 0$$

Tabel 4. Perhitungan Entropy

Atribut	Kategori	Jumlah Kasus	Terkendali	Tidak Terkendali	Entropy
Keterangan		10	7	3	0,23
Jenis Kelamin	Perempuan	7	4	3	0,25
	Laki-laki	3	3	0	0,52
Usia	Muda	1	1	0	0
	Paruh Baya	3	2	1	0,22
	Tua	6	4	2	0,27
Status	Hipertensi	8	5	3	0,25
	Normal	2	2	0	0,5
Desa	Banyubiru	2	1	1	0
	Kebondowo	1	1	0	0
	Sepakung	2	2	0	0,5
	Tegaron	1	1	0	0
	Kemambang	1	1	0	0
	Wirogomo	1	1	0	0
	Gedong	1	0	1	0
	Rowoboni	1	0	1	0

- Hitung nilai gain

Gain (jenis kelamin)

$$= 0,23 - ((\frac{7}{10}) * 0,25) + ((\frac{3}{10}) * 0,5))$$

$$= 0,09$$

Gain (usia)

$$= 0,23 - ((\frac{1}{10}) * 0) + ((\frac{3}{10}) * 0,22) + ((\frac{6}{10}) * 0,27))$$

$$= 0,12$$

Gain (status)

$$= 0,23 - ((\frac{8}{10}) * 0,25) + ((\frac{2}{10}) * 0,5))$$

$$= 0,19$$

Gain (Desa)

$$= 0,23 - ((\frac{2}{10}) * 0) + ((\frac{1}{10}) * 0) + ((\frac{2}{10}) * 0,5) + ((\frac{1}{10}) * 0) + ((\frac{1}{10}) * 0) + ((\frac{1}{10}) * 0) + ((\frac{1}{10}) * 0) + ((\frac{1}{10}) * 0)$$

$$= 0,13$$

Kesimpulan :

- Dari nilai gain tersebut pilih nilai gain maksimal dari masing-masing

atribut. Nilai gain maksimal adalah gain (status) dengan nilai 0,19 sehingga atribut status menjadi node akar.

- Cabang yang terbentuk dari atribut status berisi kategori hipertensi dan normal.

- Pada atribut status dengan kategori normal, frekuensi terkendali sejumlah 1 (satu). Dapat disimpulkan bahwa nilai keterangan pada kategori normal adalah terkendali.

Tabel 5. Confusion Matrix C4.5

	True Terkendali	True Tak Terkendali
Pred. Terkendali	73	19
Pred. Tak Terkendali	7	1

Dari tabel tersebut dapat dilakukan perhitungan :

- Accuracy

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{73+1}{73+1+7+19} = \frac{74}{100} = 0,74 = 74\%$$

- Sensitivity

$$\text{Sensitivity} = \frac{TP}{TP+FN} = \frac{73}{73+19} = \frac{73}{92} = 0,7934 = 79,34\%$$

- Specificity

$$\text{Specificity} = \frac{TN}{TN+FP} = \frac{1}{1+7} = \frac{1}{8} = 0,125 = 12,5\%$$

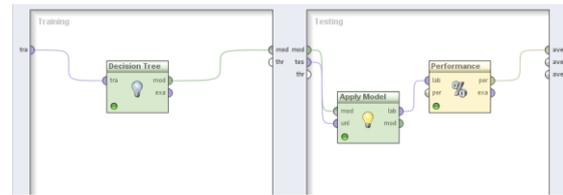
- PPV

$$\text{PPV} = \frac{TP}{TP+FP} = \frac{73}{73+7} = \frac{73}{80} = 0,9125 = 91,25\%$$

- NPV

$$\text{NPV} = \frac{TN}{TN+FN} = \frac{1}{1+19} = \frac{1}{20} = 0,05 = 5\%$$

diisi operator Apply Model dan Performance.



Gambar 8. Validasi C4.5

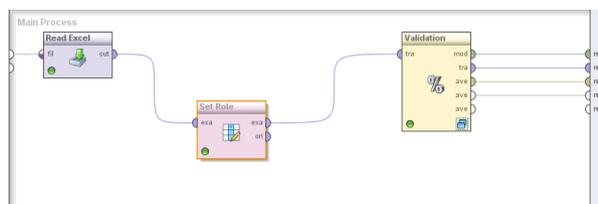
Kemudian model dijalankan sehingga didapat hasil kinerja model seperti pada gambar dan akurasinya 74%.

accuracy: 74.00% +/- 6.63% (mikro: 74.00%)			
	true Terkendali	true Tak Terkendali	class precision
pred. Terkendali	73	19	79.35%
pred. Tak Terkendali	7	1	12.50%
class recall	91.25%	5.00%	

Gambar 9. Akurasi C4.5

## 2) Pengujian dengan RapidMiner

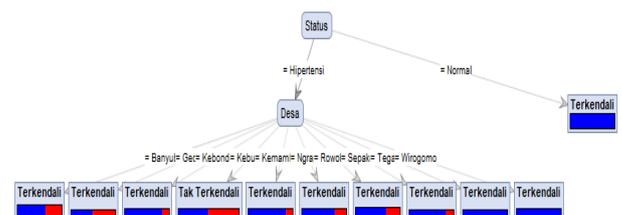
Dataset yang disimpan dalam format excel dibuka menggunakan operator Read Excel. Kemudian keluarannya dihubungkan ke Set Role terlebih dahulu untuk menentukan atribut label kemudian baru dihubungkan ke Validation (X-Validation). Susunan operator ditunjukkan pada gambar.



Gambar 7. Process C4.5

Untuk melatih dan menguji model di bagian training diisi operator Decision Tree yang merupakan turunan dari C4.5, pada testing

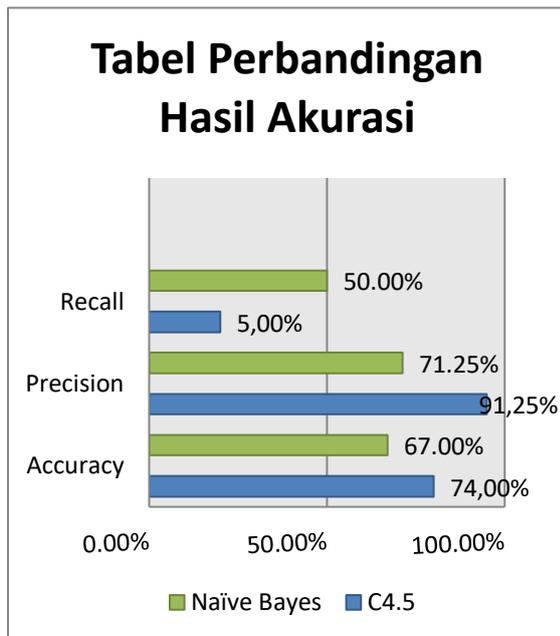
Berdasarkan hasil percobaan dalam RapidMiner dapat diketahui pohon faktornya untuk atribut (Status) sebagai node akarnya.



Gambar 10. Pohon Keputusan C4.5

### - Perbandingan Hasil Akurasi

Tabel di bawah ini memperlihatkan perbandingan akurasi dan AUC antara algoritma C4.5 dan Naive Bayes.



Gambar 11. Perbandingan Hasil Akurasi

Grafik diatas menunjukkan nilai akurasi C4.5 74,00%, Naive Bayes 67,00%. Dari ketiga metode tersebut akurasi tertinggi adalah metode C4.5 yaitu 74,00%. Hal ini membuktikan bahwa metode yang tepat untuk mendeteksi hipertensi terkendali dan tak terkendali adalah metode C4.5.

## SIMPULAN (PENUTUP)

### Kesimpulan

- Pada penelitian ini dilakukan pemodelan algoritma C4.5 dan Naive Bayes, dengan menggunakan data pasien hipertensi puskesmas Banyubiru. Fokus penelitian ini adalah mencari algoritma terbaik dalam menentukan prediksi hipertensi tak terkendali. Validasi model menggunakan X-Validation dan evaluasi model menggunakan confusion matrix.
- Hasil penelitian menunjukkan bahwa model C4.5 memiliki akurasi yang lebih baik dengan 74,00 % dibandingkan model Naive Bayes dengan akurasi 67,00 %.

- Dengan demikian terbukti bahwa untuk prediksi hipertensi tak terkendali, maka algoritma yang terbaik adalah C4.5 dengan nilai akurasi terbaik yaitu 74,00%.

### Saran

Berdasarkan kesimpulan di atas maka saran-saran yang dapat diberikan dalam penelitian ini yaitu:

1. Hasil penelitian ini diharapkan dapat dijadikan sebagai bahan masukan bagi puskesmas agar lebih meningkatkan kembali program perawatan maupun pengobatan pasien hipertensi.
2. Diharapkan pihak keluarga agar selalu mengoptimalkan dukungan yang diberikan kepada penderita hipertensi.
3. Responden hipertensi diharapkan mau dan berusaha untuk melakukan manajemen perawatan diri yang baik dengan menerapkan pola hidup sehat serta mengontrol hal-hal yang dapat memicu meningkatnya tekanan darah.

### UCAPAN TERIMAKASIH

Terimakasih segala puji bagi Allah SWT, Tuhan Yang Maha Esa yang telah memberikan rahmat, taufiq dan hidayah-Nya sehingga skripsi yang berjudul **“PERBANDINGAN ALGORITMA C4.5 DAN NAIVE BAYES DALAM MENDETEKSI HIPERTENSI TERKENDALI DAN TIDAK TERKENDALI DI PUSKESMAS BANYUBIRU”**

terselesaikan sesuai dengan rencana, tidak lupa terimakasih juga kepada berbagai pihak yang telah mendukung dalam proses penelitian ini. Oleh karena itu penulis mengucapkan terimakasih kepada:

1. Kepada Rektor Universitas Ngudi Waluyo Prof. Dr. Subyantoro M.Hum

2. Dekan dari Fakultas Komputer dan Pendidikan Teknik Informatika Iwan Setiawan Wibisono S.T., M.Kom.
  3. Ketua dari Program Studi Teknik Informatika Universitas Ngudi Waluyo Sri Mujiyono S.Kom, M.Kom.
  4. Dosen Pembimbing Skripsi Penulis Iwan Setiawan wibisono S.Kom, M.Kom.
  5. Dosen-dosen pengampu di Fakultas Komputer dan Pendidikan Universitas Ngudi Waluyo yang telah memberikan ilmu dan pengetahuannya masing-masing, sehingga penulis dapat mengimplementasikan ilmu yang telah disampaikan.
  6. Seluruh Staff di puskesmas Banyubiru yang telah memberikan data-data untuk keperluan penyusunan skripsi sampai dengan sistem aplikasi ini terbentuk.
- Henny, Chandra, 2021, Data Preprocessing adalah/ pengertian, Tahapan Kerja dan Manfaatnya,  
<https://id.linkedin.com/pulse/data-preprocessing-adalah-pengertian-tahapankerja-dan-chandra-henny>, diakses tgl 2 maret 2022.
- Nugroho, Kuncahyo Setyo, 2019, Confusion Matrix untuk Evaluasi Model pada Supervised Learning,  
<https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-padaunsupervised-machine-learning-bc4b1ae9ae3f>, diakses tgl 2 maret 2022.

#### DAFTAR PUSTAKA

- Intermedia, Beon, 2020, Data Mining: Definisi, Fungsi, Metode dan Penerapannya.
- Vidiastanta, I.G., Hidayat, N., dan Dewi, R.K., 2020, Komparasi Metode K-Nearest Neighbor (K-NN) Dengan Support Vector Machine (SVM) Untuk Klasifikasi Status Hipertensi.
- Vidiastanta, I.G., Hidayat, N., dan Dewi, R.K., 2020, Komparasi Metode K-Nearest Neighbor (K-NN) Dengan Support Vector Machine (SVM) Untuk Klasifikasi Status Kualitas Air, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, No. 1, Vol .4, 312-319.  
<https://puspitasulistyorini.blogspot.com/2020/12/uptd-puskesmas-banyubiru-kabupaten.html>
- Intermedia, Beon, 2020, Data Mining: Definisi, Fungsi, Metode dan Penerapannya,  
<https://ww.jagoanhosting.com/blog/apa-itu-data-mining/>, diakses tgl 2 maret 2022.