



Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes

*Agus Fajar Riany¹, Gusmelia Testiana²

^{1,2}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, UIN Raden Fatah Palembang

Jl. Panca Usaha No.2085, 5 Ulu, Kecamatan Seberang Ulu I, Kota Palembang

Email: ¹1930803043@radenfatah.ac.id, ²gusmeliatestiana_uin@radenfatah.ac.id

ABSTRACT

Stroke is a disturbance of brain function, both local and general, that occurs suddenly, progressively, and rapidly due to non-traumatic brain blood circulation disorders that lasts more than 24 hours or ends in death. Stroke is also one of the deadliest diseases in Indonesia. In this study, stroke data was used to explore new information or knowledge in it. The process of extracting new information from a set of data is known as data mining. Therefore, this research aims to classify data related to stroke using the Naïve Bayes algorithm to find out whether the patient has a stroke or not. There are 10 attributes that are included in the causes of stroke, among others, gender, age, history of hypertension, history of heart disease, marital status, type of work, type of residence, glucose level, body mass index and smoking status. The results showed that classification with the Naïve Bayes algorithm can be applied in classifying stroke data resulting in an accuracy value of 92.48% in the Good Classification category.

Keywords : data mining; classification; naïve bayes

ABSTRAK

Stroke merupakan suatu gangguan fungsi otak baik lokal maupun menyeluruh yang terjadi secara mendadak progresif, dan cepat akibat gangguan peredaran darah otak non traumatik yang berlangsung lebih dari 24 jam atau berakhir dengan kematian. Stroke juga merupakan salah satu jenis penyakit yang paling mematikan di Indonesia. Pada Penelitian ini data penyakit stroke dimanfaatkan untuk digali informasi atau pengetahuan baru didalamnya Proses penggalian informasi baru dari sekumpulan data dikenal dengan data mining. Maka dari itu, dengan adanya penelitian ini bertujuan untuk mengklasifikasi data terkait penyakit stroke menggunakan algoritma *Naïve Bayes* untuk mengetahui apakah pasien tersebut menderita penyakit stroke atau tidak. Terdapat 10 atribut yang termasuk dalam penyebab terjadinya stroke antara lain, jenis kelamin, usia, riwayat penyakit hipertensi, riwayat penyakit jantung, status pernikahan, jenis pekerjaan, tipe tempat tinggal, kadar glukosa, index masa tubuh dan status merokok. Hasil penelitian menunjukkan bahwa klasifikasi dengan algoritma Naïve Bayes dapat diterapkan dalam mengklasifikasi data penyakit stroke menghasilkan nilai akurasi sebesar 92,48% dalam kategori *Good Classification*.

Kata kunci : data mining; klasifikasi; *naïve bayes*

1. PENDAHULUAN

Stroke merupakan salah satu penyakit berbahaya dengan tingkat kematian yang tinggi di Indonesia. Pada saat ini, stroke semakin menjadi masalah serius yang harus dihadapi hampir diseluruh dunia. Hal tersebut dikarenakan serangan stroke yang mendadak dapat mengakibatkan kematian, kecacatan fisik, dan mental baik pada usia produktif maupun usia lanjut (Pambudi et al., 2022).

Penyakit stroke merupakan jenis penyakit yang mematikan dimana masuk kedalam 10 jenis penyakit yang paling mematikan di Indonesia. Berdasarkan Laporan Hasil Riskesdas 2018 menunjukkan prevalensi stroke nasional yang masih cukup tinggi yaitu 10.9%. Provinsi di Indonesia dengan prevalen stroke tertinggi adalah Provinsi Kalimantan Timur yaitu 14.7%. Sedangkan Provinsi di Indonesia dengan prevalensi stroke terendah adalah Provinsi Papua yaitu 4.1% (Kemenkes RI, 2018).

Kementerian Kesehatan Republik Indonesia mendefinisikan penyakit stroke sebagai kerusakan pada otak yang muncul mendadak, progresif, dan cepat akibat gangguan peredaran darah otak

non traumatik (Kesehatan, 2019). Gangguan tersebut secara mendadak menimbulkan gejala antara lain kelumpuhan pada wajah atau anggota badan, berbicara menjadi tidak jelas, perubahan kesadaran, gangguan penglihatan dan lain-lain. Faktor-faktor yang menyebabkan stroke antara lain; faktor kesehatan (hipertensi, kolesterol, obesitas dan jantung), faktor gaya hidup (merokok, kurang olahraga, konsumsi obat-obatan terlarang dan kecanduan alkohol), faktor keturunan, dan bertambahnya usia (Haryadi et al., 2021). Masalah penyakit stroke di Indonesia memerlukan perhatian yang serius karena jumlah kasus yang terus meningkat dan mempunyai angka kematian yang tinggi.

Alternatif solusi yang dilakukan pada penelitian ini adalah memanfaatkan data penyakit stroke untuk digali informasi atau pengetahuan baru didalamnya, sehingga dapat dimanfaatkan untuk membangun model baru dalam mengklasifikasi pasien untuk mengetahui apakah pasien tersebut menderita penyakit stroke atau tidak. Proses penggalian informasi baru dari sekumpulan data dikenal dengan data mining (Written et al., 2011).

Metode data mining yang digunakan pada penelitian ini yaitu klasifikasi. Metode klasifikasi dalam metode dalam data mining dengan mengklasifikasi kelas target secara akurat (Muktamar et al., 2015). Tujuan dari metode klasifikasi ini adalah untuk mengklasifikasikan nilai suatu variabel yang tidak diketahui dari variabel lain yang telah ditentukan sebelumnya (Nugraha et al., 2022).

Terdapat berbagai algoritma yang dapat digunakan pada klasifikasi seperti *Decision Tree*, *Naïve Bayes*, *K-Nearest Neighbor*, dan lainnya. penelitian ini menggunakan menggunakan algoritma *Naïve Bayes* untuk mengklasifikasi penyakit stroke. Algoritma *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik sederhana yang berdasar pada teorema Bayesian dengan asumsi independensi yang kuat (Prasetyo, 2014). *Naïve Bayes* merupakan algoritma klasifikasi sederhana dimana setiap atribut bersifat independen dan dapat mempengaruhi keputusan akhir (Testiana, 2018). Algoritma *Naïve Bayes* dapat digunakan untuk memprediksi probabilitas milik kelas. Keunggulan dari algoritma *Naïve Bayes* yaitu algoritma yang sederhana

dan mempunyai nilai performa yang baik (Siddik et al., 2020).

Terdapat penelitian terdahulu yang berkaitan dengan penerapan algoritma *Naïve Bayes* dan pengklasifikasian penyakit stroke menggunakan data mining. Penelitian untuk mengklasifikasi dan memprediksi orang-orang tentang apakah orang tersebut akan berpotensi terkena penyakit stroke atau tidak menunjukkan performa akurasi algoritma *Naïve Bayes* sebesar 93,94% (Bugis, 2022). Penelitian selanjutnya untuk mengklasifikasi penyakit diabetes mellitus menunjukkan performa akurasi algoritma *Naïve Bayes* sebesar 90,20%.

Berdasarkan uraian latar belakang permasalahan diatas, penelitian ini bertujuan untuk mengklasifikasi pasien apakah pasien tersebut terkena penyakit stroke atau tidak dengan menggunakan algoritma *Naïve Bayes* dengan perhitungan probabilitas dari setiap atributnya. Tujuan selanjutnya adalah untuk mengetahui akurasi algoritma *Naïve Bayes*.

2. METODE

Metode penelitian adalah metode yang digunakan untuk memperoleh data untuk menghasilkan suatu tujuan.

Metode penelitian kuantitatif adalah metode penelitian yang menggunakan filsafat positivisme, untuk meneliti populasi atau sampel tertentu dengan pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif dengan tujuan untuk menguji hipotesis yang telah ditetapkan (Sugiyono, 2017). Tujuan dari metode penelitian kuantitatif adalah untuk mengidentifikasi masalah yang ditetapkan sebagai tujuan penelitian. Pendekatan deskriptif merupakan suatu cara untuk mengetahui eksistensi dari variabel bebas. Dalam penelitian ini menerapkan metode kuantitatif yang memanfaatkan data mining dan perhitungan algoritma *Naïve Bayes*.

2.1. Metode Pengumpulan Data

Dalam mendukung penelitian ini dilakukan pengumpulan data. Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang akan digunakan menggunakan metode pengumpulan data sekunder. Data Sekunder adalah data yang diperoleh berdasarkan penelitian sebelumnya, artinya peneliti tidak harus mengambil datanya langsung ke lapangan melainkan sudah dikumpulkan dan diolah oleh peneliti lain (Suryani & Hendryadi,

2018). Data dari penelitian ini diambil dari situs kaggle https://www.kaggle.com/datasets/aniket_sjadhav/brain-stroke. Jumlah sampel yang digunakan dalam penelitian ini adalah sebesar 4.981 data. Data ini terdiri dari 11 atribut dengan rincian 10 sebagai atribut dan 1 sebagai label atau target.

2.2 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan klasifikasi yang mempresentasikan setiap kelas objek berdasarkan kesimpulan atau rekapitulasi probabilistik dan menemukan kemungkinan besar kelas yang sesuai untuk tiap objek-objek yang akan ditentukan kelasnya dari objek-objek uji yang ada berdasarkan atribut-atribut atau variabel yang telah diketahui nilai-nilainya (Arhami & Nasir, 2020). Algoritma Naïve Bayes dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Algoritma *Naïve Bayes* sangat cocok untuk melakukan klasifikasi pada dataset bertipe nominal (Suntoro, 2019). Teorema Bayes memiliki bentuk umum sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

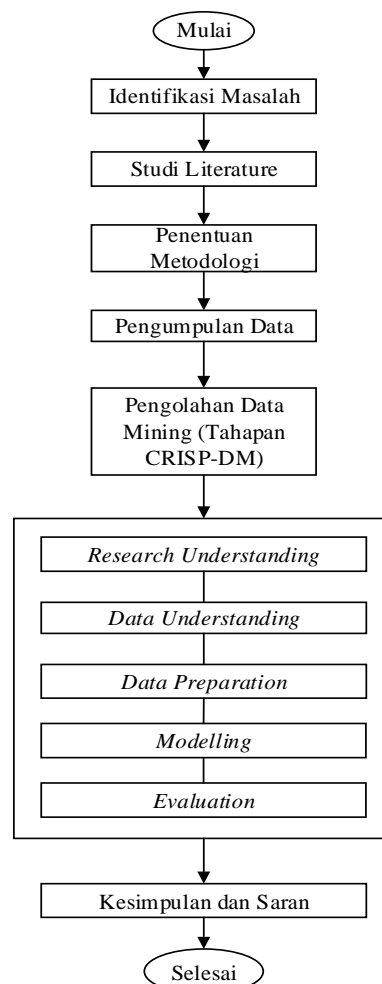
X adalah Data dengan *class* yang belum diketahui, H adalah Hipotesis data X merupakan suatu *class* spesifik, $P(H|X)$ adalah Probabilitas hipotesis H berdasar kondisi X (*posteriori probability*), $P(H)$ adalah Probabilitas hipotesis H (*pror probability*) dan $P(H|X)$ adalah Probabilitas X berdasar kondisi pada hipotesis H, $P(X)$ adalah Probabilitas dari X.

Terdapat penelitian terdahulu yang berkaitan dengan penerapan algoritma *Naïve Bayes* dan pengklasifikasian penyakit stroke menggunakan data mining. Penelitian yang dilakukan (Doni et al., 2021) untuk mengklasifikasikan tingkat kemungkinan hidup pasien yang telah di diagnosis menderita penyakit Hepatocellular Carcinoma menunjukkan performa akurasi yang dihasilkan sebesar 70,30%. Kemudian penelitian yang dilakukan (Ridwan, 2020) dalam mengklasifikasi data penyakit diabetes mellitus menunjukkan performa akurasi yang dihasilkan sebesar 90,20%.

2.3 Tahapan Penelitian

Tahapan penelitian adalah langkah-langkah yang akan dilakukan pada penelitian ini dan digambarkan dalam bentuk diagram alir. Diagram alir

tersebut menggambarkan tentang tahapan proses penelitian yang akan dilakukan sekaligus juga menggambarkan penelitian secara menyeluruh. Gambar 1 berikut ini merupakan tahapan penelitian.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

Pengolahan data mining merupakan hasil utama dalam penelitian ini yang temuan-temuannya akan dibahas dibagian hasil dan pembahasan ini. Proses pengolahan data mining

dilakukan dengan tahap sesuai *Cross Industry Standard Process for Data Mining* (CRISP-DM). Berikut merupakan uraian setiap tahapan pengolahan data mining yang dilakukan.

3.1 Pemahaman Penelitian (*Research Understanding*)

Proses data mining diawali dengan melakukan pemahaman penelitian (*research understanding*) yang bertujuan untuk memahami permasalahan dan menetapkan tujuan dilakukannya data mining. Sebagaimana permasalahan yang telah diuraikan dalam latar belakang, maka penerapan data mining pada penelitian ini bertujuan untuk mendapatkan hasil klasifikasi penyakit stroke untuk mengetahui apakah pasien tersebut menderita stroke atau tidak menggunakan algoritma Naïve Bayes menggunakan perhitungan probabilitas dan untuk mengetahui akurasi dari algoritma tersebut.

3.2 Pemahaman Data (*Data Understanding*)

Tahap setelah pemahaman penelitian adalah pemahaman data (*data understanding*). Tahap ini bertujuan untuk memahami karakteristik data yang akan diolah. Adapun data yang akan digunakan dalam proses data mining

adalah sampel data dari situs kaggle sebanyak 4981 data. Atribut dataset awal terdiri dari 11 atribut dengan rincian 10 sebagai atribut dan satu sebagai label atau target. Atribut-atribut tersebut yaitu *gender*, *age*, *hypertension*, *heart_diseasi*, *ever_married*, *work_type*, *residence_type*, *avg_glucose_level*, BMI, *smoking_status*, dan *stroke*. Tabel 1 berikut ini merupakan dataset awal.

Tabel 1. Dataset Awal

Gender	Age	Hypertension	...	Stroke
Male	67	0	...	1
Male	80	0	...	1
Female	49	0	...	1
Female	79	1	...	1
Male	81	0	...	1
Female	78	1	...	0
Male	33	0	...	0
Female	51	0	...	0
...
Female	80	1	...	0

3.3 Persiapan Data (*Data Preparation*)

Pada tahap ini dilakukan proses transformasi pada dataset awal. Transformasi dilakukan untuk mengkonversi nilai beberapa atribut yang bernilai numerik, menjadi nominal. Berikut ini beberapa atribut yang bernilai numerik yang dikonversi menjadi nominal, atribut tersebut antara lain; *age*, *hypertension*, *heart_disease*, *avg_glucose_level*, BMI dan *stroke*. Sedangkan yang bernilai nominal tidak dilakukan konversi, atribut yang tidak

dikonversi antara lain; *gender*, *ever_married*, *work_type*, *residence_type* dan *smoking_status*.

Berikut ini Tabel 2 proses transformasi data pada setiap atribut.

Tabel 2. Transformasi Atribut

No	Atribut	Sub Atribut	Konversi
1	Age	0 - 5	Balita
		5 - 11	Kanak-kanak
		12 - 16	Remaja Awal
		17 - 25	Remaja Akhir
		26 - 35	Dewasa Awal
		36 - 45	Dewasa Akhir
		46 - 55	Lansia Awal
		56 - 65	Lansia Akhir
		>65	Manula
		2	Hypertension
1	Yes		
3	Heart Disease	0	No
		1	Yes
4	Avg_Glucose_Level	> 126	Diabetes
		100 - 125	Prediabetes
		70 - 99	Normal
		<18,5	Underweight
		18,5 - 24,9	Normal
5	BMI	25 - 29,9	Overweight
		30 - 34,9	Obese
		>35	Extremely Obese
		0	No
		1	Yes

Setelah proses transformasi selesai maka dataset telah siap untuk diolah ke tahap selanjutnya. Tabel 3 berikut merupakan dataset yang telah melalui transformasi di beberapa atributnya.

Tabel 3. Dataset Hasil *Preparation Data*

Gender	Age	Hypertension	...	Stroke
--------	-----	--------------	-----	--------

Male	Manula	No	...	Yes
Male	Manula	No	...	Yes
Female	Lansia Awal	No	...	Yes
Female	Manula	Yes	...	Yes
Male	Manula	No	...	Yes
Female	Manula	Yes	...	No
Male	Dewasa Awal	No	...	No
Female	Lansia Awal	No	...	No
...
Female	Manula	Yes	...	No

3.4 Pemodelan (Modeling)

Tahap ini merupakan tahapan implementasi dari teknik dan algoritma yang dipilih. *Tools* yang digunakan untuk melakukan perhitungan manual menggunakan *Microsoft Excel*, sedangkan untuk pengujiannya juga menggunakan *Software Rapid Miner*. Pada penelitian ini akan diterapkan klasifikasi dengan menggunakan algoritma yang telah dipilih yaitu *Naïve Bayes*. Jumlah sampel pasien penyakit stroke yang dijadikan sampel untuk proses perhitungan algoritma *Naïve Bayes* untuk mengetahui pasien tersebut terklasifikasi menderita penyakit stroke atau tidak yang dilihat berdasarkan probabilitas dari setiap kelas dan probabilitas setiap kejadian perkelas dan keakurasian yang dihasilkan.

Sebelum dilakukan pemodelan data, dataset awal akan dibagi menjadi dua yaitu data latihan (*training*) dan data

uji (*testing*). Proses pembagian data dilakukan dengan *Split Train Test* yang membagi data menjadi data *training* dan data *testing*. Adapun persentase pembagian data yaitu sebesar 80% untuk data *training* dan 20% untuk data *testing*. Pembagian dataset dengan perbandingan 80:20 dapat memberikan performa model yang terbaik (Bugis, 2022). Sehingga pada penelitian ini dengan dataset sejumlah 4.981, terdapat 3.984 data *training* dan 997 data *testing*. Tahap-tahap pemodelan data menggunakan algoritma *Naïve Bayes* adalah sebagai berikut:

a. Menghitung Probabilitas Setiap Kelas

Dari data yang digunakan terdapat 2 *class* yaitu *Yes* dan *No*. *Class Yes* (penderita stroke) berjumlah 198 data, sedangkan jumlah *class No* (Tidak Menderita Stroke) berjumlah 3.786 data. Untuk menghitung probabilitas pada setiap kelasnya dilakukan perhitungan dengan membagi jumlah setiap kelas dibagi dengan seluruh data. Tabel 4 berikut merupakan hasil perhitungan probabilitas dari setiap kelas.

Tabel 4. Probabilitas Setiap Kelas

Kelas	Probabilitas Kelas P(H)
<i>Yes</i>	0,050
<i>No</i>	0,950

b. Menghitung Probabilitas Setiap Kejadian Perkelas

Untuk menghitung nilai probabilitas setiap atribut dengan cara menghitung jumlah kejadian atau atribut pada suatu kelas dibagi dengan kelas yang ada (Nurlia et al., 2021). Tabel 5 berikut merupakan hasil nilai Probabilitas setiap atribut.

Tabel 5. Probabilitas Setiap Kejadian Perkelas

Atribut	Sub Atribut	Probabilitas P(X H)	
		YES	NO
<i>Gender</i>	<i>Male</i>	0.404	0.415
	<i>Female</i>	0.596	0.585
	Balita	0.005	0.067
	Kanak-kanak	0	0.048
	Remaja Awal	0.005	0.049
<i>Age</i>	Remaja Akhir	0	0.106
	Dewasa Awal	0	0.124
	Lansia Awal	0.121	0.152
	Lansia Akhir	0.227	0.142
	Manula	0.641	0.312
<i>Hypertension</i>	<i>Yes</i>	0.273	0.084
	<i>No</i>	0.727	0.916
<i>Heart_Disease</i>	<i>Yes</i>	0.192	0.050
	<i>No</i>	0.808	0.950
<i>Ever_Married</i>	<i>Yes</i>	0.879	0.642
	<i>No</i>	0.121	0.358
<i>Work_Type</i>	<i>Children</i>	0.010	0.144
	<i>Govt_Job</i>	0.126	0.127
	<i>Private</i>	0.586	0.571
	<i>Self-Employed</i>	0.278	0.157
<i>Residence_Type</i>	<i>Rural</i>	0.449	0.498
	<i>Urban</i>	0.551	0.502
<i>Avg_Glucose_Level</i>	Normal	0.460	0.623
	Prediabetes	0.157	0.195
	Diabetes	0.384	0.182
BMI	<i>Uderweight</i>	0	0.072
	<i>Normal</i>	0.126	0.254
	<i>Overweight</i>	0.697	0.503

<i>Smoking_ Stratus</i>	<i>Extremely Obese</i>	0.177	0.171
	<i>Formerly Smoked</i>	0.273	0.167
	<i>Never Smoked</i>	0.379	0.373
	<i>Smokes</i>	0.177	0.151
	<i>Unknown</i>	0.172	0.310

c. Mengalikan Semua Variabel Perkelas

Dari nilai probabilitas yang telah diperoleh dapat digunakan untuk memprediksi atau menentukan label data *testing* dengan cara mengalikan setiap atribut perkelas.

Tabel 6. Contoh Data *Testing*

Gender	Age	Hypertensi on	...	Stroke
Male	Manula	No	...	?

Dari salah satu contoh data *Testing* pada Tabel 6, kemudian dilakukan perhitungan manual menggunakan *Microsoft Excel*. Hasil perhitungan data *testing* dapat dilihat pada Tabel 7 sebagai berikut:

Tabel 7. Perhitungan Data *Testing* Secara Manual

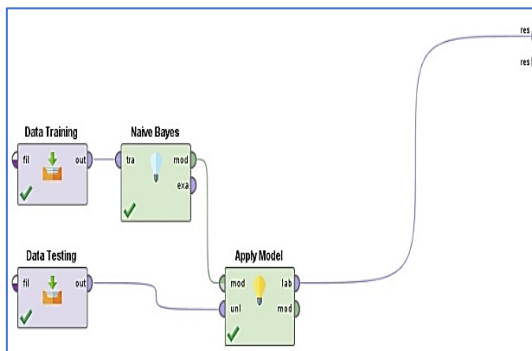
$P(X Ci)$	$P(\text{Gender} = \text{"Male"} \text{Hasil "Yes"})$	0,404
	$P(\text{Gender} = \text{"Male"} \text{Hasil "No"})$	0,4155
	$P(\text{Age} = \text{"Manula"} \text{Hasil "Yes"})$	0,6414
	$P(\text{Age} = \text{"Manula"} \text{Hasil "No"})$	0,3122
	$P(\text{Hypertension} = \text{"No"} \text{"Yes"})$	0,7273
	$P(\text{Hypertension} = \text{"No"} \text{"No"})$	0,9163

$P(\text{Heart_Disease} = \text{"Yes"} \text{"Yes"})$	0,1919	
$P(\text{Heart_Disease} = \text{"Yes"} \text{"No"})$	0,0499	
$P(\text{ever_married} = \text{"Yes"} \text{Hasil "Yes"})$	0,8788	
$P(\text{ever_married} = \text{"Yes"} \text{Hasil "No"})$	0,6424	
$P(\text{work_type} = \text{"Private"} \text{Hasil "Yes"})$	0,5859	
$P(\text{work_type} = \text{"Private"} \text{Hasil "No"})$	0,5713	
$P(\text{Residence_type} = \text{"Urban"} \text{"Yes"})$	0,5505	
$P(\text{Residence_type} = \text{"Urban"} \text{"No"})$	0,5024	
$P(\text{Glucose} = \text{"Diabetes"} \text{"Yes"})$	0,3838	
$P(\text{Glucose} = \text{"Diabetes"} \text{"No"})$	0,1817	
$P(\text{BMI} = \text{"Overweight"} \text{Hasil "Yes"})$	0,697	
$P(\text{BMI} = \text{"Overweight"} \text{Hasil "No"})$	0,5034	
$P(\text{smoking_status} = \text{"Smokes"} \text{Hasil "Yes"})$	0,1768	
$P(\text{smoking_status} = \text{"Smokes"} \text{Hasil "No"})$	0,1508	
$P(X \text{Hasil "Ya"})$	0,0005	
$P(X \text{Hasil "No"})$	2E-05	
$\frac{P(X Ci)}{*P(Ci)}$	$\frac{P(X (\text{Hasil "Ya"})) * P(\text{Ya} \text{Total})}{P(X (\text{Hasil "Tidak"})) * P(\text{Tidak} \text{Total})}$	2E-05
	Nilai Maksimal	1E-05
		2E-05

Berdasarkan hasil perhitungan manual setiap kelas, kemudian dibandingkan mana nilai probabilitas terbesar dari hasil klasifikasinya. Dimana hasil perhitungan dari contoh data *testing* pada Tabel 6. Kelas Yes

memiliki nilai probabilitas terbesar yaitu 2,409 sehingga data *testing* pada Tabel 6 pasien tersebut menderita penyakit stroke.

Setelah didapatkan hasil pengujian menggunakan excel, selanjutnya pengujian menggunakan *software* Rapid Miner. Berikut Gambar 2 hasil klasifikasi dari setiap kelompok penyakit stroke.



Gambar 2. Pemodelan di Rapid Miner

Gambar 2 merupakan model yang terbentuk pada *tools* Rapid Miner yang digunakan untuk melakukan pengujian pada data uji (*testing*). Perbandingan data pada pengujian ini yaitu 80% data *training* dan 20% data *testing*.

Row No.	Stroke	prediction\$...	confidence\$...	confidence\$...	Gender	Age	Hypertension
1	?	No	0.132	0.868	Female	Lansia Awal	No
2	?	No	0.010	0.990	Female	Manula	No
3	?	No	0.285	0.715	Female	Manula	Yes
4	?	No	0.154	0.836	Male	Manula	No
5	?	No	0.048	0.952	Male	Manula	No
6	?	Yes	0.627	0.373	Male	Manula	No
7	?	No	0.344	0.656	Female	Lansia Awal	Yes
8	?	No	0.125	0.875	Female	Manula	Yes
9	?	No	0.442	0.558	Female	Manula	Yes
10	?	No	0.043	0.957	Male	Manula	Yes
11	?	No	0.202	0.798	Female	Manula	No
12	?	No	0.023	0.977	Male	Manula	No
13	?	No	0.029	0.971	Male	Lansia Akhir	No
14	?	No	0.173	0.827	Male	Lansia Akhir	No

Gambar 3. Hasil Klasifikasi Setiap Kelompok Penyakit Stroke

Gambar 3 merupakan hasil klasifikasi setiap kelompok penyakit stoke yang telah dilakukan pengklasifikasian dengan menggunakan *tools* Rapid Miner. Hasil ini menunjukkan prediksi kelompok kelas setiap data pada data uji (*testing*). Contohnya pada data baris pertama, hasil pengujian menunjukkan bahwa pasien pada data baris pertama diprediksi tidak menderita penyakit stroke, begitupun hasil seterusnya.

3.5 Evaluasi (*Evaluation*)

Tahap evaluasi bertujuan untuk mengetahui performa akurasi dari model yang telah dibangun. Evaluasi dilakukan pada Rapid Miner dengan menambahkan fungsi *confusion matrix* dimana perbandingan data yang digunakan pada pengujian ini yaitu 80:20, dimana 80% data *training* dan 20% data *testing*. Tabel

8 berikut merupakan hasil evaluasi yang dilakukan.

Tabel 8. Evaluasi Akurasi Algoritma

Naïve Bayes

Accuracy: 92,48%

	<i>True Yes</i>	<i>True No</i>	<i>Class Precision</i>
<i>Pred. Yes</i>	5	30	14,29%
<i>Pred. No</i>	45	917	95,32%
<i>Class Recall</i>	10,00%	96,83%	

Hasil evaluasi menunjukkan bahwa akurasi dari algoritma Naïve Bayes sebesar 92,48%. Artinya akurasi algoritma Naïve Bayes dalam menentukan atau mengklasifikasi pasien penyakit stroke pada kategori *Good Classification*.

4. KESIMPULAN

Berdasarkan hasil dari serangkaian tahapan penelitian yang telah dilakukan, dapat disimpulkan bahwa proses data mining dengan teknik klasifikasi menggunakan algoritma Naïve Bayes dapat diterapkan dalam pengelompokan pasien untuk mengetahui apakah pasien tersebut menderita penyakit stroke atau tidak. Dimana berdasarkan proses pengujian yang telah dilakukan dengan menggunakan data *training* dan data *testing* dengan pembagian data 80:20, hasilnya menunjukkan bahwa pada data

testing, data yang ter-klasifikasi menderita penyakit stroke sebanyak 35 data, sedangkan data yang ter-klasifikasi tidak menderita penyakit stroke sebanyak 962 data. Dari pengujian ini menghasilkan tingkat akurasi sebesar 92,48% yang berada dalam kategori *Good Classification*.

Adapun saran bagi penelitian selanjutnya agar dapat dikembangkan dengan melakukan perbandingan beberapa algoritma klasifikasi dalam menganalisis dan mengelompokkan penyakit stroke. Sehingga dapat memberikan hasil yang bervariasi dan dapat ditemukan atribut yang memiliki kontribusi terbesar dalam memprediksi dan mengklasifikasi penyakit stroke serta dapat ditemukan model dengan algoritma mana yang memberikan performa paling baik dalam mengklasifikasi data penyakit stroke.

DAFTAR PUSTAKA

- Arhami, M., & Nasir, M. (2020). *Data Mining: Algoritma dan Implementasi*. Andi Offset.
- Bugis, H. (2022). Metode Naïve Bayes Untuk Memprediksi Penyakit Stroke. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan)*
<https://doi.org/10.47970/siskom-kb.v6i1.317>

Doni, B. T. R., Susanti, S., & Mubarak,

- A. (2021). Penerapan Data Mining Untuk Klasifikasi Penyakit Hepatocellular Carcinoma Menggunakan Algoritma Naïve Bayes. *Jurnal Responsif: Riset Sains Dan Informatika*, 3(1), 12–19. <https://doi.org/10.51977/jti.v3i1.403>
- Haryadi, D., Marini Umi Atmaja, D., Rahman Hakim, A., & Suwaryo, N. (2021). Identifikasi Tingkat Resiko Penyakit Stroke Menggunakan Algoritma Regresi Linear Berganda. *Deny Haryadi, SNTEM*, 1(November), 1198–1207. <https://doi.org/10.53026/sntem.v1i2.589>
- Kemenkes RI. (2018). Hasil Riset Kesehatan Dasar Tahun 2018. *Kementerian Kesehatan RI*, 53(9), 1689–1699.
- Kesehatan, K. (2019). Kementerian Kesehatan Republik Indonesia. *Kementerian Kesehatan RI*, 1(1), 1.
- Muktamar, B. A., Setiawan, N. A., & Adji, T. B. (2015). Analisis Perbandingan Tingkat Akurasi Algoritma Naive Bayes Classifier dengan Correlated-Naive Bayes Classifier. *Seminar Nasional Teknologi Informasi Dan Multimedia 2015*, 49–54.
- Nugraha, F. F., Sunandar, I., & Juliane, C. (2022). Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Teknologi Informatika Dan Sistem Informasi*, 9(4), 2862–2869. <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/2518/1006>
- Nurlia, E., Jajuli, M., & Purnamasari, I. (2021). Penerapan Naïve Bayes Untuk Klasifikasi Tingkat Risiko Diagnosis Gigi Di Uptd Puskesmas Cingambul. *JIKO (Jurnal Informatika Dan Komputer)*, 4(2), 127–132. <https://doi.org/10.33387/jiko.v4i2.3190>
- Pambudi, R. E., Sriyanto, S., & Firmansyah, F. (2022). Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.45. *Teknika*, 16(2), 221 – 226–221 – 226. <https://doi.org/10.5281/zenodo.7535865>
- Prasetyo, E. (2014). *Data Mining: Konsep dan Aplikasi Menggunakan Mtlab*. Andi Offset.
- Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 15–21. <https://doi.org/10.47970/siskom-kb.v4i1.169>
- Siddik, M., Hendri, H., Putri, R. N., Desnelita, Y., & Gustientiedina, G. (2020). Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naïve Bayes. *INTECOMS: Journal of Information Technology and Computer Science*, 3(2), 162–166. <https://doi.org/10.31539/intecom.s.v3i2.1654>
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Alfabeta.

- Suntoro, J. (2019). *Data Mining: Algoritma dan Implementasi dengan Pemograman PHP*. PT ELex Media Komputindo.
- Suryani, & Hendryadi. (2018). *Metode Riset Kuantitatif: Teori dan Aplikasi Pada Penelitian Bidang Manajemen dan Ekonomi Islam* (2nd ed.). Prenadamedia Group.
- Testiana, G. (2018). Perancangan model prediksi kelulusan mahasiswa tepat waktu pada UIN Raden Fatah. *JUSIFO (Jurnal Sistem Informasi)*, 4(1), 49–62. <http://jurnal.radenfatah.ac.id/index.php/jusifo/article/view/1932>
- Written, I. H., Frank, E., & A., M. (2011). *Data Mining: practical machine learning tools and techniques*. In Complementary Literature None.