



Penerapan Data Mining Pada Klasifikasi Gangguan Jiwa Menggunakan Algoritma C5.0 Di RSJ. Mahoni Kota Medan

Najla Anthira¹, Suendri²

^{1,*2}Program Studi Sistem Informasi, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

*Email Penulis Korespondensi: najlaanthira2002@gmail.com¹

Abstrak

Dalam dunia kesehatan, khususnya dalam bidang kesehatan mental, teknologi telah menjadi alat yang sangat berguna untuk mengeksplorasi berbagai metode dalam upaya mendeteksi, dan mengklasifikasikan jenis penyakit yang mungkin dialami oleh pasien. Meskipun banyak terobosan telah diterapkan di dunia kesehatan, namun dalam konteks kesehatan mental, tidak semua teknologi tersebut dapat diimplementasikan dengan efektif. Untuk memudahkan tenaga kesehatan dalam melakukan prediksi, dan klasifikasi jenis penyakit gangguan jiwa maka dibutuhkan sebuah system rekomendasi prediksi, dan klasifikasi yang efektif. Dalam penelitian ini algoritma C5.0 digunakan untuk mengolah dan mengklasifikasikan data kedalam kategori tertentu, berdasarkan hubungan antara variabel-variabel yang dimasukkan kedalamnya, sekaligus membuat pohon keputusan yang berfungsi untuk memvisualkan hubungan antara variabel-variabel tersebut. Hasil pengujian dengan metode confusion matrix didapati akurasi sebesar 94,44%. Dimana pada label Depresi didapati precision sebesar 88%, Skizofrenia sebesar 94,44%, Bipolar sebesar 100%, Psikosis sebesar 91,80%, dan PTSD sebesar 100%. Sedangkan hasil recall pada label Depresi sebesar 88%, Skizofrenia sebesar 89,47%, Bipolar sebesar 90%, Psikosis sebesar 100%, dan PTSD sebesar 100%.

Kata kunci—Algoritma C5.0, Klasifikasi, Confusion Matrix, Pohon Keputusan, Gangguan Jiwa, Prediksi

Abstract

In the field of healthcare, particularly in the realm of mental health, technology has become a valuable tool for exploring various methods to detect and classify the types of illnesses that patients may be experiencing. While many breakthroughs have been applied in the healthcare sector, not all of these technologies can be effectively implemented in the context of mental health. To facilitate healthcare professionals in making predictions and classifications of mental disorders, an effective recommendation system is needed. In this research, the C5.0 algorithm is employed to process and classify data into specific categories based on the relationships between the input variables, simultaneously creating a decision tree to visualize these relationships. The testing results using the confusion matrix method indicate an accuracy rate of 94.44%. Specifically, for the labels, the precision for Depression is 88%. for Schizophrenia is

94.44%. for Bipolar is 100%. for Psychosis is 91.80%. and for PTSD is 100%. The recall results for Depression are 88%, for Schizophrenia are 89.47%. for Bipolar are 90%. for Psychosis are 100%. and for PTSD are 100%.

Keywords—C5.0 Algorithm, Classification, Confusion Matrix, Decision Tree, Mental Disorders, Prediction

1. PENDAHULUAN

Dalam dunia kesehatan terlebih dalam kesehatan mental, banyak pengguna teknologi yang mencoba mengembangkan berbagai metode untuk mengetahui dan mengklasifikasi jenis penyakit yang diderita oleh pasien. Teknologi tersebut dapat diintegrasikan kedalam proses pendeteksian, sistem pakar, dan pengelompokkan jenis penyakit berdasarkan gejala yang diderita oleh pasien [1]. Terobosan ini telah banyak diimplementasikan di dunia kesehatan, namun untuk kasus kesehatan mental tidak semua implementasi teknologi tersebut dapat diimplementasikan secara efektif dikarenakan memerlukan sample yang besar untuk membuat keputusan terhadap jenis penyakit mental apa yang diderita oleh pasien berdasarkan gejala yang mereka hadapi [2].

Pada penelitian ini, penulis bertujuan untuk mengimplementasikan algoritma C5.0 dalam sebuah aplikasi pengambilan keputusan untuk mengklasifikasikan jenis penyakit gangguan jiwa [3]. Algoritma C5.0 sendiri adalah algoritma yang berfungsi untuk memprediksi, dan mengklasifikasikan data kedalam kategori tertentu, berdasarkan hubungan antara variabel-variabel yang dimasukkan kedalamnya, sekaligus membuat pohon keputusan yang berfungsi untuk memvisualkan hubungan antara variabel-variabel tersebut. Data mining dapat menjadi sebuah dasar atau pedoman untuk menentukan kebijakan bisnis dalam upaya peningkatan pesaing bisnis perusahaan [4].

Pada tahun 2020, Utomo, et al melakukan penelitian tentang aplikasi klasifikasi terhadap penyakit jantung menggunakan algoritma C5.0. Dengan memanfaatkan dataset pasien yang mengidap penyakit jantung dengan ciri-ciri tertentu, mereka berhasil mengimplementasikan algoritma C5.0 untuk mengklasifikasikan penyakit jantung yang dialami oleh pasien sekaligus memvisualkannya dalam pohon keputusan. Akurasi ketepatan yang didapatkan pada penelitian ini sebesar 89,11% [5].

Saidata Aesyi et al mengimplementasikan algoritma C5.0 dalam proses diagnosa penyakit disk hernia dan *spondylolisthesis* berbasis website. Penelitian ini bertujuan untuk mencegah lebih dini terjadinya penyakit diskhernia, dan *spondylolisthesis* pada pasien. Pada penelitian ini algoritma C5.0 tidak hanya digunakan untuk membuat pohon keputusan dari variabel yang berkaitan, namun algoritma ini juga dipakai untuk memangkas variabel dari dataset yang tidak seimbang, dikarenakan terdapat beberapa variabel yang memiliki pemicu yang sama, namun menghasilkan hasil yang berbeda [6].

Pada penelitian yang dilakukan pada tahun 2019 oleh Nurhaningsih et al, mereka melakukan penelitian untuk mengklasifikasi kemungkinan terjangkitnya pasien terhadap penyakit gagal ginjal kronik, dengan membandingkan kemungkinan terbesar apabila seorang penyakit memiliki gejala-gejala tertentu. Penelitian ini sendiri berhasil membuat pohon klasifikasi yang memiliki akurasi sangat memuaskan sebesar 99,3% [7].

Khanif Zyen & Mulyo pada tahun 2022 melakukan sebuah penelitian tentang klasifikasi pasien demam berdarah dengan menggunakan algoritma decision tree C5.0 berbasis web. Tujuan utama dari penelitian ini adalah untuk mendeteksi penyakit demam berdarah berdasarkan gejala yang dialami oleh pasien. Sehingga diharapkan pasien bisa mendapatkan penanganan lebih awal terhadap penyakit ini, dan mengurangi angka kematian yang disebabkan oleh demam berdarah. Hasil akurasi yang didapat dari penelitian ini terbilang cukup baik, yakni sebesar 73,44% [8].

Pada tahun 2023 Meila Azzahra Sofyan et al, melakukan sebuah penelitian yang berfungsi untuk memprediksi penyakit stroke. Dataset yang dipakai pada penelitian ini adalah

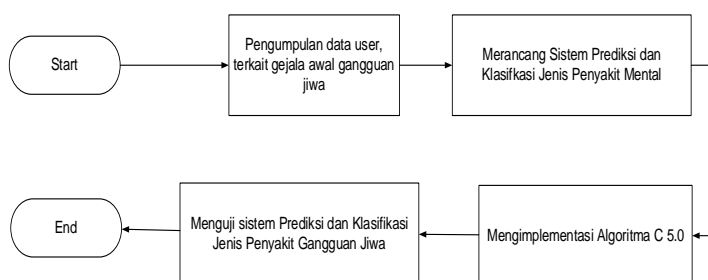
informasi awal terkait gejala medis dan faktor risiko yang terkait dengan stroke sejak dini. Penelitian ini dilakukan dengan mengumpulkan subjek, dan gejala dini yang mungkin terjadi kemudian mengelompokkannya, dan mengelompokkannya berdasarkan tingkat keparahan penyakit stroke itu sendiri. Tujuan utama dari penelitian ini adalah untuk mengembangkan model prediksi yang dapat membantu dalam identifikasi dini risiko terjadinya stroke. Hasil yang didapat dari penelitian ini adalah phpn keputusan dengan akurasi sebesar 95% [9].

Berdasarkan artikel-artikel yang telah dipaparkan diatas, dapat disimpulkan bahwa algoritma *decision tree* C5.0 adalah algoritma yang paling umum, dan paling efektif untuk menyelesaikan masalah klasifikasi, dan prediksi terhadap masalah yang dijabarkan pada artikel-artikel sebelumnya. Dan setiap artikel yang dipaparkan diatas menghasilkan akurasi, dan efisiensi kerja yang terbilang cukup baik sehingga algoritma ini sangat cocok untuk diimplementasikan pada penelitian ini. Oleh sebab itu fokus dari penelitian ini adalah memaksimalkan kinerja dari algoritma untuk menciptakan sebuah sistem yang dapat berjalan dengan mulus, dan efisien dalam proses pembuatan pohon keputusan, dan sistem pengklasifikasian jenis penyakit jiwa yang di derita oleh pasien, berdasarkan gejala-gejala yang dialami.

Penelitian ini akan dilakukan di salah satu rumah sakit jiwa bernama RSJ Mahoni yang berlokasi di kota Medan. Aplikasi yang dibuat nantinya tidak hanya berfungsi sebagai alat klasifikasi jenis penyakit saja, tapi juga sebagai alat untuk mengenalkan terhadap pasien (*user*) tentang jenis-jenis penyakit jiwa beserta gejala-gejala yang mungkin muncul, beserta cara penanganannya. Sehingga aplikasi ini dapat lebih bermanfaat bagi pasien, agar meningkatkan kesadaran akan pentingnya menjaga kesehatan jiwa. Cara kerja sistem pengklasifikasian ini nantinya akan meminta pasien (*user*) menginput gejala apa yang dialami selama beberapa waktu belakangan, dan data yang diinput oleh pasien (*user*) tersebut akan menghasilkan *decision tree* klasifikasi jenis penyakit apa yang dialami oleh pasien (tersebut) beserta penjelasannya. Manfaat yang diperoleh oleh pihak RSJ Mahoni nantinya adalah memudahkan mereka dalam mengklasifikasikan jenis penyakit gangguan jiwa tanpa harus melalui proses wawancara yang panjang dengan pasien, sehingga waktu yang digunakan dapat lebih efisien, dan efektif.

2. METODE PENELITIAN

Pada penelitian ini, peneliti akan menerapkan algoritma prediksi, dan klasifikasi C 5.0 untuk membuat sistem prediksi, dan klasifikasi jenis penyakit gangguan jiwa di RSJ Mahoni Medan. Berikut ini adalah gambaran umum tentang proses yang terjadi di dalam penelitian dari awal hingga akhir.



Gambar 1 Tahapan Penelitian

Pada Gambar 1 dikembangkan terlebih dahulu, dimulai dari Pengumpulan data mentah yang akan digunakan untuk proses klasifikasi sehingga mendapatkan data merancang tampilan aplikasi, dan merancang sistem prediksi dan klasifikasi jenis penyakit gangguan jiwanya. Selanjutnya, sistem prediksi dan klasifikasi diterapkan menggunakan algoritma C 5.0. Terakhir, pengujian dilakukan pada sistem prediksi dan klasifikasi jenis penyakit untuk mengevaluasi

kinerja aplikasi, dan menentukan apakah sistem tersebut berfungsi dengan baik. Berikut adalah deskripsi tahap-tahap penelitian.

2.1 Pengumpulan Data

Selama melakukan pengumpulan data, penulis mengumpulkan data dari basis data yang ada pada RSJ Mahoni, Medan. Penulis mencari informasi mengenai data diri pasien, mempelajari jenis penyakit ganggu, jenis penyakit gangguan jiwa apa yang di alami oleh pasien, gejala-gejala apa yang mungkin muncul pada diri pasien. Untuk mendapatkan data tersebut, penulis melakukan wawancara dengan beberapa perawat ahli, juga pasien yang sedang di rawat di RSJ Mahoni. Kemudian untuk membuat aplikasi prediksi, dan klasifikasi jenis penyakit gangguan jiwa, penulis memerlukan data jenis_penyakit, dan gejala_penyakit. Nilai dari jenis_penyakit akan berisi gejala_penyakit, sehingga apabila user menginput data gejala penyakit yang mirip atau sesuai dengan pola dari suatu jenis_penyakit maka sistem akan mengelola data inputan tersebut, dan mencari pola yang paling sesuai dengan gejala yang di input, kemudian menyajikannya dalam bentuk pohon keputusan.

Total data yang digunakan dalam penelitian ini adalah sebanyak 900 data pasien, yang terdiri dari 720 data training, dan 180 data testing. Sedangkan untuk jenis penyakit yang akan diteliti, dan di jadikan acuan untuk proses prediksi, dan klasifikasi adalah sebanyak 5 jenis penyakit gangguan jiwa, yaitu Skizofrenia, Depresi, PTSD, Bipolar, dan Psikosis.

2.2. Data Preprocessing

Data *preprocessing* adalah tahap untuk melakukan sebuah proses awal dalam pengolahan data. Pada tahap ini data yang akan diolah bertujuan untuk menghindarkan dari data yang mengganggu (*noise*) atau data yang tidak konsisten

2.3 Implementasi Algoritma C5.0

Dalam penelitian ini, algoritma C5.0 akan berfungsi untuk memproses prediksi, dan klasifikasi dataset yang ada. Algoritma C5.0 sendiri merupakan hasil perkembangan dari algoritma C4.5 yang sudah ada sebelumnya. Cara kerja algoritma ini adalah membagi dataset kedalam beberapa kelompok berbeda berdasarkan pola fitur yang mengikutinya, kemudian algoritma ini mencari pola fitur tersebut secara berulang-ulang agar dapat membuat model prediksi yang akurat. Berikut ini adalah langkah-langkah yang dilakukan pada proses implementasi algoritma C5.0 [10].

1. Menghitung nilai *entropy*, dengan tujuan untuk mengukur tingkat ketidakpastian pada *dataset*, dengan rumus persamaan seperti pada formula (1).

$$Entropy(S) = \sum_{i=1}^m p_i \times \log_2(p_i) \quad (1)$$

Keterangan :

S = Jumlah dari kasus

n = Jumlah dari partisi S

Pi = Rasio Si terhadap S

A = Atribut

2. Kemudian menghitung nilai *information gain*, dengan rumus persamaan (2).

$$Gain(S, A) = p_{total} - \frac{D_j}{D} \times Entropy(S_i) \quad (2)$$

Keterangan :

S = Jumlah dari kasus

A = Atribut

N = Jumlah partisi atribut A

|Si| = Jumlah kasus pada partisi ke-i

$|S|$ = Jumlah kasus dalam S
 $|S|$ = Jumlah kasus dalam S

3. Tahap terakhir adalah menghitung nilai *gain ratio* menggunakan formula (3), dengan tujuan untuk mendapatkan nilai acuan dalam menentukan *node* yang akan digunakan berikutnya.

$$\text{Gain Ratio} = \frac{\text{Gain}(S,A)}{\sum_{i=1}^m \text{Entropy}(S_i)} \quad (3)$$

Keterangan:

$\text{Gain}(S,A)$ = The gain of a variable

$\sum_{i=1}^n \text{Entropy}(S_i)$ = The amount of entropy in a variable

Pengujian sisem bertujuan untuk memastikan bahwa aplikasi dan sistem yang dibuat berfungsi dengan baik, sekaligus mencegah adanya *bug* atau gangguan pada aplikasi. Untuk menguji aplikasi prediksi, dan klasifikasi yang dibuat dengan algoritma C5.0, maka teknik pengujian yang akan digunakan adalah teknik pengujian *confusion matrix* [11]. Teknik *confusion matrix* sendiri sering digunakan untuk mengevaluasi kinerja sistem yang menggunakan model prediksi, dan klasifikasi dengan cara mengukur performa klasifikasi dengan membandingkan akurasi dari nilai prediksi, dan nilai aktual. Formula (4), formula (5), dan formula (6) adalah rumus untuk melakukan pengujian dengan *confusion matrix*.

$$\text{Akurasi} = \frac{\text{jumlah data benar}}{\text{jumlah keseluruhan data}} \times 100\% \quad (4)$$

$$\text{Precision} = \frac{\sum TP}{\sum (TP+FP)} \times 100\% \quad (5)$$

$$\text{Recall} = \frac{\sum TP}{\sum (TP+FN)} \times 100\% \quad (6)$$

2. 4 Pengujian Algoritma C5.0

Pengujian hasil prediksi, dan klasifikasi dengan algoritma C5.0 dilakukan dengan menggunakan aplikasi *Rapidminer Studio*, yang sering dipakai untuk pengujian data. Berikut ini adalah rangkaian kegiatan yang dilakukan dalam pengujian menggunakan aplikasi *rapidminer studio*. Hal pertama yang harus dilakukan adalah melakukan persiapan dataset (*Data Preperation*). Tahap ini bertujuan untuk mengenalkan atribut apa saja yang digunakan, dan label apa saja yang ada di dataset.

Row No.	Label	Nama Lengk...	Initial Jenis ...	Klasifikasi U...	Perubahan ...	Halusinasi	Delusi	Gangguan Pl...	Gangguan K...
1	Depresi	Muhammad I...	L	Dewasa Muda	Ya	Tidak	Tidak	Ya	Tidak
2	Skizofrenia	Ratu Maharani	P	Remaja	Ya	Ya	Ya	Ya	Ya
3	Bipolar	Rizky Pratama	L	Remaja	Ya	Tidak	Ya	Ya	Ya
4	Skizofrenia	Dian Novianti	P	Dewasa Muda	Ya	Ya	Ya	Ya	Ya
5	Psikosis	Bambang Su...	L	Dewasa Tua	Ya	Ya	Tidak	Ya	Tidak
6	PTSD	Sari Puspita	P	Dewasa Tua	Tidak	Tidak	Tidak	Ya	Ya
7	Psikosis	Hendra Wijaya	L	Remaja	Ya	Tidak	Ya	Tidak	Tidak
8	Depresi	Maya Anggraini	P	Dewasa Muda	Ya	Tidak	Tidak	Ya	Tidak
9	PTSD	Antonius Seti...	L	Dewasa Tua	Tidak	Ya	Tidak	Ya	Ya
10	Psikosis	Ratna Sari	P	Dewasa Muda	Ya	Tidak	Ya	Ya	Tidak
11	Psikosis	Budi Santoso	L	Dewasa Muda	Ya	Tidak	Ya	Tidak	Tidak
12	Skizofrenia	Indah Permata	P	Dewasa Tua	Ya	Ya	Ya	Ya	Ya
13	Bipolar	Darmawan H...	L	Remaja	Ya	Tidak	Ya	Ya	Ya
14	Depresi	Tri Utami	P	Dewasa Tua	Ya	Tidak	Tidak	Ya	Tidak
15	PTSD	Andi Firmans...	L	Remaja	Tidak	Ya	Tidak	Ya	Ya
16	Psikosis	Yulia Rahma...	P	Remaja	Ya	Ya	Ya	Ya	Tidak
17	Psikosis	Elko Nugroho	L	Dewasa Tua	Ya	Ya	Ya	Ya	Tidak
18	Bipolar	Fitriani Susanti	P	Dewasa Muda	Ya	Tidak	Ya	Ya	Ya
19	Skizofrenia	Joko Pramono	L	Remaja	Ya	Ya	Ya	Ya	Ya
20	Skizofrenia	Rina Anggraini	P	Dewasa Muda	Ya	Ya	Ya	Ya	Ya
21	Psikosis	Agus Wibowo	L	Dewasa Tua	Ya	Ya	Ya	Ya	Tidak
22	Bipolar	Dewi Lestari	P	Remaja	Ya	Tidak	Ya	Ya	Ya
23	Skizofrenia	Wahyu Nugro...	L	Dewasa Tua	Ya	Ya	Ya	Ya	Ya
24	Bipolar	Sinta Puspita	P	Dewasa Tua	Ya	Tidak	Ya	Ya	Ya
25	Bipolar	Ahmad Firdaus	L	Dewasa Muda	Ya	Tidak	Ya	Ya	Ya
26	Skizofrenia	Ratih Setiawati	P	Dewasa Muda	Ya	Ya	Ya	Ya	Ya

ExampleSet (720 examples, 1 special attribute, 8 regular attributes)

Gambar 2 Dataset Penelitian

Pada Gambar 2 dapat dilihat bahwa dataset yang digunakan adalah data *training* yang berjumlah 720 set data, dengan 131 data yang terkalsifikasi sebagai label Depresi, 144 data terkalasifikasi sebagai label Skizofrenia, 141 data terkalasifikasi Bipolar, 173 data terkalasifikasi

sebagai Psikosis, dan 124 data terklasifikasi sebagai PTSD. Tahap Selanjutnya adalah membuat design process kerja yang digunakan untuk menghasilkan decision tree dengan algoritma C5.0. Fitur-fitur yang akan digunakan antara lain retrieve data, dan cross validation, dimana pada *cross validation* terdapat beberapa fitur lain yang menyertainya, yaitu *decision tree*, *apply model*, dan *performance*.

3. HASIL DAN PEMBAHASAN

3.1 Dataset Pasien RSJ Mahoni.

Pada penelitian ini data yang digunakan adalah sebanyak 900 baris data. Kemudian dilakukan pembagian data training, dan data testing dengan data training sebanyak 80% dari total data, dan data testing sebanyak 20% dari total data.

Tabel 1 Data Training

Nama	Perubahan Suasana Hati	Halusinasi	Delusi	Gangguan Pikiran	Label
User(1)	Ya	Tidak	Tidak	Ya	Depresi
User(2)	Ya	Ya	Ya	Ya	Skizofrenia
User(3)	Ya	Tidak	Ya	Ya	Bipolar
User(4)	Ya	Ya	Ya	Ya	Skizofrenia
User(5)	Ya	Ya	Tidak	Ya	Psikosis
....
User(720)	Ya	Ya	Ya	Ya	Depresi

Pada Tabel 1 ditampilkan data *training* yang penulis gunakan sebanyak 720 *dataset*. Jumlah tersebut sejalan dengan aturan pembagian data *training* dengan data *testing* di penjelasan sebelumnya. Dimana 720 merupakan 80% dari 900 *dataset*. Data training tersebutlah yang akan digunakan untuk mengimplementasikan algoritma C5.0 pada aplikasi prediksi, dan klasifikasi jenis penyakit gangguan jiwa.

Tabel 2 Data Testing

Nama	Perubahan Suasana Hati	Halusinasi	Delusi	Gangguan Pikiran	Gangguan Kognitif	Label
User(1)	Ya	Ya	Ya	Ya	Ya	Skizofrenia
User(2)	Ya	Tidak	Ya	Ya	Ya	Bipolar
User(3)	Tidak	Ya	Tidak	Ya	Ya	PTSD
User(4)	Tidak	Ya	Ya	Tidak	Tidak	Psikosis
User(5)	Ya	Tidak	Tidak	Ya	Tidak	Depresi
.....
User(180)	Ya	Tidak	Tidak	Ya	Tidak	Depresi

Sedangkan pada Tabel 2 data testing yang digunakan berjumlah 180 *dataset*. Jumlah ini sejalan dengan aturan pembagian yang sudah dijelaskan sebelumnya, dimana data yang digunakan untuk testing adalah 20% dari total keseluruhan dataset. Data *testing* ini nantinya akan berfungsi untuk menguji performa dari algoritma C5.0 yang penulis gunakan, dengan metode pengujian *confusion matrix*.

3.2 Penerapan Algoritma C5.0

Penerapan algoritma C5.0 ini akan melalui beberapa tahapan, diantaranya adalah menghitung jumlah kasus berdasarkan gejala (atribut) yang dialami oleh pasien, kemudian menghitung nilai *entropy* dari setiap atribut, selanjutnya menghitung nilai *information gain*, dan *gain ratio* [12]. Dari hasil perhitungan *gain ratio* tersebut akan terbentuk node yang akan saling terhubung hingga menciptakan *decision tree*, yang berfungsi sebagai acuan dalam pembentukan pola prediksi, dan klasifikasi jenis penyakit gangguan jiwa. Berikut ini adalah tahapan-tahapan perhitungan yang penulis lakukan

Hal pertama yang dilakukan adalah menghitung jumlah kasus berdasarkan atribut (gejala) yang bertujuan untuk menemukan jumlah kasus yang terjadi pada setiap pasien, beserta label yang mengikutinya. Pengujian sisem bertujuan untuk memastikan bahwa aplikasi dan sistem yang dibuat berfungsi dengan baik, sekaligus mencegah adanya bug atau gangguan pada aplikasi. Untuk menguji aplikasi prediksi, dan klasifikasi yang dibuat dengan algoritma C5.0, maka teknik pengujian yang akan digunakan adalah teknik pengujian *confusion matrix* [13]. Seperti yang diketahui terdapat 5 gejala (atribut) yang digunakan sebagai acuan, yaitu perubahan susasana hati, halusinasi, delusi, gangguan pikiran, gangguan kognitif. Sedangkan label yang digunakan adalah 5 jenis penyakit yang berkaitan dengan gejala (atribut) yang ada. Kelima label tersebut antara lain Bipolar, Depresi, Psikosis, PTSD, Skizofrenia.

Tabel 3 Perhitungan Jumlah Kasus Berdasarkan Gejala (atribut)

Gejala (atribut)		Jumlah Kasus	Bipolar	Depresi	Psikosis	PTS D	Skizofrenia
Gangguan Kognitif	Ya	445	141	36	0	124	144
	Tidak	275	0	102	173	0	0
Halusinasi	Ya	346	0	36	146	20	144
	Tidak	374	141	102	27	104	0
Delusi	Ya	420	130	0	143	3	144
	Tidak	300	11	138	30	121	0
Perubahan Suasana Hati	Ya	573	141	138	150	0	144
	Tidak	147	0	0	23	124	0
Gangguan Pikiran	Ya	675	141	122	149	124	139
	Tidak	45	0	16	24	0	5

Dari Tabel 3 dapat diketahui kemunculan setiap gejala (atribut) yang dialami oleh setiap pasien yang ada. Kemudian tahap selanjutnya adalah menghitung nilai *entropy* dari setiap gejala (atribut), dan nilai *entropy* total dari keseluruhan atribut. Sebagai contoh penulis akan menghitung nilai *entropy* dari gejala (atribut) gangguan kognitif, dan *entropy* total dari keseluruhan data.

Maka langkah selanjutnya adalah menghitung *entropy* pada setiap gejala (atribut) yang ada agar dapat melakukan perhitungan *information gain* dari hasil perhitungan keseluruhan *entropy*. Untuk hasil perhitungan *entropy* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Perhitungan Nilai *Entropy*

Gejala (atribut)	Entropy
Gangguan Kognitif	Ya 1.86
	Tidak 0.95
Halusinasi	Ya 1.63
	Tidak 1.83
Delusi	Ya 1.63
	Tidak 1.55
Perubahan Suasana Hati	Ya 2.08
	Tidak 0.63
Gangguan Pikiran	Ya 1.87
	Tidak 1.37
Total Entropy	2.31

Selanjutnya dilakukan perhitungan *information gain*. Sebagai contoh penulis akan menggunakan *entropy* dari gejala (atribut) gangguan kognitif, dan halusinasi, sebagai acuan untuk perhitungan *information gain* pada *entropy* lainnya.

Maka langkah selanjutnya adalah menghitung *information gain* pada setiap gejala (atribut) yang ada agar dapat melakukan perhitungan *gain ratio* pada tahap selanjutnya. Untuk

hasil perhitungan information gain keseluruhan gejala (atribut) dapat dilihat pada Tabel 5 dibawah ini.

Tabel 5 Hasil Perhitungan *Information Gain*

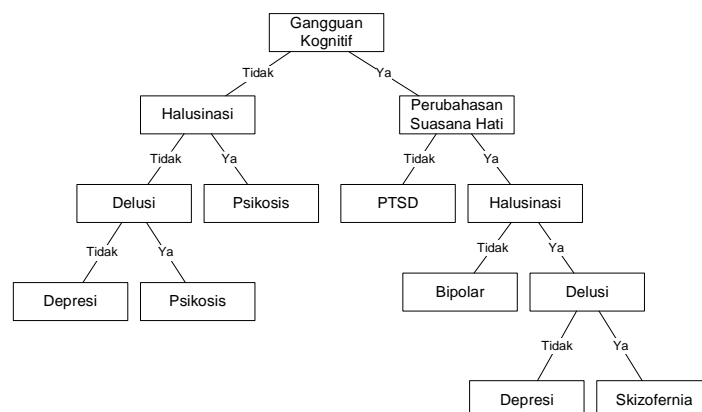
Gejala (atribut)			Entropy	Information Gain
Kognitif	Gangguan	Ya	1.86	0.80
		Tidak	0.95	
Halusinasi		Ya	1.63	0.58
		Tidak	1.83	
Delusi		Ya	1.63	0.71
		Tidak	1.55	
Perubahan Suasana Hati		Ya	2.08	0.53
		Tidak	0.63	
Gangguan Pikiran		Ya	1.87	0.48
		Tidak	1.37	
Total <i>Entropy</i>			2.31	

Tahap terakhir adalah menghitung nilai dari gain ratio yang akan menjadi acuan pada penentuan *node* pada proses pembuatan decision tree dengan algoritma C5.0.

Tabel 6. Hasil Perhitungan *Gain Ratio*

Gejala (atribut)		Entropy	Information Gain	Gain Ratio
Gangguan Kognitif	Ya	1.86	0.80	0.28
	Tidak	0.95		
Halusinasi	Ya	1.63	0.58	0.17
	Tidak	1.83		
Delusi	Ya	1.63	0.71	0.22
	Tidak	1.55		
Perubahan Suasana Hati	Ya	2.08	0.53	0.20
	Tidak	0.63		
Gangguan Pikiran	Ya	1.87	0.48	0.15
	Tidak	1.37		
Total <i>Entropy</i>		2.31		

Jika diperhatikan pada Tabel 6, nilai *gain ratio* tertinggi adalah atribut “Gangguan Kognitif”. Maka berdasarkan aturan dari algoritma C5.0, dimana apabila terdapat nilai *gain ratio* yang lebih tinggi pada suatu atribut, maka atribut tersebut yang akan menjadi *node* akar (*root*), dimana *node* akar merupakan *node* awal, sekaligus acuan untuk melakukan percabangan *node* berikutnya. Sehingga dapat disimpulkan bahwa gejala (atribut) “Gangguan Kognitif” yang akan menjadi *node* awal pada *decision tree* ini.

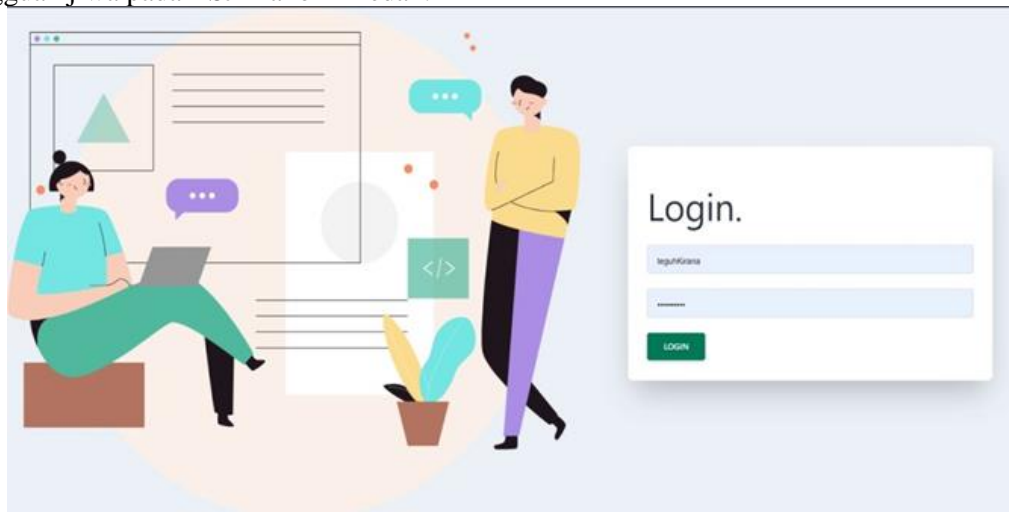


Gambar 3 *Decision Tree*

Setelah melalui tahap perhitungan keseluruhan node atau biasa disebut dengan *gase training* dengan *decision tree*[14], maka akan terbentuk sebuah *decision tree* yang dapat dilihat pada Gambar 2. Sehingga setelah *tree* udah terbentuk maka pola prediksi, dan klasifikasi dapat direpresentasikan dalam bentuk list. Setiap pola yang terbentuk merupakan hubungan antar gejala (atribut), yang mengacu pada terbentuknya sebuah keputusan pada *leaf node* untuk menentukan jenis penyakit gangguan jiwa apa yang diderita oleh pasien.

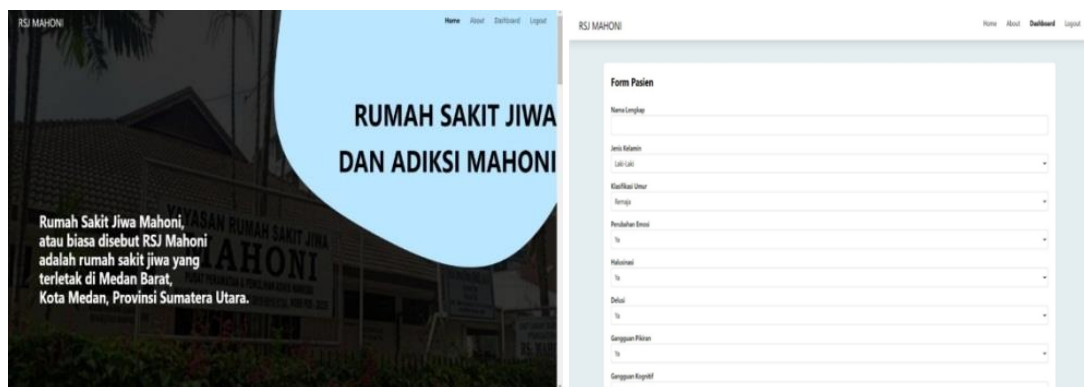
3.3 Implementasi Aplikasi

Implementasi aplikasi ini memberikan gambaran halaman antarmuka aplikasi berbasis web baik dari sisi pengguna atau pasien, maupun dari sisi tenaga kesehatan. Berikut adalah halaman antarmuka aplikasi sistem prediksi, dan klasifikasi untuk mendiagnosa jenis penyakit gangguan jiwa pada RSJ Mahoni Medan.



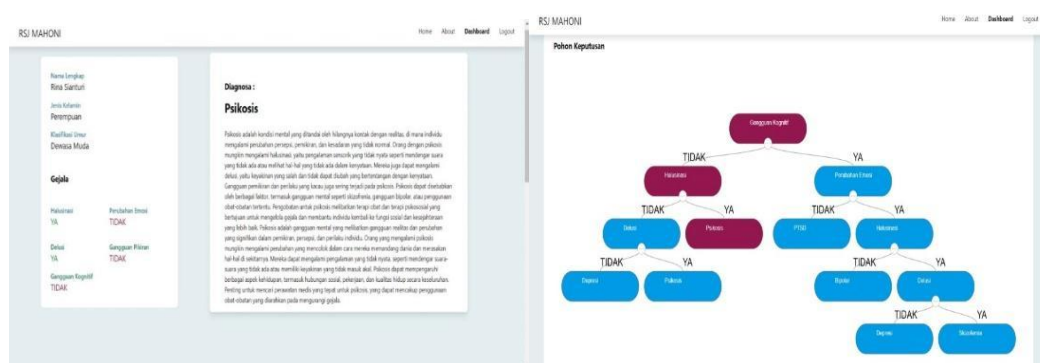
Gambar 4 Halaman Login

Gambar 4 menampilkan halaman Login utama untuk pengguna dan admin. Halaman Login berfungsi sebagai antarmuka awal bagi pengguna dan admin untuk mengakses dashboard aplikasi prediksi, dan klasifikasi jenis penyakit gangguan jiwa yang telah diimplementasikan sesuai dengan tujuan penelitian.



Gambar 5 Tampilan Halaman Awal

Gambar 5 menampilkan halaman awal yang akan muncul ketika pengguna berhasil melakukan login. Pada halaman ini pengguna dapat mengakses berbagai hal seperti data tentang rumah sakit jiwa mahoni, ilmuilmu dasar tentang penyakit gangguan jiwa, dan tentunya *form* input data gejala pasien untuk proses prediksi, dan klasifikasi jenis penyakit jiwa yang diderita.



Gambar 6 Tampilan Hasil Prediksi dan Klasifikasi

Gambar 6 merupakan halaman yang akan muncul ketika pengguna selesai menginput form gejala, kemudian sistem akan memberikan output berupa pohon keputusan, dan penjelasan atas jenis penyakit gangguan jiwa apa yang diderita oleh pasien.

3.4 Evaluasi dan Validasi Dengan Confusion Matrix

Penelitian ini menggunakan teknik *confusion matrix*. Teknik *confusion matrix* mengevaluasi kinerja sistem yang menggunakan model prediksi, dan klasifikasi dengan cara mengukur performa klasifikasi dengan membandingkan akurasi dari nilai prediksi, dan nilai aktual [15]. Dengan menggunakan aplikasi rapidminer pengujian data testing dapat dilakukan dengan lebih mudah, fitur yang cocok untuk melakukan evaluasi dan validasi performa dari algoritma C5.0 dengan teknik confusion matrix adalah fitur apply model, dan *performance*. Hasil dari pembuatan model pada fitur yang ada di *rapidminer* menghasilkan output seperti Gambar 7.

accuracy: 94.44%

	true Depresi	true Skizofrenia	true Bipolar	true Psikosis	true PTSD	class precision
pred. Depresi	22	2	1	0	0	88.00%
pred. Skizofrenia	1	34	1	0	0	94.44%
pred. Bipolar	0	0	27	0	0	100.00%
pred. Psikosis	2	2	1	56	0	91.80%
pred. PTSD	0	0	0	0	31	100.00%
class recall	88.00%	89.47%	90.00%	100.00%	100.00%	

Gambar 7 Performa Pengujian

Pada gambar 7 Berdasarkan evaluasi dan validasi dengan teknik *confusion matrix* pada Gambar 6 tersebut, didapatkan akurasi sebesar 94,44%. Dimana pada label Depresi didapati precision sebesar 88%, sedangkan label Skizofrenia sebesar 94,44%, label Bipolar didapati precision sebesar 100%, sedangkan label Psikosis didapati precision sebesar 91,80%, dan label PTSD didapati precision sebesar 100%. Disisi lain didapati hasil recall pada label Depresi sebesar 88%, Skizofrenia sebesar 89,47%, Bipolar sebesar 90%, Psikosis sebesar 100%, dan PTSD sebesar 100%.

4. KESIMPULAN

Berdasarkan hasil penelitian, dan pengujian aplikasi prediksi, dan klasifikasi penyakit gangguan jiwa pada pasien di RS Mahoni Medan dengan algoritma C5.0. Hasil akhir dari perhitungan dengan algoritma C5.0 adalah membuat rule dengan decision tree, sehingga proses prediksi, dan klasifikasi jenis penyakit gangguan jiwa dapat dengan mudah di deteksi. Penerapan algoritma C5.0 menunjukkan performa yang baik, dikarenakan pada saat dilakukan evaluasi, dan validasi dengan metode pengujian confusion matrix didapati akurasi sebesar 94,44%. Dimana pada label Depresi didapati precision sebesar 88%, Skizofrenia sebesar 94,44%, Bipolar sebesar 100%, Psikosis sebesar 91,80%, dan PTSD sebesar 100%. Sedangkan hasil *recall* pada label

Depresi sebesar 88%, Skizofrenia sebesar 89,47%, Bipolar sebesar 90%, Psikosis sebesar 100%, dan PTSD sebesar 100%.

5. SARAN

Aplikasi prediksi, dan klasifikasi jenis penyakit gangguan jiwa berdasarkan gejala berbasis web dapat dikembangkan lebih lagi kedepannya, berikut ini adalah saran untuk pengembangan aplikasi ini.

1. Melakukan pengujian objektif pada kesesuaian objek penelitian dengan atribut yang lebih bervariasi.
2. Mencoba menerapkan algoritma prediksi lain seperti algoritma prediksi, dan klasifikasi *K-Nearest Neighbour*, *Random Forest*, *Naïve Bayes*, *Support Vector Machine*, dan algoritma lainnya.
3. Mencoba mengganti platform atau basis dari aplikasi menjadi basis android, ataupun basis berbasis desktop.

UCAPAN TERIMA KASIH

Segala puji syukur penulis panjatkan bagi Allah SWT, atas selesainya penelitian ini. Penulis juga ingin berterima kasih kepada Departemen Sistem Informasi Universitas Islam Negeri Medan atas dukungannya sepanjang penelitian ini, baik secara langsung maupun tidak langsung. Terima kasih khusus kepada dosen pembimbing yang ikut serta dalam penelitian ini selama salah satu studi di RSJ Mahoni Medan. Semoga penelitian ini bermanfaat dalam meningkatkan prediksi, dan klasifikasi jenis penyakit bagi pasien, serta meningkatkan efisiensi kerja sekaligus meningkatkan kualitas RSJ Mahoni Medan. Selain itu penulis berharap penelitian ini dapat memberikan pengetahuan kepada para pembaca, dan juga mampu berkontribusi pada pemahaman mereka dalam penelitian di masa depan.

DAFTAR PUSTAKA

- [1] D. Nofriansyah and G. W. Nurchayo, *Algoritma Data Mining dan Pengujiannya*, vol. 2. 2015.
- [2] and M. K. A. Nurzahputra, M. A. Muslim, "Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa," *Techno.Com*, vol. 16, no. 17–24, 2017.
- [3] Triase and Samsudin, "Implementasi Data Mining dalam Mengklasifikasikan UKT (Uang Kuliah Tunggal) pada UIN Sumatera Utara Medan. Jurnal Teknologi Informasi," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 370–376, 2020.
- [4] E. Buulolo, "Data Mining Untuk Perguruan Tinggi," 2020.
- [5] D. P. Utomo, P. Sirait, and R. Yunis, "Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5.0," *J. Media Inform. Budidarma*, vol. 4, no. 4, 2020.
- [6] U. S. Aesyi, T. W. Diwangkara, and R. T. Kurniawan, "DIAGNOSA PENYAKIT DISK HERNIA DAN SPONDYLOLISTHESIS MENGGUNAKAN ALGORITMA C5," *Telematika*, vol. 16, no. 2, 2020, doi: 10.31315/telematika.v16i2.3181.
- [7] S. Nurhaningsih, Y. Susanti, and S. S. Handajani, "Implementasi Algoritma C5.0 Untuk Klasifikasi Penyakit Gagal Ginjal Kronik," *INTEK J. Inform. dan Teknol. Inf.*, vol. 2, no. 1, 2019, doi: 10.37729/intek.v2i1.89.
- [8] H. M. Ahmad Khanif Zyen, "Implementasi Algoritma Decision Tree C5.0 untuk Klasifikasi Pasien Demam Berdarah di Kabupaten Rembang," *AMRI J. Nas. Anal. Metod. Rekayasa Inform.*, vol. 1, no. 1, pp. 42–51, 2020.
- [9] F. M. A. Sofyan, A. P. Riyandoro, D. F. Maulana, and J. H. Jaman, "No Title," *J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 6, no. 2, pp. 619–625, 2023.

- [10] R. Ella Sari, Solikhun, and F. Rizky, "Penerapan Algoritma C5.0 dalam Memprediksi Persediaan Buah pada UD. Bunda Syafira Buah," *JUKI J. Komput. dan Inform.*, vol. 3, no. 2, 2021, doi: 10.53842/juki.v3i2.62.
- [11] R. N. Amalda, N. Millah, and I. Fitria, "IMPLEMENTASI ALGORITMA C5.0 DALAM MENGANALISA KELAYAKAN PENERIMA KERINGANAN UKT MAHASISWA ITK," *Teorema Teor. dan Ris. Mat.*, vol. 7, no. 1, 2022, doi: 10.25157/teorema.v7i1.6692.
- [12] D. D. Jantce, T. J. Sitinjak, and J. Suwita, "ANALISA DAN PERANCANGAN SISTEM INFORMASI ADMINISTRASI KURSUS BAHASA INGGRIS PADA INTENSIVE ENGLISH COURSE DI CILEDUG TANGERANG," *Insa. Pembang. Sist. Inf. dan Komput.*, 2020, [Online]. Available: <https://api.semanticscholar.org/CorpusID:255673058>
- [13] C. N. Dengen, K. Kusriani, and E. T. Luthfi, "Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *SISFOTENIKA*, vol. 10, no. 1, p. 1, 2020, doi: 10.30700/jst.v10i1.484.
- [14] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0192-5.
- [15] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Mach. Learn.*, vol. 16, no. 3, pp. 235–240, 1994, doi: 10.1007/BF00993309.