

KOMPARASI ALGORITMA C4.5 BERBASIS PSO DAN GA UNTUK DIAGNOSA PENYAKIT STROKE

Ramadhan Saepul Rohman¹, Rizal Amegia Saputra², Dasya Arif Firmansaha³

Page | 155

^{1,2,3} *Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika
Jl. Cemerlang No. 8 Sukakarya Sukabumi, Jawa Barat
¹ ramdhan.rpe@bsi.ac.id, ² rizal.rga@bsi.ac.id, ³ dasya.daf@bsi.ac.id*

Abstrak— Stroke merupakan gangguan fungsi otak baik lokal maupun menyeluruh yang disebabkan karena pasokan darah ke otak terganggu yang terjadi secara cepat dan berlangsung lebih dari 24 jam atau berakhir dengan kematian. Stroke Merupakan 1 dari 10 jenis penyakit yang paling mematikan di Indonesia. Hal ini berdasarkan pada data yang dikumpulkan dari sampel yang mewakili Indonesia, meliputi 41.590 kematian sepanjang 2014 dan pada semua kematian itu dilakukan autopsi verbal, sesuai pedoman Badan Kesehatan Dunia. Pentingnya mengetahui gejala sejak dini merupakan langkah awal dalam pencegahan terjadinya stroke. Maka itu, dilakukan penelitian untuk menganalisa data terkait dengan penyebab stroke. Adapun atribut yang terlibat dalam penyebab terjadinya stroke yakni, usia, jenis kelamin, kadar glukosa, riwayat penyakit jantung, hipertensi, tipe pekerjaan, tipe tempat tinggal, status merokok, index masa tubuh dan status pernikahan. Diperlukan suatu algoritma tertentu untuk mengklasifikasikan semua atribut tersebut. C45 merupakan Algoritma yang paling banyak digunakan, dalam kasus ini akurasi dari algoritma C4.5 sebesar 99.07%. Selanjutnya Algoritma C4.5 dioptimasi dengan menggunakan *Particle Swarm Optimization* sehingga memperoleh akurasi sebesar 99.28% dan Algoritma C4.5 juga dioptimasi dengan menggunakan *Genetic Algorithm* sehingga memperoleh akurasi sebesar 99.38%.

Kata Kunci— Stroke, C4.5, Algoritma C4.5, *Particle Swarm Optimization*, *Genetic Algorithm*.

Abstract— Stroke is a disorder of brain function both locally and comprehensively due to a disrupted blood supply to the brain that occurs quickly and lasts more than 24 hours or ends with death. Stroke is one of the 10 most deadly diseases in Indonesia. This is based on data collected from samples representing Indonesia, covering 41,590 deaths throughout 2014 and in all deaths a verbal autopsy was carried out, according to World Health Organization guidelines. The importance of knowing symptoms early on is the first step in preventing stroke. Therefore, research was conducted to analyze data related to the cause of stroke. The attributes involved in the cause of stroke are age, sex, glucose level, history of heart disease, hypertension, type of work, type of residence, smoking status, body mass index and marital status. A certain algorithm is needed to classify all of these attributes. C45 is the most widely used algorithm, in this case the accuracy of the C4.5 algorithm is 99.07%. Furthermore, the C4.5 Algorithm was optimized using *Particle Swarm Optimization* to obtain an accuracy of 99.28% and the C4.5 Algorithm was also optimized by using the *Genetic Algorithm* to obtain an accuracy of 99.38%.

Keywords— Stroke, C4.5, Algoritma C4.5, *Particle Swarm Optimization*, *Genetic Algorithm*.

I. PENDAHULUAN

Pentingnya mengetahui gejala penyakit merupakan langkah awal dalam mengantisipasi timbulnya suatu penyakit yang dapat membahayakan kesehatan dan bahkan menyebabkan kematian.

Stroke merupakan gangguan fungsi otak baik lokal maupun menyeluruh yang disebabkan karena pasokan darah ke otak terganggu yang terjadi secara cepat dan berlangsung lebih dari 24 jam atau berakhir dengan kematian[1].

Stroke Merupakan 1 dari 10 jenis penyakit yang paling mematikan di Indonesia. Hal ini berdasarkan pada data yang dikumpulkan dari sampel yang mewakili Indonesia, meliputi 41.590 kematian sepanjang 2014 dan pada semua kematian itu

dilakukan autopsi verbal, sesuai pedoman Badan Kesehatan Dunia (WHO) secara real time oleh dokter dan petugas terlatih [2].

Stroke menjadi salah satu penyakit yang paling banyak diderita oleh masyarakat Indonesia dan juga menjadi salah satu penyebab kematian tertinggi di Indonesia pada urutan pertama disusul oleh diabetes dan hipertensi[3].

Telah banyak penelitian yang sudah dilakukan diantaranya: Menurut jurnal Sigit Abdillah yang berjudul Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi Data Mining Pada Rumah Sakit Santa Maria Pemalang, mengemukakan bahwa untuk studi kasus penyakit stroke dapat memanfaatkan teknik klasifikasi data

mining dengan algoritma C4.5 sebagai klasifikasi stroke atau non-stroke. Dari metode klasifikasi data mining dengan algoritma C4.5 dan pengaplikasian pohon keputusan yang membentuk aturan tersebut terdapat akurasi pada data training yang berjumlah 130 dari 156 data pasien sebesar 82,31% sedangkan akurasi pada data testing yang berjumlah 26 dari 156 data pasien sebesar 76,92%. Perhitungan keduanya menggunakan confusion matrix. Menurut jurnal Vina Adelina, Dian Eka Ratnawati, M. Ali Fauzi dalam penelitiannya menentukan tingkat resiko stroke menggunakan metode Metode GA-Fuzzy Tsukamoto dengan tingkat akurasi yang didapatkan sebesar 86.66%. Menurut jurnal Jeena R S dan Suresh Kumar dalam penelitiannya memprediksi stroke menggunakan metode SVM dan akurasi yang diperoleh sebesar 90% dengan menggunakan 12 atribut berupa Usia, Jenis Kelamin, gejala Berjalan, Fibrilasi Atrium, Defisit wajah, Lengan / Tangan defisit, Kaki / Kaki defisit, Infark terlihat pada CT, Dyphasi, Hemianopia, Gangguan Visuospasial, tanda Cerebellar.

Penelitian ini bertujuan untuk menerapkan algoritma C4.5 sebagai algoritma data mining yang paling banyak digunakan serta komparasi antara optimasi *Genetic Algorithm* (GA) dan optimasi *Particle Swarm Optimization* (PSO) dalam memprediksi terjadinya penyakit *stroke*.

II. METODE PENELITIAN

A. Jenis Penelitian

Dalam Penelitian ini dilakukan dua pendekatan utama penelitian, yaitu pendekatan kualitatif dan pendekatan kuantitatif. Pendekatan Kualitatif digunakan untuk menganalisa kajian literatur yang berkenaan dengan variable-variabel yang digunakan dalam pengumpulan data[5].

Sedangkan pendekatan kuantitatif merupakan metode penelitian yang digunakan untuk meneliti pada populasi atau sampel tertentu, pengumpulan data menggunakan instrumen penelitian dan analisis data bersifat statistik atau kuantitatif dengan tujuan untuk menguji hipotesis yang telah ditetapkan[6].

B. Metode Pengumpulan Data

Pada tahap ini metode pengumpulan data dibagi menjadi dua sumber data yakni data primer dan data sekunder. Data primer yaitu data yang dikumpulkan dari sumbernya langsung, sedangkan data sekunder yaitu data yang dikumpulkan dari peneliti sebelumnya. Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang digunakan adalah metode pengumpulan data sekunder. Data utama diperoleh dari kaggle dataset repository yang dapat diperoleh melalui alamat web <https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data> sedangkan data pendukung didapatkan dari buku, jurnal dan publikasi lainnya

C. Validasi Data

Data dalam penelitian ini diambil dari *Kaggle Dataset* yang terdiri dari 10 atribut *predictor* dan 1 atribut hasil. Pada dataset tersebut akan dilakukan validasi data dengan menghilangkan data *missing value*. Data yang didapat dari *Kaggle Dataset* sebanyak 43401 *record*, dari data tersebut terdapat data *missing value* sebanyak 14335 *record*. Kemudian data *missing value* tersebut akan dihilangkan agar data menjadi valid sehingga data yang diolah tanpa *missing value* adalah sebanyak 29066 *record* setelah itu dilakukan *remove duplicate* sehingga data yang diperoleh sebanyak 2906. Pengolahan data juga dapat berupa konversi nilai agar mempermudah pembentukan model

TABEL I
ATRIBUT DAN NILAI KATEGORI DALAM MEMPREDIKSI
PENYAKIT STROKE

| Atribut | Nilai | Sumber |
|--------------------------------|-----------------|----------------------------------|
| AGE | <15 | (Kementerian Kesehatan RI, 2014) |
| | 15-24 | |
| | 25-34 | |
| | 35-44 | |
| | 45-54 | |
| | 55-64 | |
| | 65-74 | |
| GENDER | FEMALE | <i>Kaggle Dataset</i> |
| | MALE | |
| HYPERTENSION | 0 | <i>Kaggle Dataset</i> |
| | 1 | |
| HEART DISEASE | 0 | <i>Kaggle Dataset</i> |
| | 1 | |
| EVER MARRIED | YES | <i>Kaggle Dataset</i> |
| | NO | |
| WORK TYPE | Children | <i>Kaggle Dataset</i> |
| | Private | |
| | Self Employed | |
| | Govt job | |
| | Never Work | |
| RESIDENCE TYPE | Urban | <i>Kaggle Dataset</i> |
| | Rural | |
| AVG CLUCOSE LEVEL (After Meal) | <180 mg/dl | (Noya, 2016) |
| | >180 mg/dl | |
| BMI | <30 Kg/m | (Quamila, 2017) |
| | 30 Kg/m | |
| | >30 Kg/m | |
| SMOKING STATUS | Formerly smoked | <i>Kaggle Dataset</i> |
| | Smoked | |
| | Never Smoked | |

Berikut merupakan deskripsi dari atribut diatas :

1. Age/Umur

Mencakup usia dari usia pasien, menurut data kementerian kesehatan batasan usia yang berpotensi mengidap penyakit stroke diantaranya :

- 15-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65-74

- h. >75
2. *Gender/ Jenis kelamin*
Mencakup jenis kelamin dari pasien dengan kategori *Male* (Laki-laki) dan *Female* (Perempuan)
 3. *Hypertension*
Salah satu atribut yang mengindikasikan apakah pasien yang bersangkutan memiliki riwayat penyakit hipertensi/ darah tinggi ataukah tidak
 4. *Heart Disease*
Salah satu atribut yang mengindikasikan apakah pasien yang bersangkutan memiliki riwayat penyakit penyakit jantung ataukah tidak
 5. *EverMarried*
Salah satu atribut yang mengindikasikan apakah pasien yang bersangkutan pernah menikah ataukah tidak
 6. *Work Type*
Salah satu atribut yang mengindikasikan type pekerjaan dari pasien yang merupakan salah satu penyebab stres yang berpotensi terjadinya penyakit stroke diantaranya:
 - a. *Children* (Anak-anak)
 - b. *Private* (Pribadi)
 - c. *Self Employed* (Bekerja Sendiri)
 - d. *Govt_job* (Pekerja Pemerintahan)
 - e. *Never Work* (Tidak Bekerja)
 7. *Residence Type*
Mencakup tempat tinggal dari pasien yang meliputi *Urban* (Perkotaan) dan *Rural* (Pedesaan)
 8. *AVG Glucose Level*
Mencakup rata-rata tingkat glukosa (Kadar gula dalam darah) yang dapat menyebabkan timbulnya penyakit stroke. pada dataset yang diteliti ini mencakup “measured after meal” (Diukur setelah makan).
Kadar glukosa yang berpotensi mengakibatkan stroke pada pasien, diantaranya adalah pasien dengan kadar gula berlebih (hiperglikemia). Berikut kisaran kadar gula darah normal pada tubuh (Noya, 2016):
 - a. Sebelum makan: sekitar 70-130 mg/dL
 - b. Dua jam setelah makan: kurang dari 180 mg/dL
 - c. Setelah tidak makan (puasa) selama setidaknya delapan jam: kurang dari 100 mg/dL
 - d. Menjelang tidur: 100 – 140 mg/dL

Menjaga kadar gula darah agar dalam angka normal sangat penting. Gula darah terlalu rendah (hipoglikemia) atau tinggi (hiperglikemia) bisa berdampak negatif pada tubuh. Jika gula darah

Anda di bawah 70 mg/dL maka Anda mengalami hipoglikemia. Anda dikatakan mengalami hiperglikemia jika kadar gula darah Anda lebih dari 200 mg/dL.

9. *BMI (Body Mass Index)*
Obesitas dengan tingkat BMI > 30 Kg/m Sangat berpotensi mengalami resiko stroke secara independen (Li, et al., 2015)
Didefinisikan sebagai indeks massa tubuh (BMI) 30 atau lebih, obesitas sendiri merupakan faktor risiko untuk stroke, terlepas apakah seseorang memiliki faktor risiko lain seperti tekanan darah tinggi atau diabetes (Samiadi, 2017). Indeks massa tubuh alias BMI membandingkan berat badan dengan tinggi badan, dihitung dengan membagi berat badan dalam kilogram dengan tinggi badan dalam meter kuadrat (Quamila, 2017).

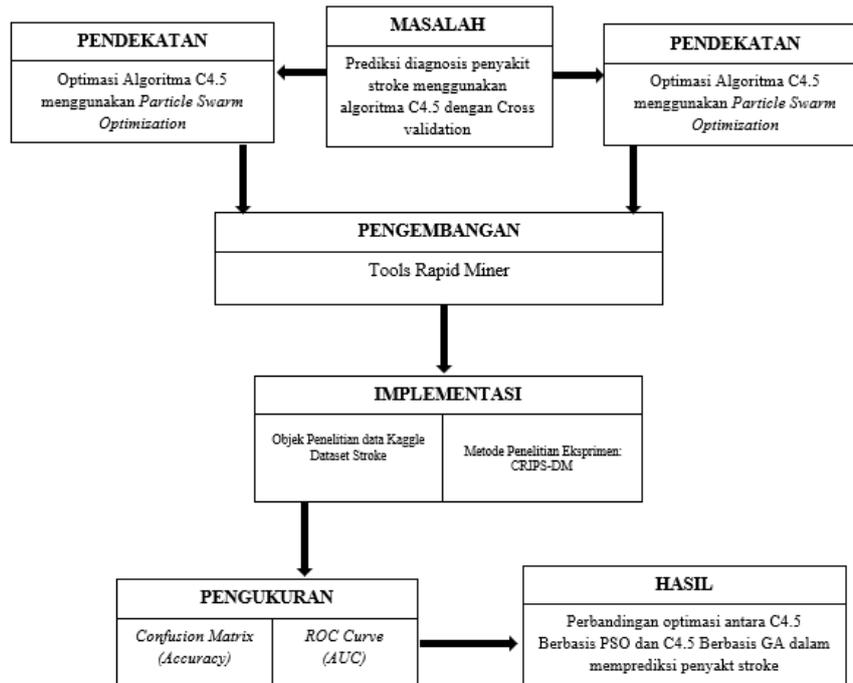
$$BMI = \frac{(\text{weight in kilograms})}{\text{height in meters}^2} \quad (1)$$

10. *Smoking Status*
Salah satu atribut yang mengindikasikan status perokok/bukan perokok dari pasien dengan kategori :
 - a. Formerly smoked (Sebelumnya merokok)
 - b. Smoked (Merokok)
 - c. Never Smoked (Tidak pernah merokok)
11. *Stroke*
Atribut ini merupakan atribut *predictor* atau penentu dengan kategori Suffered stroke (Menderita Stroke) dan No Stroke (Tidak Stroke)

D. Kerangka Pemikiran

Penelitian ini terdiri dari beberapa tahapan seperti terlihat pada gambar kerangka pemikiran 2.9. Permasalahan pada penelitian ini adalah belum begitu banyak penggunaan metode dalam memprediksi penyakit stroke serta belum dilakukannya komparasi algoritma C4.5 berbasis PSO dan C4.5 berbasis GA

Atas dasar alasan tersebut diatas, maka dilakukan penelitian menggunakan metode klasifikasi algoritma C4.5 berbasis PSO (*Particle Swarm Optimization*) dan juga C4.5 berbasis *Genetic Algorithm* dalam memprediksi penyakit stroke. Pengujian metode dilakukan dengan cara *confusion matrix* dan kurva ROC. Untuk mengembangkan aplikasi berdasarkan metode, digunakan *tools RapidMiner studio* serta didukung dengan penggunaan pembuatan aplikasi berbasis web. Berikut dibawah ini kerangka pemikiran yang penulis buat untuk penelitian ini:



Gbr1. Kerangka Pemikiran

E. Data Mining

Data mining adalah suatu disiplin ilmu yang bertujuan untuk menemukan, menggali atau menambahkan pengetahuan dari data atau informasi yang kita miliki[7][8]. Menurut Gartner Group menyebutkan bahwa Data Mining adalah proses menelusuri pengetahuan baru, pola dan tren yang dipilih dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan menggunakan teknik pengenalan pola serta statistik dan tehnik matematika [8].

F. Algoritma C4.5

Algoritma C4.5 merupakan salah satu solusi pemecahan kasus yang sering digunakan dalam pemecahan masalah pada teknik klasifikasi[9]. Pemilihan atribut sebagai simpul, baik simpul akar (*root*) atau simpul internal didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Penghitungan nilai *Gain* digunakan rumus seperti dalam Persamaan 1.

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

S : Himpunan kasus

A : Atribut

n : Jumlah partisi himpunan atribut A

|S_i| : Jumlah kasus pada partisi ke- i

|S| : Jumlah kasus dalam S

Untuk menghitung nilai *Entropy* dapat dilihat pada Persamaan 2

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2(p_i) \quad (3)$$

n : Jumlah partisi S

pi : Proporsi dari S_i terhadap S

G. Particle Swarm Optimization (PSO)

Teknik perhitungan evolusi yang dikembangkan oleh Kennedy dan Eberhart tahun 1995, PSO dikembangkan dari kecerdasan berkelompok dan didasarkan pada penelitian perilaku gerakan kawanan burung dan ikan. Sementara mencari makanan, burung-burung yang baik tersebar atau pergi bersama-sama sebelum mereka menemukan tempat di mana mereka dapat menemukan makanan. Sementara burung mencari makanan dari satu tempat ke tempat[10].

Secara garis besar, algoritma PSO dapat dijabarkan dalam rumus berikut :

$$v_i = w * v_i + r \text{ and } (0,1) * c_1 * (x_{pBest} - x_i) + r \text{ and } (0,1) * c_2 * (x_{gBest} - x_i) \quad (4)$$

dimana *v_i* merupakan kecepatan dari partikel ke-*i*, *w* merupakan parameter inertia weight, *c₁* merupakan parameter cognitive factor, *c₂* merupakan parameter social factor, *x_i* adalah koordinat posisi dari partikel ke-*i* yang melambangkan kandidat solusi dari suatu permasalahan, *x_{pBest}* adalah koordinat posisi dari *pBest*, *x_{gBest}* adalah koordinat posisi dari *gBest*.

Modifikasi kecepatan dan posisi tiap partikel dapat dihitung menggunakan kecepatan saat ini dan jarak *pbest_{i,d}* ke *gbest_d* seperti ditunjukkan pada persamaan berikut:

$$v_{i,m} = w.v_{i,m} + c_1 * R * (pbest_{i,m} - x_{i,m}) + c_2 * R * (gbest_m - x_{i,m}) \quad (5)$$

$$x_{i,d} = x_{i,m} + v_{i,m} \quad (6)$$

Dimana:

n : jumlah partikel dalam kelompok

d : dimensi

v_{i,m} : kecepatan partikel ke-*i* pada iterasi ke-*i*

w : faktor bobot *inersia*
 c_1, c_2 : konstanta akselerasi (*learning rate*)
 R : bilangan *random* (0-1)
 $x_{i,d}$: posisi saat ini dari partikel ke- i pada iterasi ke- i
 $pbest_i$: posisi terbaik sebelumnya dari partikel ke- i
 $gbest$: partikel terbaik diantara semua partikel dalam satu kelompok atau populasi.

H. Genetic Algorithm(GA)

Genetic Algorithm atau Algoritma Genetika adalah salah satu tipe algoritme evolusi yang populer dan sudah banyak digunakan untuk masalah-masalah kompleks pada berbagai bidang. Algoritme Genetika bekerja dengan sebuah populasi yang terdiri dari suatu individu-individu yang dimana masing-masing dapat merepresentasikan sebuah solusi yang mungkin bagi persoalan yang ada[4].

Berikut merupakan parameter dan siklus dari algoritma genetika[4]:

a. Parameter Algoritma Genetika

Dalam Algoritme Genetika terdapat beberapa parameter antara lain:

1. Ukuran Populasi (Popsize)
2. Tingkat pindah silang atau Crossover rate (Cr)
3. Tingkat Mutasi atau Mutation Rate (Mr)
4. Kondisi berhenti atau stop

b. Siklus dalam Algoritme Genetika yaitu :

1. Inisialisasi
2. Reproduksi (Crossover, Mutasi)
3. Evaluasi (Fitness & Offspring)
4. Populasi generasi selanjutnya

Pendekatan yang diambil oleh algoritma ini adalah dengan menggabungkan secara acak berbagai pilihan solusi terbaik di dalam suatu kumpulan untuk mendapatkan generasi solusi terbaik berikutnya yaitu pada suatu kondisi yang memaksimalkan kecukannya atau lazim disebut fitness[11].

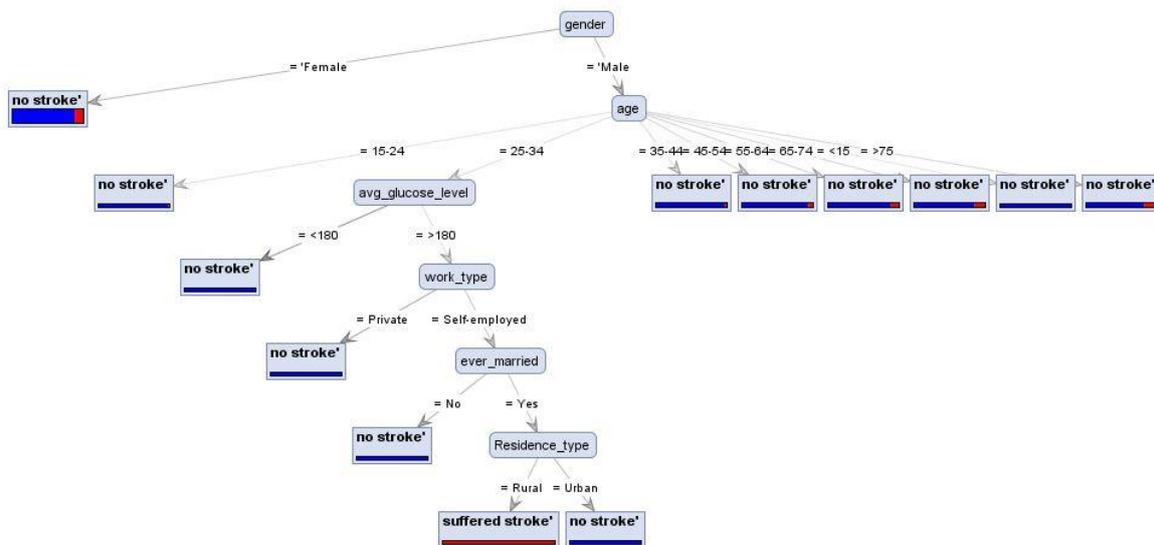
III. PEMBAHASAN

Penelitian ini bertujuan untuk menerapkan dan membandingkan *Particle Swarm Optimizing* dan *Genetic Algorithm* dalam meningkatkan tingkat akurasi prediksi penyakit stroke dengan menggunakan metode C4.5. Hasil dari penelitian ini sendiri berupa hasil pengolahan data kualitatif dan kuantitatif dengan perhitungan berdasarkan model yang diusulkan. Penelitian dilakukan pada sebuah dataset yang sudah tersedia yang bersifat publik. Eksperimen dan pengujian dalam penelitian ini dilakukan dengan menggunakan metode C4.5 tanpa PSO atau GA dan metode menggunakan PSO atau GA.

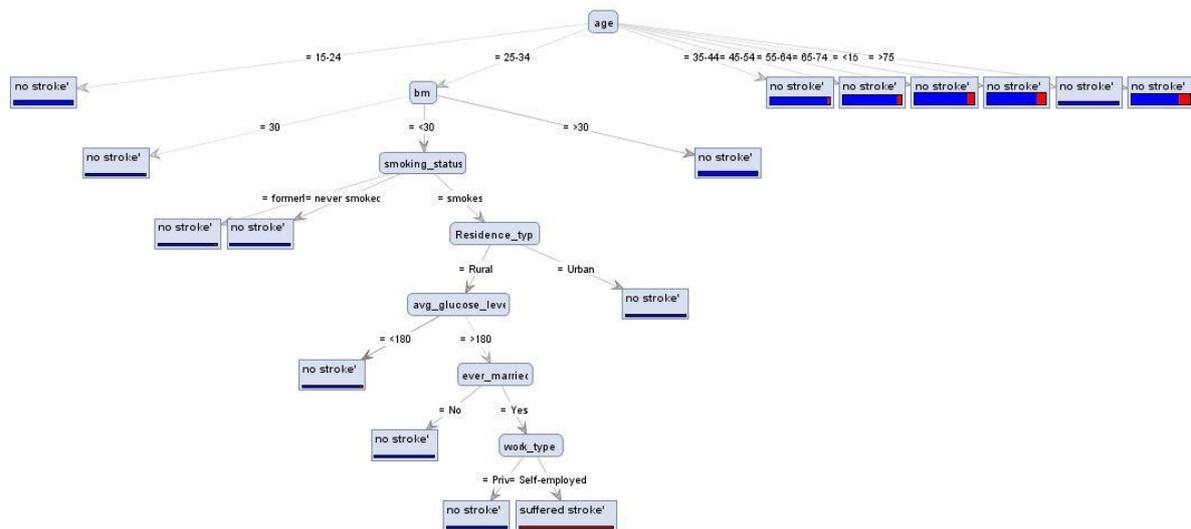
1. Eksperimen dan pengujian model C4.5

Pengelolaan data menggunakan model C4.5 dilakukan pada dataset yang terdiri dari 11 atribut yang merupakan atribut dari diagnosis penyakit stroke dan class yang merupakan hasil akhir prediksi.

Model dari algoritma C4.5 yaitu berupa pohon keputusan, agar lebih mudah dalam membuat pohon keputusan, langkah pertama adalah menghitung jumlah *class* yang berpotensi terkena penyakit stroke dan tidak stroke dari masing-masing *class* berdasarkan atribut yang telah ditentukan dengan menggunakan data *training*.



Gbr 2. Pohon Keputusan Klasifikasi Penyakit Stoke Menggunakan Algoritma C4.5



Gbr 3. Pohon Keputusan Klasifikasi Penyakit Stoke PSO-C4.5

Dari pohon keputusan tersebut maka diperoleh aturan-aturan atau *rule* sebagai berikut:

- R1 : if Age = 15-24 then class = no Stroke
- R2 : if Age = 25-34 and bmi = 30 then class = no stroke
- R3 : if Age = 25-34 and bmi <30 and smoking status=former/ never smoker then class = no stroke
- R4 : if Age = 25-34 and bmi <30 and smoking status=Never smoked then class = no stroke
- R5 : if Age = 25-34 and smoking status = smokes and Residence Type = Rural and Avg Glukosa <=180 then class no stroke
- R6 : if Age = 25-34 and Avg Glukosa >=180 and ever married = No then class no stroke
- R6 : if Age = 25-34 and ever Married = Yes and Work Type = Private then class no stroke
- R7 : if Age = 25-34 and Work Type = Self Employed then class Suffered Stroke
- R8 : if Age = 25-34 and Residence Type = Urban then class No Stroke
- R9 : if Age = 25-34 and bmi => 30 then class No Stroke
- R10 : if Age = 35-44 then class = no stroke
- R11 : if Age = 45-54 then class = no stroke
- R12 : if Age = 55-64 then class = no stroke
- R13 : if Age = 65-74 then class = no stroke
- R14 : if Age = <15 then class = no stroke
- R15 : if Age = >75 then class = no stroke

Terdapat 16 rule yang dihasilkan dari pohon keputusan algoritma klasifikasi C4.5, dengan jumlah class *no stroke* sebanyak 15 rule dan 1 rule untuk class *Suffered Strokes*. Sedangkan atribut yang terseleksi dari 11 atribut menjadi 9 atribut yang terdiri dari Age, Heart disease, Ever Married, work Type, Avg Glukose Level, BMI, Residence Status, Smoking Status, Stroke.

2. Komparasi Model Algoritma C4.5 dengan Algoritma C4.5 – PSO dan C4.5 – GA

TABLE II
PERBANDINGAN AKURASI DAN NILAI AUC

| | Accuracy | AUC |
|----------|----------|-------|
| C4.5 | 89,93% | 0.610 |
| C4.5-PSO | 91,63% | 0.500 |
| C4.5-GA | 92,02% | 0.500 |

Hasil dari pengujian diatas menjelaskan bahwa pengujian C4.5 menjadi lebih baik ketika menggunakan optimasi *Particle Swarm Optimization* dan *Genetik Algorithm*. Terbukti dengan menggunakan C4.5 berbasis PSO maka akurasi yang diperoleh sebesar 91,63% dibanding dengan C4.5 saja yang hanya mendapatkan 89,93%. Namun meskipun demikian tingkat akurasi yang didapatkan jauh lebih tinggi menggunakan GA dibanding dengan menggunakan PSO dengan tingkat akurasi sebesar 92,02%

3. Pengembangan Aplikasi Model dari Algoritma C4.5

Berdasarkan hasil eksperimen didapatkan model terbaik dari metode algoritma C4.5 dan algoritma C4.5 dan *Particle Swarm Optimization*, yang dikembangkan menjadi sebuah aplikasi. Berikut merupakan rancangan antarmuka aplikasi dalam memprediksi penyakit stroke :

Gbr 4 Pengujian Aplikasi Model Algoritma C4.5 dan Particle Swarm Optimization

IV. PENUTUP

Dari pembahasan-pembahasan diatas dapat ditarik kesimpulan bahwa Penelitian dengan menerapkan *particle swarm optimization* dan *Genetic Algorithm* pada algoritma C4.5 dilakukan agar dapat diketahui seberapa besar tingkat akurasi yang dihasilkan dalam mendiagnosa penyakit stroke.

Dalam menyeleksi atribut dibutuhkan suatu algoritma *particle swarm optimization* pada C4.5, untuk proses pembelajarannya dieksperimenkan pada Rapidminer. Sedangkan proses pengujiannya menggunakan *Accuracy* dan *ROC Curve* untuk melihat tingkat keakuratan prediksi diagnosis penyakit hepatitis. Atribut yang digunakan yaitu *gender, age, hypertension, heart disease, ever married, work type, avg glukose level, bmi, residence status, smoking status, stroke*.

Dari 11 atribut yang terdapat pada dataset *kaggle*, kemudian selanjutnya diseleksi menjadi hanya 10 atribut yang digunakan dalam menentukan prediksi penyakit hepatitis, atribut-atribut tersebut yaitu : *gender, age, hypertension, heart disease, ever married, work type, avg glukose level, bmi, residence status, stroke*.

Sehingga dapat disimpulkan bahwa penerapan teknik optimasi *particle swarm optimization* mampu menyeleksi atribut pada C4.5, sehingga menghasilkan tingkat akurasi diagnosis penyakit hepatitis yang lebih baik dibanding dengan menggunakan metode individual algoritma C4.5.

REFERENSI

- [1] LF, N. (2013). Stroke Non Hemoragik Pada Laki-Laki Usia 65 Tahun . *Medula*, 1-9.
- [2] Widowati, U. (2015, Mei 14). Berita Kesehatan : 10 Penyakit Paling Mematikan di Indonesia . Retrieved juni 29, 2018, from CNN Indonesia:

- <https://www.cnnindonesia.com/gayahidup/20150513163407-255-53129/10-penyakit-paling-mematikan-di-indonesia>.
- [3] Albirra, F. (2017, November 21). *Kesehatan : Inilah Penyakit yang Paling Banyak Menyerang Masyarakat Indonesia*. Retrieved Juni 29, 2018, from Jawa Pos: <https://www.jawapos.com/kesehatan/21/11/2017/inilah-penyakit-yang-paling-banyak-menyerang-masyarakat-indonesia>
- [4] Adelina, V., Ratnawati, D. E., & Fauzi, M. (2018).Klasifikasi Tingkat Risiko Penyakit Stroke Menggunakan Metode GA-Fuzzy Tsukamoto . *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3015-3021.
- [5] Fauzi, A., & Suharjo, B. (2016). Pengaruh Sumber Daya Finansial, Aset Tidak Berwujud dan Keunggulan Bersaing yang Berimplikasi Terhadap Kinerja Usaha Mikro, Kecil dan Menengah di Lombok NTB. *Manajemen IKM*, 151-158.
- [6] Lestari , Y. A., & Nuzula , N. F. (2017). Analisis pengaruh financial leverage dan operating leverage terhadap profitabilitas perusahaan. *Jurnal Administrasi Bisnis*, 1-10.
- [7] Susanto, S., & Suryadi, D. (2010). Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data. Yogyakarta: C.V ANDI OFFSET.
- [8] Saputra, R. A. (2014). Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (Tb): Studi Kasus Puskesmas Karawang Sukabumi. In *Seminar Nasional Inovasi dan Tren (SNIT)* (pp. 1-8).
- [9] Nofriansyag, D & Nurcahyo, G.W.(2015). Algoritma Data Mining Dan Pengujian. Yogyakarta. Deepublish.
- [10] Ramdhani, L. S. (2016). Penerapan Particle Swarm Optimization (PSO) untuk Seleksi Atribut dalam Meningkatkan Akurasi Prediksi Diagnosis Penyakit Hepatitis dengan Metode Algoritma C4. 5. *Swabumi*, 4(1), 1-15.
- [11] Nugroho, D. P. (2015). Optimasi Solusi Permasalahan Rute Kendaraan Dengan Pemerataan Beban Menggunakan Genetic Algorithm. *Jurnal Penelitian Transportasi Multimoda*, 1-10.