

## PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA C4.5 DALAM PREDIKSI PENYAKIT ANGIN DUDUK

Salman Alfaridzi<sup>1</sup>, Agung Nugroho<sup>2</sup>, Muhammad Rizki Sani<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Universitas Pelita Bangsa

<sup>1)</sup>[salmanalfaridzi1107@gmail.com](mailto:salmanalfaridzi1107@gmail.com), <sup>2)</sup>[agung@pelitabangsa.ac.id](mailto:agung@pelitabangsa.ac.id), <sup>3)</sup>[m.rizki.sani@gmail.com](mailto:m.rizki.sani@gmail.com)

### ABSTRAK

Penyakit angin duduk (*Angina Pectoris*) merupakan penyakit yang terjadi karena gangguan pada aliran darah menuju jaringan otot jantung yang menyebabkan terjadinya nyeri pada dada. Angin duduk terjadi karena adanya penyempitan pembuluh coroner yang menyebabkan suplai oksigen untuk otot jantung mengalami gangguan sehingga jantung tidak dapat memompa darah dengan maksimal. Kurangnya pengetahuan masyarakat dalam mendeteksi gejala penyakit ini maka dengan memanfaatkan data tersebut penulis ingin menerapkan salah satu teknik data mining dalam melakukan prediksi atau mendiagnosis penyakit angin duduk (*angina pectoris*). Metode yang digunakan adalah *Algoritma C4.5* dan *Particle Swarm Optimization (PSO)* dengan alat bantu *RapidMiner* dengan menggunakan sebanyak 200 data. Hasil analisis menunjukkan bahwa gejala kolesterol, diabetes, hipertensi, obesitas dan merokok bisa menjadi indikator untuk mendiagnosis penyakit angin duduk (*angina pectoris*). Hasil nilai yang didapatkan dari penelitian ini yaitu nilai *Accuracy* yang didapatkan meningkat sebanyak 7,5% dari 76,50% menjadi 84,00%, nilai *Precision* yang didapatkan meningkat sebanyak 7,64% dari 80,50% menjadi 88,14%, dan nilai *Recall* yang didapatkan meningkat sebanyak 9% dari 72,00% menjadi 81,00%.

Kata Kunci: *Data Mining, Angin Duduk, Algoritma C4.5, Particle Swarm Optimization, RapidMiner*

### ABSTRACT

Sitting wind disease (*Angina Pectoris*) is a disease that occurs due to disruptions in blood flow to heart muscle tissue that causes chest pain. Wind sitting occurs due to a narrowing of the coroner vessels that cause the oxygen supply to the heart muscle to be disrupted so that the heart cannot pump blood optimally. Lack of public knowledge in detecting the symptoms of this disease then by utilizing the data the author wants to apply one of the data mining techniques in predicting or diagnosing sitting wind disease (*angina pectoris*). The methods used are *Algorithm C4.5* and *Particle Swarm Optimization (PSO)* with *RapidMiner* tools using as much as 200 data. The results of the analysis showed that the symptoms of cholesterol, diabetes, hypertension, obesity and smoking could be indicators for diagnosing sitting wind disease (*angina pectoris*). The results of the value obtained from this study are that the accuracy value obtained increased by 7.5% from 76.50% to 84.00%, the precision value obtained increased by 7.64% from 80.50% to 88.14%, and the recall value obtained increased by 9% from 72.00% to 81.00%.

Kata Kunci: *Data Mining, Sitting Wind, C4.5 Algorithm, Particle Swarm Optimization, RapidMiner.*

## 1. Pendahuluan

Angin duduk adalah nyeri di dada karena minimnya suplai darah dan oksigen menuju ke jantung. Angin duduk, atau yang dalam bahasa medisnya dikenal sebagai *angina* merupakan penyakit yang marak terjadi di masyarakat. Siapapun bisa mengalami penyakit ini, baik laki-laki ataupun perempuan. Namun, risiko penyakit ini umumnya akan semakin meningkat seiring bertambahnya usia dan bisa menjadi gejala penyakit jantung koroner. Ketika ini terjadi, biasanya akan terjadi penumpukan kolesterol dan lemak (*plak*) di dalam arteri koroner jantung (*aterosklerosis*). Angin Duduk (*Angina*) juga dapat disebabkan oleh kejang otot di area arteri coroner [1]. Menurut ilmu kesehatan atau ilmu medis, tidak ada istilah medis atau penyakit angin duduk, karena di dalam ilmu kedokteran penyakit ini termasuk dalam kelompok penyakit tertentu yang gejalanya dapat dikenali dari keluhan pasien, hasil pemeriksaan laboratorium atau ditelusuri dari jejak riwayat terdahulu penyakit yang diderita oleh pasien maupun keluarganya [2].

Pada penelitian yang dilakukan oleh Bayu Sugara, Dany Widyatmoko, Bobby Suryo Prakoso Dan Doddy Mulyadi Saputro untuk prediksi penyakit autisme pada anak dengan metode *Algoritma C4.5* karena dapat memberikan prediksi gangguan autisme pada anak berdasarkan gejala yang dialami anak, dan dapat diperoleh hasil nilai akurasi sebesar 72% [3]. Lalu penelitian selanjutnya yang dilakukan oleh Ramdhan Saepul Rohman, Rizal Amegia Saputra Dan Dasya Arif Firmansaha untuk diagnosa penyakit stroke menggunakan metode *Algoritma C4.5* dengan alat bantu *RapidMiner* dengan menerapkan *particle swarm optimization* dan Genetic Algorithm karena metode ini yang paling banyak digunakan dan memberikan hasil yang akurat dengan nilai akurasi 99.07% [4]. Sedangkan penelitian selanjutnya oleh Ari Muzakir Dan Rika Anisa Wulandari untuk prediksi penyakit hipertensi kehamilan dengan metode *Algoritma C4.5* menggunakan teknik *Decision Tree* dapat memberikan hasil tingkat akurasi 92.6573% dan menghasilkan kesalahan (error) 7.3427% [5].

Salah satu wujud nyata dari upaya ini yaitu memberikan hasil prediksi kepada masyarakat akan gejala yang sering terjadi pada penyakit angin duduk ini agar bisa meminimalisir terjadinya hal-hal yang tidak diinginkan dan bisa membantu masyarakat dalam mengenali gejala yang sering terjadi.

Berdasarkan penjelasan di atas penelitian ini akan menggunakan metode Algoritma C4.5 karena merupakan salah satu algoritma decision tree yang tepat untuk klasifikasi dan *particle swarm optimization* (PSO) sebagai perbandingan hasil Accuracy, Precision dan Recall. Salah satu wujud nyata dari upaya ini yaitu memberikan hasil prediksi kepada masyarakat akan gejala yang sering terjadi pada penyakit angin duduk ini agar bisa meminimalisir terjadinya hal-hal yang tidak diinginkan.

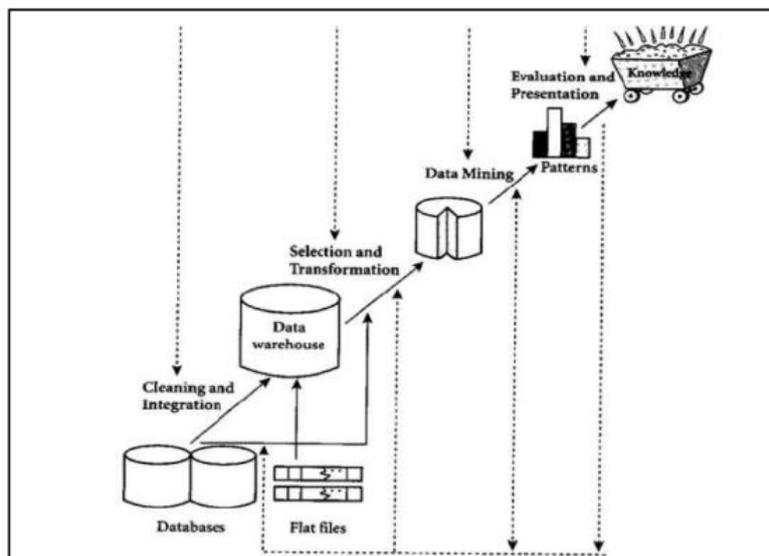
## 2. Metode Penelitian

Penelitian ini terdiri dari beberapa tahap. Masalah dengan penelitian ini adalah bahwa tidak ada algoritma yang akurat untuk mendiagnosis penyakit hati

inflamasi. Untuk alasan ini, pendekatan (model), algoritma C4.5 menggunakan, PSO digunakan [6]. Objek Penelitian yang dilakukan dalam penelitian ini mengenai penyakit angin duduk. Peneliti memilih objek penelitian ini karena masih awamnya pengetahuan pada umumnya masyarakat dalam mendeteksi atau mengetahui gejala awal yang diderita dari penyakit angin duduk.

## 2.1 Data Mining

Data mining adalah serangkaian proses yang digunakan untuk menemukan informasi dari data *warehouse* dengan menggunakan teknik statistika, matematika, kecerdasan buatan, dan lainnya untuk mengolah informasi mentah dari berbagai gudang data. Hasil data mining bisa berupa informasi pola/rumus atau hubungan khusus dari gudang data yang diteliti. Data mining juga digunakan untuk menggambarkan penemuan pengetahuan dalam database [7]. Informasi ini nantinya yang akan digunakan untuk sesuatu yang bermanfaat di masa depan. Data mining juga banyak yang menyebutnya sebagai Knowledge Discovery in Database (KDD) [8].



Gambar 2. 1 Tahapan Data Mining

## 2.2 Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma pohon keputusan yang paling cocok untuk klasifikasi. Algoritma C4.5 dilakukan dengan cara memisahkan data secara rekursif sehingga sebagian data berasal dari kelas yang sama. Dan yang membuat algoritma ini berbeda yakni, Algoritma C4.5 menggunakan kriteria split yang dimodifikasi yang disebut Gain Ratio dalam proses pemilihan split tiap atribut [9]. Algoritma ini memiliki input berupa data latih dan sampel. Contoh pelatihan data sampel yang digunakan untuk membangun pohon yang divalidasi. Sampel adalah bidang data yang nantinya akan digunakan sebagai parameter dalam klasifikasi data [10].

## 2.3 Particle Swarm Optimization

*Particle Swarm Optimization* (PSO) adalah metode optimasi yang memiliki proses sederhana dibandingkan algoritma genetika, yang terinspirasi dari

perilaku kawanan ikan dan unggas dalam mencari sumber makanannya [11]. Studi ini akan melakukan analisis komparatif dan fusi dari dua kategori teknik data mining [12]. Partikel merupakan setiap individu yang berada didalam suatu kelompok. Setiap partikel dalam *Particle Swarm Optimization* dapat menyesuaikan posisi dan kecepatannya masing-masing sehingga memiliki kecenderungan untuk berpindah tempat menuju posisi yang dianggap terbaik selama proses pencarian [13].

Adapun rumusnya sebagai beriku [14]t:

$$V_i(t) = V_i(t - 1) + c_1r_1 [Xpbest_i - X_i(t)]$$

$$+ c_2r_2 [Xgbest - X_i(t)]$$

$$X_i(t) = X_i(t - 1) + V_i(t)$$

#### 2.4 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder kualitatif yang didapatkan dari proses download yang diperoleh dari situs web <https://www.kaggle.com/> dataset yang di upload di kaggle dengan link <https://www.kaggle.com/kagglersr/angina> data ini berjumlah 200 data dan 9 atribut.

Tabel 2. 1 Dataset Angin Duduk

| Status | Age | Smoke | Cig | Hyper | Fam<br>ang | Fam<br>myo | Fam<br>stroke | Diabetes |
|--------|-----|-------|-----|-------|------------|------------|---------------|----------|
| No     | 37  | 1     | 15  | 1     | No         | No         | No            | No       |
| No     | 45  | 3     | 0   | 1     | No         | No         | No            | No       |
| Yes    | 60  | 3     | 0   | 1     | No         | Yes        | No            | No       |
| No     | 57  | 3     | 0   | 1     | No         | No         | No            | No       |
| yes    | 65  | 1     | 20  | 2     | No         | yes        | No            | No       |
| Yes    | 56  | 1     | 20  | 1     | No         | No         | No            | No       |
| No     | 42  | 3     | 0   | 1     | No         | No         | No            | No       |
| Yes    | 52  | 1     | 30  | 1     | No         | No         | No            | No       |
| Yes    | 61  | 2     | 0   | 1     | No         | Yes        | No            | No       |
| yes    | 59  | 2     | 0   | 1     | No         | Yes        | No            | No       |
| No     | 49  | 2     | 0   | 1     | No         | No         | Yes           | No       |
| No     | 56  | 2     | 0   | 1     | No         | No         | Yes           | No       |
| No     | 38  | 3     | 0   | 1     | No         | No         | No            | No       |
| Yes    | 66  | 1     | 12  | 1     | No         | Yes        | No            | No       |
| No     | 49  | 3     | 0   | 1     | No         | No         | No            | No       |
| Yes    | 48  | 1     | 12  | 1     | Yes        | No         | No            | No       |
| Yes    | 61  | 1     | 10  | 1     | No         | Yes        | No            | No       |
| Yes    | 57  | 1     | 30  | 1     | Yes        | Yes        | No            | No       |
| No     | 52  | 1     | 0   | 1     | No         | No         | No            | No       |
| No     | 43  | 3     | 0   | 1     | No         | No         | No            | No       |

Tabel 2. 2 Atribut Dataset

|    |        |                      |
|----|--------|----------------------|
| a. | Status | : Status ( yes, no ) |
| b. | Age    | : Usia ( 31-73 )     |

|    |            |  |
|----|------------|--|
| c. | Smoke      | : Merokok ( 1:Saat ini, 2:Mantan, 3:Sedang)                                    |
| d. | Cig        | : Rata-rata merokok perhari  |
| e. | Hyper      | : Hipertensi ( 1:Tidak ada, 2:Ringan, 3:Sedang )                               |
| f. | Fam ang    | : Riwayat keluarga yang terkena penyakit angina duduk ( yes, no )              |
| g. | Fam myo    | : Riwayat keluarga dengan <i>infark miokard</i> "serangan jantung" ( yes, no ) |
| h. | Fam stroke | : Riwayat keluarga terkena stroke ( yes, no )                                  |
| i. | Diabetes   | : Apakah menderita penyakit diabetes ( yes, no )                               |

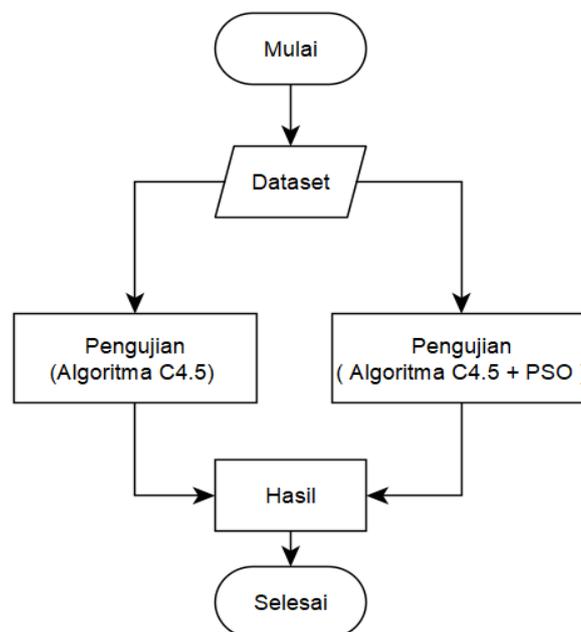
## 2.5 Teknik Analisa Data

Dalam penelitian ini, penulis menggunakan metode *Cross Industry Standar Process for Data Mining* (CRISP-DM) yang digunakan untuk pemrosesan data mining, CRISP-DM memiliki keunggulan dibandingkan dengan metode lainnya karena setiap fase terstruktur dan jelas sehingga mudah untuk dipahami dan diterapkan. Tahap-tahap CRISP-DM [15].

Tahapan metode ini meliputi fase pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi dan penyebaran. Karena dataset yang digunakan ini dataset jadi yang sudah melalui processing data, jadi tidak lagi perlu melakukan processing data.

Teknik pengujian data mining yang dipakai dalam mengolah data yang telah disiapkan sebelumnya yaitu dengan teknik *Algoritma C4.5 Cross Validasion* dan menambahkan *Particle Swarm Optimization* (PSO) sebagai perbandingan [6].

Berikut ini adalah pemodelan penelitian :



Gambar 2. 2 Tahap Pemodelan

Penjelasan tahap pemodelan diatas:

1. Dataset  
Pada tahap ini dilakukan persiapan dataset yang akan dipakai dalam penelitian ini. Dataset yang dipakai ini merupakan dataset public yang didapat dari kaggle.com.
2. Pengujian  
Pada tahap ini dilakukan proses pengujian untuk mengetahui hasil dari metode *Algoritma C4.5* dan *Particle Swarm Optimization (PSO)* dengan analisa *Cross validation* K=10 pengulangan menggunakan alat bantu *RapidMiner*.
3. Hasil  
Pada tahap ini yaitu membandingkan nilai *Accuracy*, *Precision* dan *Recall*, antara *Algoritma C4.5* dengan *Algoritma C4.5 + PSO*.

### 3. Hasil Dan Analisis

Tahap Proses pengujian penelitian ini dengan menerapkan metode *Algoritma C4.5* dan *Particle Swarm Optimization (PSO)* dengan menggunakan analisa *Cross Validation*. Proses pengujian metode dilakukan sebanyak 2 kali, untuk mengetahui hasil perbandingan antara *Accuracy*, *Precision* dan *Recall* dari penggabungan metode. Pada proses pengujian pertama dilakukan hanya dengan menggunakan metode *Algoritma C4.5* dan pada proses pengujian kedua menggabungkan antara metode *Algoritma C4.5* dengan *Particle Swarm Optimzation (PSO)*.

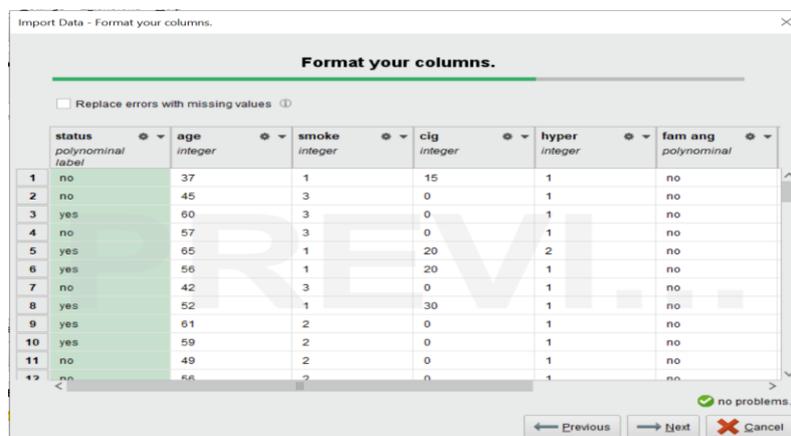
#### 3.1 Pengujian Pertama

Pengujian dilakukan dengan menerapkan metode *Algoritma C4.5* dengan analisa *Cross Validation* tanpa *Particle Swarm Optimization (PSO)* dengan menggunakan 200 data. Proses pengujian dilakukan dengan menggunakan alat bantu *RapidMiner*.

Berikut proses pengujiannya:

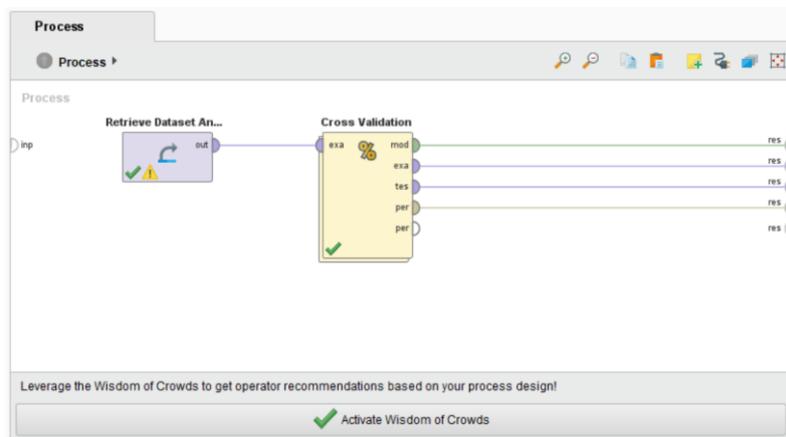
1. Import Data

Data yang sudah diberikan label berjumlah 200 data di import kedalam alat bantu *RapidMiner*.



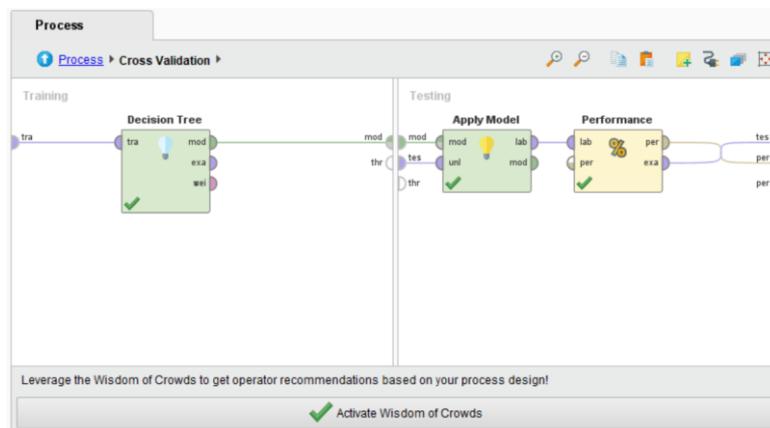
Gambar 2. 3 Import Dataset

## 2. Proses Pengujian



Gambar 2. 4 Proses Algoritma C4.5

Gambar diatas menunjukkan serangkaian model pengujian dimana pada dataset disesuaikan dengan dataset angin duduk. Metode validasi silang digunakan untuk memvalidasi model Algoritma C4.5. Dengan menggunakan *Confussion Matrix* sebagai model penilaian untuk menentukan *Performance* dari *Algoritma C4.5*. penjelasannya dapat dilihat pada gambar berikut:



Gambar 2. 5 Proses Cross Validation

Pada proses *Cross Validation* terdapat dua bagian yaitu ada data training 90% dan data testing 10% dengan 10 kali percobaan, dimana pada bagian training terdapat *Algoritma C4.5 (Decision Tree)*. Kemudian pada data testing terdapat dua fungsi yaitu *Apply model* yang digunakan untuk menerapkan model data yang sebelumnya dilatih dengan data uji. Terakhir, pada bagian testing terdapat fitur *Performance* dimana fungsi ini digunakan untuk mengevaluasi hasil kinerja dari *Algoritma C4.5* dengan parameter pengukuran *Confussion Matrix (accuracy, recall, precision)*.

## 3. Hasil ExampleSet

Dibawah ini merupakan gambar hasil exampleset dari pengujian 200 data dengan menggunakan metode *Algoritma C4.5* pada alat bantu *RapidMiner*.

| Row No. | status | age | smoke | cig | hyper | fam ang | fam myo | fa |
|---------|--------|-----|-------|-----|-------|---------|---------|----|
| 1       | no     | 37  | 1     | 15  | 1     | no      | no      | no |
| 2       | no     | 45  | 3     | 0   | 1     | no      | no      | no |
| 3       | yes    | 60  | 3     | 0   | 1     | no      | yes     | no |
| 4       | no     | 57  | 3     | 0   | 1     | no      | no      | no |
| 5       | yes    | 65  | 1     | 20  | 2     | no      | yes     | no |
| 6       | yes    | 56  | 1     | 20  | 1     | no      | no      | no |
| 7       | no     | 42  | 3     | 0   | 1     | no      | no      | no |
| 8       | yes    | 52  | 1     | 30  | 1     | no      | no      | no |
| 9       | yes    | 61  | 2     | 0   | 1     | no      | yes     | no |
| 10      | yes    | 59  | 2     | 0   | 1     | no      | yes     | no |
| 11      | no     | 49  | 2     | 0   | 1     | no      | no      | ye |
| 12      | no     | 56  | 2     | 0   | 1     | no      | no      | ye |

Gambar 2. 6 ExampleSet Algoritma C4.5

#### 4. Hasil Pengujian

Pengujian di lakukan dengan menggunakan teknik analisa *Cross Validation* dengan nilai  $k=10$ . Dibawah ini merupakan hasil *Confusion Matrix* dengan menggunakan metode *Algoritma C4.5* pada alat bantu *RapidMiner*.

##### a. Accuracy

Hasil pengukuran menunjukkan bahwa nilai *Accuracy* dari metode *Algoritma C4.5* sebesar 76.50%.

accuracy: 76.50% +/- 8.51% (micro average: 76.50%)

|              | true no | true yes | class precision |
|--------------|---------|----------|-----------------|
| pred. no     | 81      | 28       | 74.31%          |
| pred. yes    | 19      | 72       | 79.12%          |
| class recall | 81.00%  | 72.00%   |                 |

Gambar 2. 7 Accuracy Algoritma C4.5

##### b. Precision

Hasil pengukuran menunjukkan bahwa nilai *Precision* dari metode *Algoritma C4.5* sebesar 80.50%.

precision: 80.50% +/- 11.06% (micro average: 79.12%) (positive class: yes)

|              | true no | true yes | class precision |
|--------------|---------|----------|-----------------|
| pred. no     | 81      | 28       | 74.31%          |
| pred. yes    | 19      | 72       | 79.12%          |
| class recall | 81.00%  | 72.00%   |                 |

Gambar 2. 8 Precision Algoritma C4.5

##### c. Recall

Hasil pengukuran menunjukkan bahwa nilai *Recall* dari metode *Algoritma C4.5* sebesar 72.00%.

recall: 72.00% +/- 16.87% (micro average: 72.00%) (positive class: yes)

|              | true no | true yes | class precision |
|--------------|---------|----------|-----------------|
| pred. no     | 81      | 28       | 74.31%          |
| pred. yes    | 19      | 72       | 79.12%          |
| class recall | 81.00%  | 72.00%   |                 |

Gambar 2. 9 Recall Algoritma C4.5

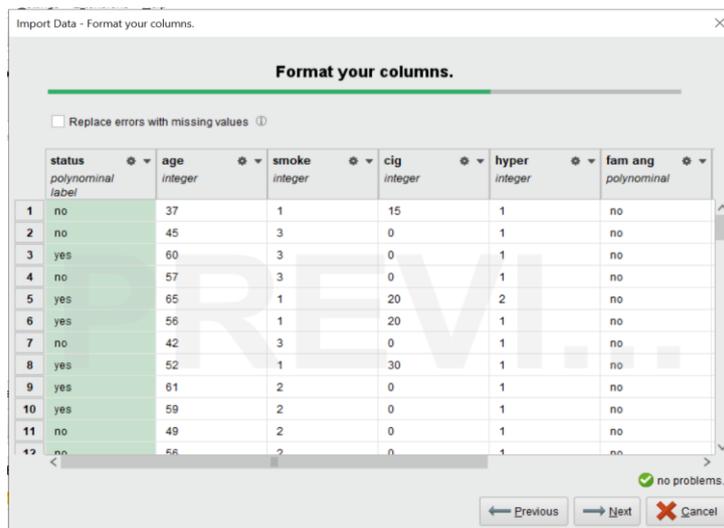
### 3.2 Pengujian kedua

Pengujian ini dilakukan dengan menggunakan metode *Algoritma C4.5* dengan *Particle Swarm Optimization* (PSO) dan analisa *Cross Validation*, dengan menggunakan 200 data. Proses pengujian dilakukan dengan menggunakan alat bantu RapidMiner.

Berikut proses pengujiannya:

#### 1. Import Data

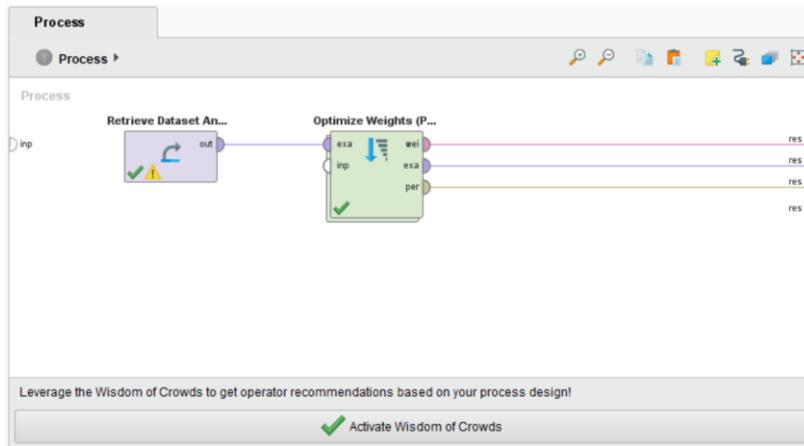
Data yang digunakan pada pengujian kedua ini sama dengan data pengujian pertama, yang sudah diberikan label dan berjumlah 200 data.



Gambar 2. 10 Import Dataset

#### 2. Proses Pengujian

Pada proses pengujian kedua ini menggabungkan metode *Algoritma C4.5* dengan *Particle Swarm Optimization* (PSO).

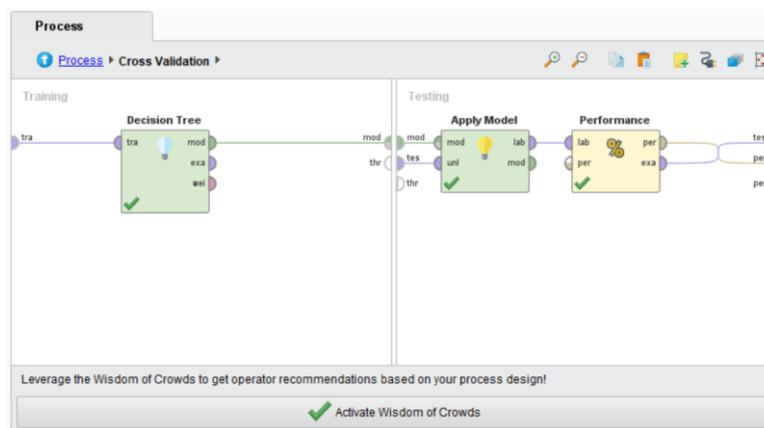


Gambar 2. 11 Proses Algoritma C4.5 + PSO

| Process Parameters            |     |
|-------------------------------|-----|
| <b>Optimize Weights (PSO)</b> |     |
| population size               | 5   |
| maximum number of generations | 30  |
| inertia weight                | 1.0 |
| local best weight             | 1.0 |
| global best weight            | 1.0 |
| min weight                    | 0.0 |
| max weight                    | 1.0 |

Gambar 2. 12 Parameter PSO

Gambar 2.11 diatas menunjukkan serangkaian model pengujian dimana pada dataset disesuaikan dengan dataset angin duduk. Metode validasi silang digunakan untuk memvalidasi model Algoritma C4.5 dengan PSO. Dengan menggunakan *Confussion Matrix* sebagai model penilaian untuk menentukan *Performance* dari *Algoritma C4.5*. penjelasannya dapat dilihat pada gambar berikut:



Gambar 2. 13 Proses Cross Validation di PSO

Pada proses *Cross Validation* di PSO terdapat dua bagian yaitu ada data training 90% dan data testing 10% dengan 10 kali percobaan, dimana pada bagian training terdapat *Algoritma C4.5 (Decision Tree)*. Kemudian pada data testing terdapat dua fungsi yaitu *Apply model* yang digunakan untuk menerapkan model data yang sebelumnya dilatih dengan data uji. Terakhir, pada bagian testing terdapat fitur *Performance* dimana fungsi ini digunakan untuk mengevaluasi hasil kinerja dari *Algoritma C4.5* dengan parameter pengukuran *Confussion Matrix (accuracy, recall, precision)*.

3. Hasil ExampleSet  
 Hasil dari proses pengujian data.

| Row No. | status | age | smoke | cig | hyper | fam ang | fam myo | fa |
|---------|--------|-----|-------|-----|-------|---------|---------|----|
| 1       | no     | 37  | 1     | 15  | 1     | no      | no      | no |
| 2       | no     | 45  | 3     | 0   | 1     | no      | no      | no |
| 3       | yes    | 60  | 3     | 0   | 1     | no      | yes     | no |
| 4       | no     | 57  | 3     | 0   | 1     | no      | no      | no |
| 5       | yes    | 65  | 1     | 20  | 2     | no      | yes     | no |
| 6       | yes    | 56  | 1     | 20  | 1     | no      | no      | no |
| 7       | no     | 42  | 3     | 0   | 1     | no      | no      | no |
| 8       | yes    | 52  | 1     | 30  | 1     | no      | no      | no |
| 9       | yes    | 61  | 2     | 0   | 1     | no      | yes     | no |
| 10      | yes    | 59  | 2     | 0   | 1     | no      | yes     | no |
| 11      | no     | 49  | 2     | 0   | 1     | no      | no      | ye |
| 12      | no     | 56  | 2     | 0   | 1     | no      | no      | ye |

Gambar 2. 14 Hasil ExampleSet Algoritma C4.5 + PSO

4. Attribute Weight  
 Hasil dari pembobotan nilai attribute.

| attribute  | weight |
|------------|--------|
| age        | 0.684  |
| smoke      | 0      |
| cig        | 0.669  |
| hyper      | 0      |
| fam ang    | 0.264  |
| fam myo    | 1      |
| fam stroke | 0.668  |
| diabetes   | 0.710  |

Gambar 2. 15 Attribute Weight

5. Hasil Pengujian  
 Pengujian di lakukan dengan menggunakan teknik analisa *Cross Validation* dengan nilai k=10. Dibawah ini merupakan hasil *Confusion*

Matrix dengan menggunakan metode Algoritma C4.5 dan *Particle Swarm Optimization* (PSO) pada alat bantu *RapidMiner*.

a. Accuracy

Hasil pengukuran menunjukkan bahwa nilai *Accuracy* dari metode Algoritma C4.5 dengan PSO sebesar 84.00%.

accuracy: 84.00% +/- 6.15% (micro average: 84.00%)

|              | true no | true yes | class precision |
|--------------|---------|----------|-----------------|
| pred. no     | 87      | 19       | 82.08%          |
| pred. yes    | 13      | 81       | 86.17%          |
| class recall | 87.00%  | 81.00%   |                 |

Gambar 2. 16 Accuracy Algoritma C4.5 + PSO

b. Precision

Hasil pengukuran menunjukkan bahwa nilai *Precision* dari metode Algoritma C4.5 dengan PSO sebesar 88.14%.

precision: 88.14% +/- 11.66% (micro average: 86.17%) (positive class: yes)

|              | true no | true yes | class precision |
|--------------|---------|----------|-----------------|
| pred. no     | 87      | 19       | 82.08%          |
| pred. yes    | 13      | 81       | 86.17%          |
| class recall | 87.00%  | 81.00%   |                 |

Gambar 2. 17 Precision Algoritma C4.5 + PSO

c. Recall

Hasil pengukuran menunjukkan bahwa nilai *Recall* dari metode Algoritma C4.5 dengan PSO sebesar 81.00%.

recall: 81.00% +/- 9.94% (micro average: 81.00%) (positive class: yes)

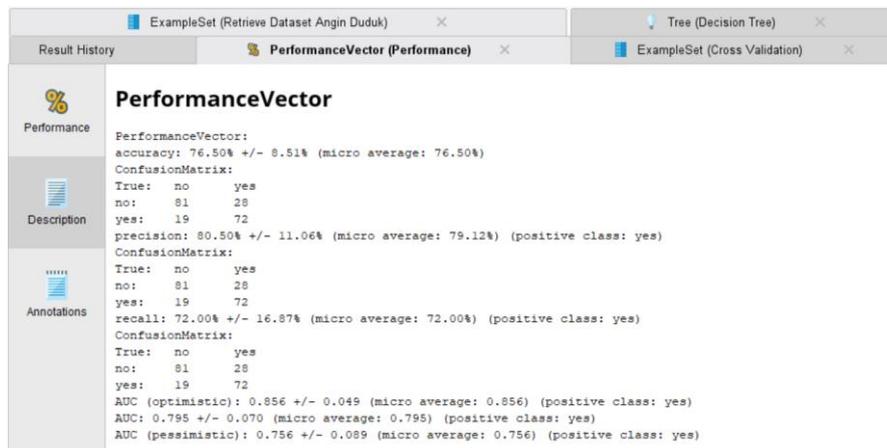
|              | true no | true yes | class precision |
|--------------|---------|----------|-----------------|
| pred. no     | 87      | 19       | 82.08%          |
| pred. yes    | 13      | 81       | 86.17%          |
| class recall | 87.00%  | 81.00%   |                 |

Gambar 2. 18 Recall Algoritma C4.5 + PSO

### 3.3 Performance Vector

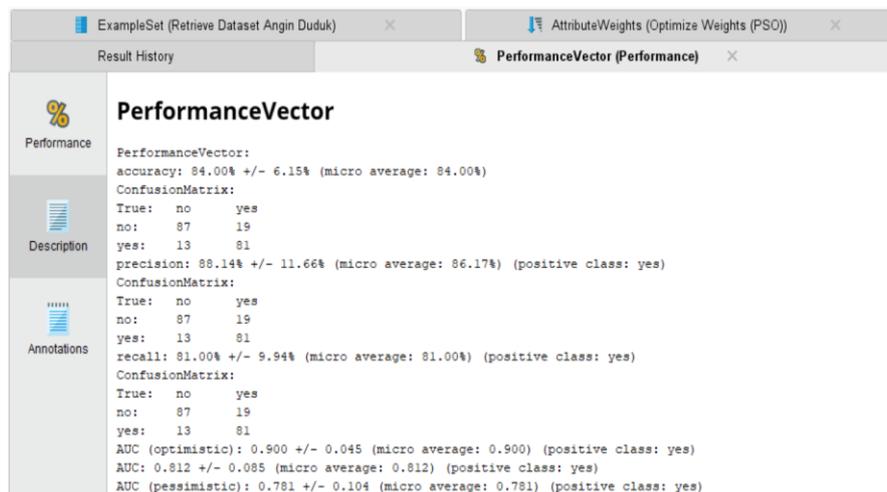
Perbandingan hasil Performance Vektor dari Algoritma C4.5 dengan Algoritma C4.5 dengan *Particle Swarm Optimzation* (PSO). Berikut ini hasil dari perbandingan keduanya:

## 1. Algoritma C4.5



Gambar 2. 19 Performance Vektor Algoritma C4.5

## 2. Algoritma C4.5 + PSO



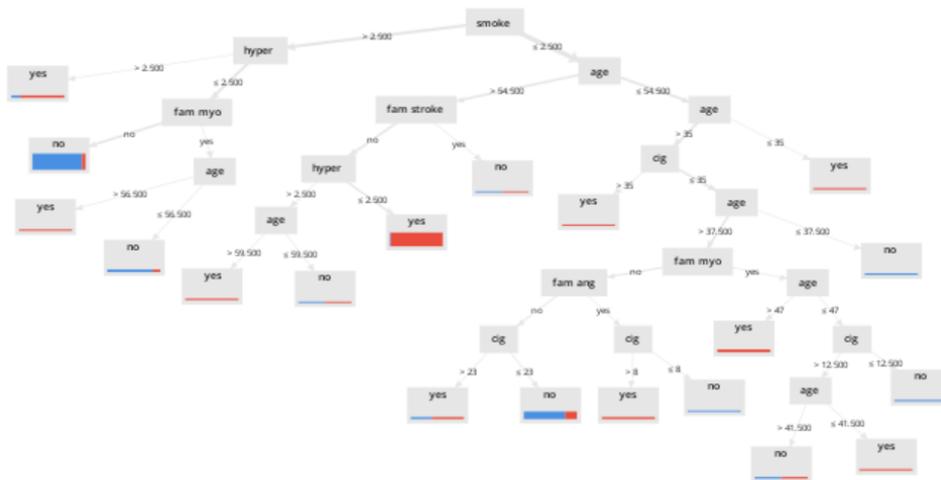
Gambar 2. 20 Performance Vektor Algoritma C4.5 + PSO

Dari hasil performanceVektor diatas menunjukkan bahwa Algoritma C4.5 dengan *Particle Swarm Optimzation* (PSO) lebih tinggi nilai *Accuracy*, *Precision* dan *Recall* nya.

### 3.4 Decision Tree

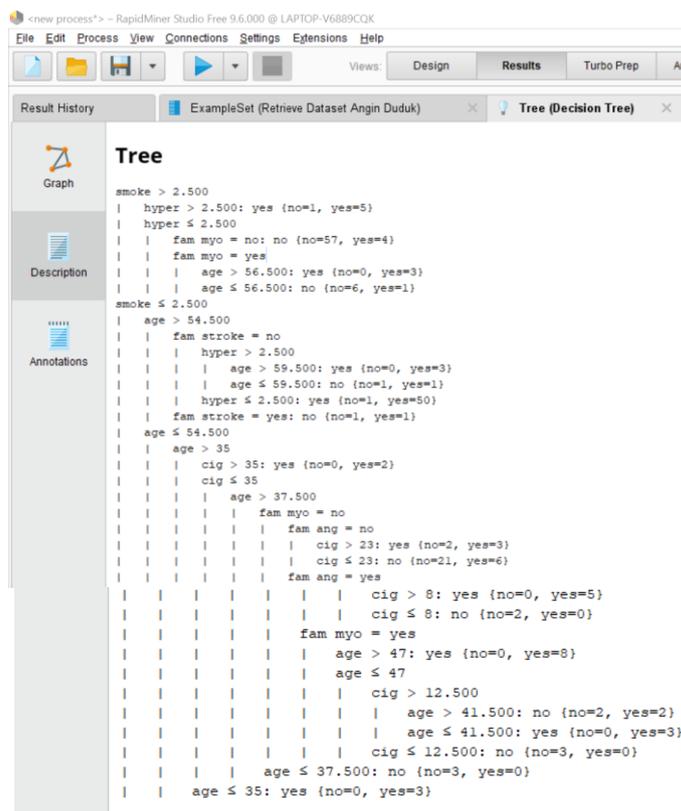
Dalam algoritma decision tree terdapat yang namanya pola pohon keputusan. Inilah pola pohon keputusan dari data angin duduk penulis. Dapat dilihat bahwa atribut yang paling berpengaruh adalah "Smoke". Dengan demikian nilai Smoke bertindak sebagai root node/akar, sementara atribut lainnya menjadi branches/cabang-cabang.

Berikut ini adalah gambar pola pohon keputusan dari penelitian penyakit angin duduk:



Gambar 2. 21 Pohon Keputusan

Untuk lebih jelas berikut deskripsi dari pohon keputusan *Decision Tree*:



Gambar 2. 22 Deskripsi Pohon Keputusan

### 3.5 Pembahasan

Hasil pengujian diatas menjelaskan bahwa pengujian Algoritma C4.5 semakin meningkat ketika saat menggunakan optimasi *Particle Swarm Optimization* (PSO). Berikut tabel perbandingan pengujian antara *Algoritma C4.5* tanpa *Particle Swarm Optimization* (PSO) dan pengujian menggunakan metode *Algoritma C4.5* dengan *Particle Swarm Optimization* (PSO).

Tabel 4. 1 Perbandingan Hasil Pengujian

| Confusion Matrix | Pengujian 1    | Pengujian 2        | Keterangan  |           |
|------------------|----------------|--------------------|-------------|-----------|
|                  | Algoritma C4.5 | Algoritma C4.5+PSO | Peningkatan | Hasil     |
| Accuracy         | 76,50%         | 84,00%             | 7.5%        | Meningkat |
| Precision        | 80,50%         | 88,14%             | 7,64%       | Meningkat |
| Recall           | 72,00%         | 81,00%             | 9%          | Meningkat |

#### 4. Penutup

Berdasarkan hasil penelitian yang telah dilakukan maka dapat disimpulkan bahwa:

1. Pengujian yang dilakukan dengan menggunakan teknik optimasi *Particle Swarm Optimization* (PSO) dapat menghasilkan tingkat akurasi yang lebih besar dalam mendiagnosa penyakit angin duduk dibandingkan dengan menggunakan metode individual Algoritma C4.5
2. Pengujian yang dilakukan dengan menggunakan *Particle Swarm Optimization* (PSO) pada metode *Algoritma C4.5* dalam menganalisis prediksi penyakit angin duduk yang sering terjadi di masyarakat dapat meningkatkan nilai *Accuracy*, *Precision*, dan *Recall*. Nilai *Accuracy* yang didapatkan meningkat sebanyak 7,5% dari 76,50% menjadi 84,00%, nilai *Precision* yang didapatkan meningkat sebanyak 7,64% dari 80,50% menjadi 88,14%, dan nilai *Recall* yang didapatkan meningkat sebanyak 9% dari 72,00% menjadi 81,00%.

##### 4.1 Saran

Adapun beberapa saran dan masukan dari penulis yang dapat digunakan pada penelitian selanjutnya, yaitu sebagai berikut:

1. Pada penelitian selanjutnya dapat mencoba memanfaatkan data–data dengan jenis data yang berbeda dan dengan jumlah data yang lebih banyak lagi, sehingga dapat menghasilkan tingkat *Accuracy*, *Precision* dan *Recall* yang lebih tinggi.
2. Menggunakan metode atau teknik analisa lain dalam upaya mencari yang terbaik dan yang berbeda supaya dapat diketahui perbandingan dari tingkat *Accuracy*, *Precision* dan *Recall* yang dihasilkan.
3. Penelitian selanjutnya di sarankan bisa diimplementasikan kedalam Aplikasi android ataupun website untuk memprediksi penyakit angin duduk agar mempermudah masyarakat dalam mengaksesnya.

#### REFRENSI

- [1] T. Dr. Willy, “Angin Duduk,” 2019. <https://www.alodokter.com/angin-duduk>.

- [2] A. Turmudi Zy and L. Adji Ardiansyah, "Implementasi Algoritma Naive Bayes Dalam Mendiagnosa Penyakit Angin Duduk," *Pelita Teknol.*, vol. 16, no. 1, pp. 52–65, 2021.
- [3] B. Sugara, D. Widyatmoko, B. S. Prakoso, and D. M. Saputro, "Penerapan Algoritma C4.5 untuk Deteksi Dini Autisme Pada Anak," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2018, no. Sentika, pp. 87–96, 2018.
- [4] R. S. Rohman, R. A. Saputra, and D. A. Firmansaha, "Komparasi Algoritma C4.5 Berbasis PSO Dan GA Untuk Diagnosa Penyakit Stroke," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 1, p. 155, 2020, doi: 10.24114/cess.v5i1.15225.
- [5] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016, doi: 10.15294/sji.v3i1.4610.
- [6] O. Pahlevi and I. Satriadi, "Optimasi Algoritma C4 . 5 dan Naïve Bayes Berbasis Particle Swarm Optimization Untuk Diagnosa Penyakit Peradangan Hati," vol. 2, no. 1, pp. 10–14, 2021.
- [7] G. B. Sulisty, "Penerapan Decision Tree C . 45 Berbasis Particle Swarm Optimization untuk Prediksi Hasil Pemilu Legeslatif," *Sentra Penelit. Eng. dan edukasi*, vol. 12, no. 1, pp. 21–26, 2020.
- [8] N. Nanni and A. Sudransyah, "Perbandingan Kinerja Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Harga Pangan," *PROtek J. Ilm. Tek. Elektro*, vol. 7, no. 1, pp. 20–24, 2020, doi: 10.33387/protk.v7i1.1710.
- [9] N. Yahya and A. Jananto, "Komparasi Kinerja Algoritma C.45 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaanmahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang)," *Pros. SENDI*, no. 2014, pp. 978–979, 2019, [Online]. Available: <https://www.unisbank.ac.id/ojs/index.php/sendu/article/view/7389/2369>.
- [10] D. M. B. Tarigan, D. P. Rini, and Samsuryadi, "Seleksi Fitur pada Klasifikasi Penyakit Gula Darah Menggunakan Particle," vol. 4, no. 3, pp. 569–575, 2020.
- [11] E. Ermawati, "Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Penerima Bantuan Pangan Non Tunai," *Sistemasi*, vol. 8, no. 3, p. 513, 2019, doi: 10.32520/stmsi.v8i3.576.
- [12] M. A. Muslim, S. H. Rukmana, E. Sugiharti, B. Prasetyo, and S. Alimah, "Optimization of C4.5 algorithm-based particle swarm optimization for breast cancer diagnosis," *J. Phys. Conf. Ser.*, vol. 983, no. 1, 2018, doi: 10.1088/1742-6596/983/1/012063.
- [13] K. J. Prayoga, A. Nugroho, and N. Wiyatno, "Komparasi feature selection particle swarm optimization (pso) dengan genetic algorithm (ga) terhadap algoritma naïve bayes pada analisis sentimen twitter," *Pros. Semin. Nas. Teknol. dan Sains*, no. September, pp. 1–12, 2019.
- [14] D. M. B. Tarigan, D. Palupi Rini, and Sukemi, "Particle swarm optimization-based on decision tree of c4.5 algorithm for upper respiratory tract infections (URTI) prediction," *J. Phys. Conf. Ser.*, vol. 1196, no. 1, 2019, doi: 10.1088/1742-6596/1196/1/012077.
- [15] M.Tuga and B.Faisal, "CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)," 2020. <https://mmsi.binus.ac.id/2020/09/18/cross-industry-standard-process-for->

data-mining-crisp-dm/.