

Penerapan Decision Tree C.45 Berbasis Particle Swarm Optimization untuk Prediksi Hasil Pemilu Legeslatif

Gunawan Budi Sulistyو
Universitas Bina Sarana Informatika
gunawan.gnw@bsi.ac.id

Abstract - General Election is a manifestation of democratic freedom in determining the people's representatives in the Unitary State of the Republic of Indonesia, which is held every 5 years. These democratic parties include legislative elections, which from time to time have changing rules, in determining the elected council. This time the researchers took data from the South Sulawesi Provincial KPU in 2019. The data was processed using Decision Tree C.45 resulting in an accuracy rate of 92.91% while the accuracy value using the C4.5 classification algorithm based on PSO (Particle Swarm Optimization) method is 93.70% with a difference of 0.79% While for evaluation using the ROC curve that is based on the AUC value, the C4.5 classification algorithm is worth 0.953 while the C4.5 classification algorithm based on PSO (Particle Swarm Optimization) is worth 0.961 with a difference of 0.008.

Keywords: C4.5 Algorithm, Particle Swarm Optimization, Elections, Datamining, RapidMiner

Abstrak – Pemilihan Umum merupakan salah satu perwujudan kebebasan demokrasi dalam menentukan wakil rakyat di Negara Kesatuan Republik Indonesia, yang dilaksanakan setiap 5 tahun 1 kali. Pesta demokrasi ini diantaranya adalah pemilihan umum legeslatif, yang dari masa ke masa memiliki aturan yang berubah-ubah, dalam menentukan dewan yang terpilih. Kali ini peneliti mengambil data dari KPU Provinsi Sulawesi Selatan tahun 2019. Data tersebut diolah menggunakan Decision Tree C.45 menghasilkan tingkat akurasi sebesar sebesar 92,91% sedangkan nilai akurasi dengan menggunakan metode algoritma klasifikasi C4.5 berbasis PSO (*Particle Swarm Optimization*) yaitu sebesar 93,70% dengan selisih 0,79% Sedangkan untuk evaluasi menggunakan *ROC curve* yaitu berdasarkan nilai AUC, algoritma klasifikasi C4.5 bernilai 0,953 sedangkan algoritma klasifikasi C4.5 berbasis PSO (*Particle Swarm Optimization*) bernilai 0,961 dengan selisih 0,008.

Kata kunci: Algoritma C4.5, Particle Swarm Optimization, Pemilu, Datamining, RapidMiner

1. Latar Belakang

Indonesia merupakan negara demokrasi, yang menerapkan pemilihan umum secara langsung setiap 5 tahun sekali untuk memilih pemimpin yang akan duduk di daerah, provinsi maupun di pusat. Pemilihan Umum legeslatif adalah salah satu pelaksanaan kedaulatan rakyat diantaranya diselenggarakannya pemilihan umum (selanjutnya disebut pemilu). Ketentuan mengenai pemilu diatur di dalam Undang- Undang Dasar Negara Republik Indonesia Tahun 1945 Pasal 22E ayat (1) sampai dengan ayat (6). Adapun bunyi pasal tersebut yaitu: (1) Pemilihan umum dilaksanakan secara langsung, umum, bebas, rahasia, jujur, dan adil setiap lima tahun sekali, (2) Pemilihan umum diselenggarakan untuk memilih anggota Dewan Perwakilan Rakyat, Dewan Perwakilan Daerah, Presiden dan Wakil Presiden, dan Dewan Perwakilan Rakyat Daerah, (3) Peserta pemilihan umum untuk memilih anggota Dewan Perwakilan Rakyat dan Dewan Perwakilan Rakyat Daerah adalah partai politik, (4) Peserta pemilihan umum untuk memilih Dewan Perwakilan Daerah adalah

peseorangan, (5) Pemilihan umum diselenggarakan oleh suatu komisi pemilihan umum yang bersifat nasional, tetap dan mandiri, (6) Ketentuan lebih lanjut tentang pemilihan umum diatur dengan undang- undang.

Kebijakan afirmasi (*affirmative action*) terhadap perempuan dalam bidang politik setelah berlakunya perubahan UUD 1945 dimulai dengan disahkannya UU Nomor 12 Tahun 2003 tentang Pemilu DPR, DPD, dan DPRD. Peningkatan keterwakilan perempuan berusaha dilakukan dengan cara memberikan ketentuan agar partai politik peserta Pemilu memperhatikan keterwakilan perempuan sekurang-kurangnya 30% di dalam mengajukan calon anggota DPR, DPD, dan DPRD. Pasal 65 ayat (1) UU Nomor 12 Tahun 2003 tentang Pemilu DPR, DPD, dan DPRD menyatakan: "*Setiap Partai Politik Peserta Pemilu dapat mengajukan calon Anggota DPR, DPRD Provinsi, dan DPRD Kabupaten/Kota untuk setiap Daerah Pemilihan dengan memperhatikan keterwakilan perempuan sekurang-kurangnya 30%.*"

Pada UU No 10/2008 Pasal 214, pada pokoknya, menentukan bahwa anggota legislatif terpilih ditentukan berdasarkan suara terbanyak secara berurutan dari antara para calon anggota legislatif (caleg) yang meraih suara "sekurang-kurangnya 30 persen" dari bilangan pemilih pembagi (BPP) di daerah pemilihan yang bersangkutan.

Jika tidak ada yang mencapai lebih dari 30 persen atau ada lebih dari satu calon yang mendapat lebih dari 30 persen dari BPP, keterpilihan anggota legislatif ditentukan berdasarkan nomor urut dari yang terkecil.

Metode prediksi hasil pemilihan umum legislatif sudah pernah dilakukan menggunakan Algoritma K-Nearest Neighbor, Neural Network berbasis Particle Swarm, Naive Bayes menggunakan algoritma genetika sebagai fitur seleksi.

Pada penelitian ini peneliti menggunakan metode *Decision Tree*. Dalam *data mining*, *Decision Tree* adalah cara yang umum untuk mengatur skema klasifikasi. Klasifikasi menggunakan *Decision Tree* dilakukan dengan routing dari simpul akar hingga simpul daun (Galathya, 2012). *Decision tree* mempunyai kelebihan yaitu mempunyai kelebihan dalam prediksi karena struktur algoritmanya mudah dimengerti dan tingkat kesalahannya cukup kecil sedangkan kelemahan algoritma *decision tree* adalah keandalan cabang yang lebih rendah menjadi lebih buruk dari cabang di atasnya, pohon keputusan yang dihasilkan tidak optimal dan tidak bisa menggunakan sampel yang lebih besar, karena itu tidak mudah untuk memahami pohon keputusan besar dan masalah *overfitting* data bisa terjadi dengan target data terbatas yang ditetapkan (Nurrahman, 2017).

Particle Swarm Optimization (PSO) adalah metode pencarian PSO memiliki struktur sederhana, kemampuan manuver yang kuat, mudah direalisasi, dan karakteristik lainnya. Hingga kini, PSO sudah telah berhasil diterapkan di berbagai daerah (Sun, Song, & Chen, 2019). Metode optimasi sangat mudah diterapkan dan ada beberapa parameter untuk menyesuainya. Berdasarkan latar belakang yang ada, penulis mencoba melakukan penelitian prediksi hasil pemilihan umum legislatif dengan menggunakan metode C.45 kemudian dioptimisasi dengan *Particle Swarm Optimization* untuk meningkatkan nilai akurasi dari prediksi pemilu itu sendiri. Hasil yang diharapkan dari penelitian ini adalah untuk memperoleh hasil prediksi yang lebih akurat, agar dapat digunakan keputusan oleh masyarakat yang mencalonkan menjadi anggota legislatif.

2. Kajian Pustaka

2.1. Pemilihan Umum

Pemilihan Umum (pemilu), yang diselenggarakan oleh Negara Kesatuan Republik Indonesia, yang meliputi pemilihan presiden dan wakil presiden, kepala daerah dan wakilnya (PILKADA), pemilihan DPD, merupakan sarana pelaksanaan kedaulatan rakyat yang diselenggarakan dengan prinsip jujur (jujur adil) serta luber (langsung umum bebas dan rahasia), sesuai dengan ketentuan dan perundang-undangan yang berlaku. Pemisahan sistem pemilu tersebut, dinilai kurang efektif dan efisien dalam pelaksanaan pemilu yang menganut pemerintahan sistem presidensial, karena menimbulkan berbagai permasalahan, seperti konflik yang terus terjadi antara berbagai kepentingan kelompok maupun individu, pemborosan anggaran dalam penyelenggaraannya, maraknya politik uang, politisasi birokrasi, serta tingginya intensitas pemilu di Indonesia. Intensitas penyelenggaraan pemilu, pilpres dan pilkada yang terlampaui sering tersebut berdampak pada rendahnya tingkat partisipasi sebagai akibat kejenuhan publik (Solihah & Padjadjaran, 2019).

Persoalan lain dari format pemilu tersebut adalah fakta bahwa penyelenggaraan pemilu legislatif selalu mendahului pemilu presiden, padahal pada saat yang sama kita sepakat untuk memperkuat sistem presidensial. Pemilu legislatif yang mendahului pemilu presiden dalam skema presidensial jelas sebuah anomali, mengingat di dalam sistem presidensial lembaga eksekutif terpisah dari lembaga legislatif. Di sisi lain, penyimpangan ini beresiko pada implementasi sistem presidensial itu sendiri, baik dalam praktek politik dan pemerintahan. Salah satu resiko itu adalah berlangsungnya pencalonan pilpres yang "didikte" oleh hasil pemilihan legislatif. Artinya, tidak semua parpol bisa mengajukan pasangan calon untuk pemilihan umum presiden. Hanya parpol atau gabungan parpol yang memenuhi syarat ambang batas perolehan suara atau kursi minimal tertentu yang dapat mengajukan pasangan calon presiden dan wakil presiden. Untuk Pilpres 2009 dan 2014 misalnya, hanya parpol atau gabungan parpol yang memperoleh suara sekurang-kurangnya 25% atau perolehan kursi DPR sekurang-kurangnya 20% yang dapat mengajukan pencalonan dalam pilpres.

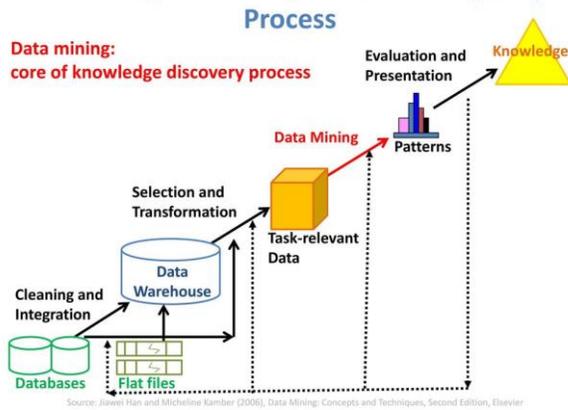
Dalam kepustakaan ilmu politik terdapat prinsip-prinsip pilar sistem politik yang demokrasi, yaitu (1) jaminan hak dan kebebasan warga Negara; (2) partisipatory democracy; (3) sistem memilih dan mengganti penyelenggara Negara; (4) rule of law; (5) check and balance separation of power; (6) pemerintah dan oposisi yang efektif; (7) sistem pemerintahan daerah berdasarkan desentralisasi; (8) Korelasi Pemilu Serentak dengan Multi Partai Sederhana Sebagai Penguatan Sistem

Presidensial paham konstitusionalisme; (9) pemerintahan oleh partai mayoritas; (10); budaya demokrasi (civic culture) sebagai sikap dan perilaku warga Negara (Solihah & Padjadjaran, 2019).

2.2. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Nurrahman, 2017). Data mining adalah proses untuk menemukan dan membaca pola berdasarkan data; karenanya, ia memiliki fungsi sebagai alat untuk membantu menemukan informasi penting dari data (Agustina dkk, 2018).

Knowledge Discovery in Databases (KDD)



Gambar 1. Knowledge Discovery in Databases

Tahapan proses Knowledge Discovery terdiri dari :

1. Seleksi Data
 Nilai Gain tertinggi menjadi root utama. Untuk mengetahui nilai Gain, kita harus tahu nilai Entropy-nya terlebih dahulu. Rumus untuk menghitung entropi disajikan sebagai berikut:

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

Information:

- S : the set of cases
- A : attribute
- n : the number of partitions S
- pi : the proportion of Si towards S

2. Kalkulasi dari nilai Gain

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

Information:

- S : The set of cases

- A : attribute
- n : The number of attribute partition A
- |Si| : the number of cases on the i-th partition
- |S| : the number of cases in S

3. Ulangi Langkah 2 sampai semua tupel dipartisi.
4. Proses partisi akan berhenti jika mengalami kondisi berikut.
 - a. Sebuah. Semua tupel dalam simpul mendapatkan kelas yang sama.
 - b. Semua atribut sudah dipartisi.
 - c. Tidak ada tupel di cabang kosong.

2.3. Swarm Intelligence

Disebut Swarm Intelligence, karena Swarm Optimization Algorithm adalah metode optimasi berdasarkan agen terdistribusi yang berfungsi untuk memecahkan masalah. Swarm Intelligence mempelajari perilaku seperangkat organisme social(Purnomo, 2014).

2.4. Particle Swarm Optimization

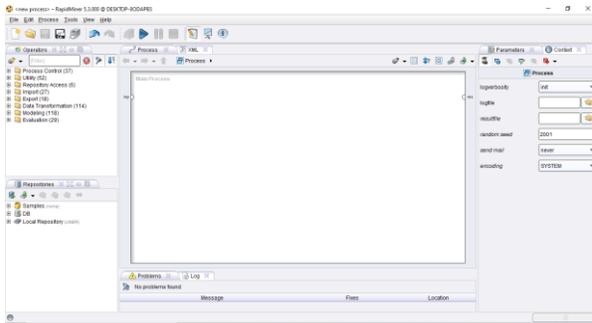
Particle Swarm Optimization adalah model sederhana dari teori evolusi yang didasarkan pada prinsip-prinsip yang berasal dari perilaku sekelompok lebah dalam mencari sumber makanan. Ketika mereka pertama kali mencari makanan, mereka tidak tahu di mana lokasi sumber makanan itu, tetapi hewan akhirnya bisa mencapai lokasi sumber makanan terbaik dengan berkomunikasi satu sama lain (Dutta, Roy, & Choudhury, 2013).

Ada 3 langkah yang harus dilakukan Partikel Swarm Optimization yaitu:

1. Mengevaluasi nilai kebugaran dari partikel yang ada.
2. Memperbarui nilai kebugaran dan posisi individu terbaik yang biasa disebut global terbaik.
3. Memperbarui kecepatan dan posisi masing-masing partikel. (Purnomo, 2014)

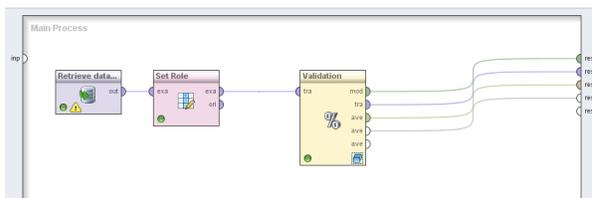
2.5. Software Rapid Miner

Rapid Miner banyak digunakan untuk memproses data. Perangkat lunak ini dapat diunduh di www.rapidminer.com. Layoutnya terlihat seperti ini:



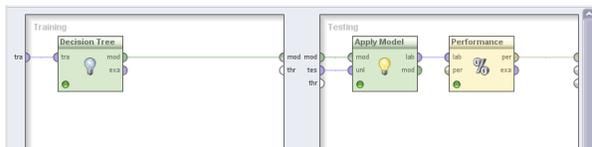
Gambar 2. Software RapidMiner

Model yang dibangun pada penelitian kali ini adalah sebagai berikut:



Gambar 3. Model yang diusulkan

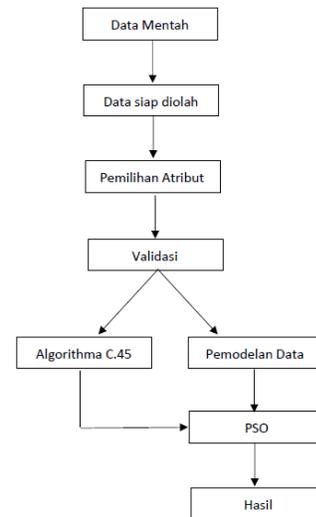
Data diproses untuk ditentukan role nya (aturan), kemudian divalidasi menggunakan *Decision Tree*, yang selanjutnya di optimasi menggunakan *Particle Swarm*, seperti gambar dibawah ini:



Gambar 4. *Decision Tree* PSO

2.6. Kerangka pemikiran

Dalam melaksanakan penelitian, penulis membuat sebuah kerangka pemikiran yang digunakan sebagai pedoman atau acuan penelitian ini, agar dapat menghasilkan yang terbaik. Permasalahan pada penelitian ini adalah untuk menerapkan metode *Decision Tree C.45* berbasis *Particle Swarm Optimization* untuk mengetahui prediksi hasil pemilihan umum legislative tingkat provinsi. Pada penelitian ini, data yang diolah mengambil dari hasil PEMILU Legislatif DPRD Provinsi Sulawesi Selatan Tahun 2019, yang datanya diunduh melalui <https://sulsel.kpu.go.id/pileg-2019/>. Data tersebut masih berupa data mentah, yang harus diolah terlebih dahulu agar dapat diproses.



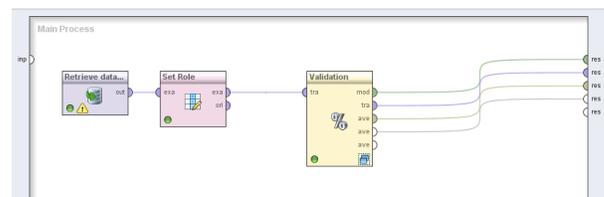
Gambar 5. Kerangka penelitian

Data mentah berupa data dalam format rekapitulasi hasil penghitungan suara tingkat Provinsi dari 9 Kabupaten dan Kota. Data tersebut diolah disesuaikan format nya dengan mengambil dan menyusun beberapa atribut yang diperlukan, maka data sudah menjadi siap diolah. Selanjutnya pemilihan atribut yang menjadi keputusan Terpilih atau Tidak Terpilih yang divalidasi dengan *Algorithm C.45* untuk menjadi data Training, serta pemodelan data yang dioptimasi dengan *Particle Swarm Optimization (PSO)* menjadi data Testing yang elanjutnya diproses untuk menghasilkan keputusan.

3. Implementasi Sistem dan Hasil

3.1. Implementasi *Algorithm C.45*

Adapun desain model menggunakan Rapidminer adalah sebagai berikut :



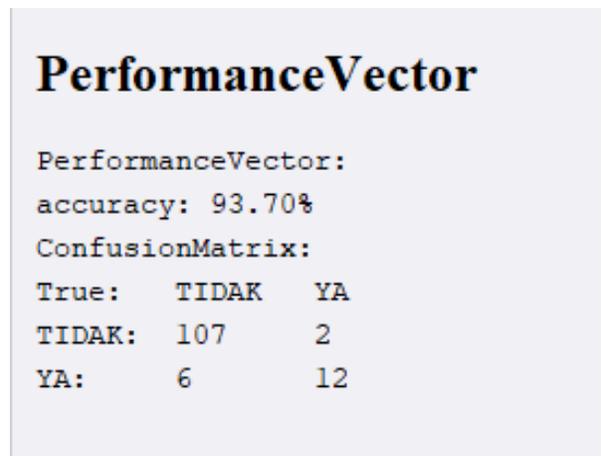
Gambar 6. Desain model C.45

Gambar 6 menunjukkan proses data siap diolah yang di tentukan terlebih dahulu aturannya, selanjutnya diklasifikasi dan diproses, sehingga hasilnya disajikan dalam bentuk table sebagai berikut :

Table 1. Accuracy C.45

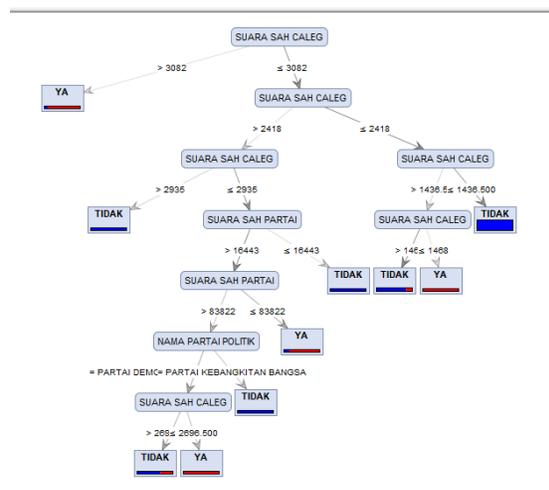
	true TIDAK	true YA	class precision
pred. TIDAK	107	2	98.17%
pred. YA	6	12	66.67%
class recall	94.69%	85.71%	

Tabel ini menunjukkan dari 109 yang diprediksi TIDAK, ternyata hanya ada 2 yang dinyatakan YA, atau telah mencapai 98.17%, sedangkan dari 18 data yang diperkirakan YA ternyata ada 12 yang sesuai atau sekitar 66.67% .



Gambar 7. PerformanceVector

Pada performacneVector menunjukkan accuracy 93.70%. Sedangkan grafik ditunjukkan pada gambar berikut



Gambar 8. Graph View Decision Tree

Dari gambar 8 menunjukkan pohon keputusan terpilih (YA) atau TIDAK nya seorang calon legeslatif yang turut berkompetensi.

```

Tree
SUARA SAH CALEG > 3082: YA (TIDAK=3, YA=23)
SUARA SAH CALEG ≤ 3082
| SUARA SAH CALEG > 2418
| | SUARA SAH CALEG > 2935: TIDAK (TIDAK=5, YA=0)
| | SUARA SAH CALEG ≤ 2935
| | | SUARA SAH PARTAI > 16443
| | | | SUARA SAH PARTAI > 83822
| | | | | NAMA PARTAI POLITIK = PARTAI DEMOKRASI INDONESIA PERJUANGAN
| | | | | SUARA SAH CALEG > 2696.500: TIDAK (TIDAK=2, YA=1)
| | | | | SUARA SAH CALEG ≤ 2696.500: YA (TIDAK=0, YA=3)
| | | | | NAMA PARTAI POLITIK = PARTAI KEBANGHITAN BANGSA: TIDAK (TIDAK=3, YA=0)
| | | | | SUARA SAH PARTAI ≤ 83822: YA (TIDAK=2, YA=10)
| | | | | SUARA SAH PARTAI ≤ 16443: TIDAK (TIDAK=2, YA=0)
| | | SUARA SAH CALEG ≤ 2418
| | | SUARA SAH CALEG > 1436.500
| | | SUARA SAH CALEG > 1468: TIDAK (TIDAK=47, YA=9)
| | | SUARA SAH CALEG ≤ 1468: YA (TIDAK=0, YA=3)
| | SUARA SAH CALEG ≤ 1436.500: TIDAK (TIDAK=313, YA=0)
    
```

Gambar 9. TextView Decision Tree

Pada gambar 8 menghasilkan bahwa suara sah caleg yang lebih besar dari 3082 suara diprediksi YA ternyata benar 23, sedangkan 3 yang lainnya TIDAK.

4. Penutup

Hasil penelitian dari nilai akurasi dengan menggunakan metode algoritma klasifikasi C4.5 yaitu sebesar 92,91% sedangkan nilai akurasi dengan menggunakan metode algoritma klasifikasi C4.5 berbasis PSO (*Particle Swarm Optimization*) yaitu sebesar 93,70% dengan selisih 0,79%. Sedangkan untuk evaluasi menggunakan *ROC curve* yaitu berdasarkan nilai AUC, algoritma klasifikasi C4.5 bernilai 0,953 sedangkan algoritma klasifikasi C4.5 berbasis PSO (*Particle Swarm Optimization*) bernilai 0,961 dengan selisih 0,008. Sehingga dapat disimpulkan bahwa penerapan optimasi Particle Swarm Optimization dapat meningkatkan nilai akurasi pada algoritma klasifikasi C4.5.

5. Saran

Berdasarkan proses pengujian dan kesimpulan yang telah dilakukan, dibawah ini merupakan beberapa saran dari penelitian ini, antara lain :

1. Mengembangkan dengan optimasi yang lain guna membandingkan optimasi mana yang lebih akurat tingkat akurasi seperti *Ant Colony Optimization (ANT)*, *Genetik Algorithm(GA)*, dan lain sebagainya.
2. Menggunakan data yang lebih besar lagi dan dengan atribut yang lebih banyak sehingga hasil yang akan didapatkan akan lebih baik.

6. Pustaka

- Agustina. (2018). Decision Tree Optimization Algorithm Using Particle Swarm Optimization To Improve The Accuracy Of Detection Malnutrition In Toddler.
- D. Dutta, R. Roy and K. Choudhury, Training .International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no 3, pp. 430- 434, 2013.
- Galathiya, A. S. (2012). Classification with an improved Decision Tree Algorithm, *46(23)*, 1–6.
- H.D.Purnomo (2014), Soccer Game Optimization: Fundamental Concept, *Jurnal Sistem Komputer*, Vol 4, No 1, pp. 30-41
- Nurrahman, D. (2017). Algoritma klasifikasi c4.5 berbasis particle swarm optimization untuk prediksi hasil pemilihan legislatif dprd karawang, *II(April)*, 28–37.
- Solihah, R., & Padjadjaran, U. (2019). Peluang dan tantangan pemilu serentak 2019 dalam perspektif politik, *3(1)*, 73–88. <https://doi.org/10.14710/jiip.v3i1.3234>
- Sun, L., Song, X., & Chen, T. (2019). An Improved Convergence Particle Swarm Optimization Algorithm with Random Sampling of Control Parameters, *2019*.