

IJCIT

(Indonesian Journal on Computer and Information Technology)

Journal Homepage: <http://ejournal.bsi.ac.id/ejurnal/index.php/ijcit>

Optimasi Algoritma C4.5 Untuk Mengukur Keputusan Pembelajaran Daring Berbasis Particle Swarm Optimization (PSO)

Dewi Ayu Nurwulandari¹, Siti Masripah², Rizal Amegia Saputra³

¹Sistem Informasi, Universitas Bina Sarana Informatika
Bogor, Indonesia
e-mail: dewi.dan@bsi.ac.id

²Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika
Bogor, Indonesia
e-mail: siti.stm@bsi.ac.id

³Sistem Informasi Akuntansi Kampus Kota Sukabumi, Universitas Bina Sarana Informatika
Sukabumi, Indonesia
e-mail: rizal.rga@bsi.ac.id

ABSTRAK

Algoritma yang populer dan modern dalam pengolahan data dengan teknik data mining adalah Algoritma C4.5. Algoritma C4.5 banyak digunakan untuk melakukan pengklasifikasian data karena algoritma C4.5 dapat menghasilkan sebuah pohon keputusan yang mudah dipahami dan mudah dimengerti. Pada penelitian ini metode yang digunakan adalah dengan menambahkan teknik optimasi menggunakan algoritma Particle Swarm Optimization (PSO) dengan tujuan meningkatkan nilai akurasi pada information gain algoritma C4.5 untuk mengukur keputusan pembelajaran daring. PSO merupakan salah satu metode dan teknik untuk mengklasifikasi dan meningkatkan akurasi, dimana PSO terdiri dari sekumpulan partikel yang mencari posisi yang terbaik. Hasil penelitian menunjukkan akurasi dan Kappa pada nilai information gain yaitu sebesar 91,76 dan 0,834, akurasi dan Kappa Gain Ratio sebesar 89,41 dan 0,788, akurasi dan Kappa Gini Index sebesar 90,59 dan 0,811. Sehingga diperoleh kesimpulan penerapan algoritma PSO dapat berpengaruh terhadap nilai akurasi pada setiap criteria splitting algoritma C4.5

Kata Kunci: algoritma c4.5, data mining, particle swarm optimization, pso

ABSTRACTS

A popular and modern algorithm in data processing with data mining techniques is the C4.5 Algorithm. The C4.5 algorithm is widely used to classify data because the C4.5 algorithm can produce a decision tree and easy to understand. In this study, the author made a comparison between previous studies using the C4.5 algorithm by adding optimization techniques using the Particle Swarm Optimization (PSO) algorithm with the aim of increasing the accuracy value of the C4.5 algorithm information gain in measuring Online Learning Decisions. PSO Technique is one of the methods and techniques for classifying and improving accuracy, where PSO consists of a set of particles that are looking for the best position. The results of this study showed the results of accuracy and Kappa on the value of information gain, namely 91.76 and 0.834, accuracy and Kappa Gain Ratio of 89.41 and 0.788, accuracy and Kappa Gini Index of 90.59 and 0.811. So that it can be concluded that the application of the PSO algorithm can affect the accuracy value in each criteria splitting the C4.5 algorithm



Jurnal ini dapat diakses secara terbuka dan memiliki lisensi CC-BY-SA

(<https://creativecommons.org/licenses/by-sa/4.0/>) ©2022 by penulis dan IJCIT

Keywords: *c4.5 algorithm, data mining, particle swarm optimization, pso*

1. PENDAHULUAN

Pandemi COVID-19 telah banyak memberikan dampak terhadap berbagai aktifitas (Martin & Eliza, 2020), berbagai cara dan peraturan dikeluarkan oleh pemerintah untuk membatasi penyebaran virus COVID-19 (Irhandayaningsih, 2020). Dunia Pendidikan merupakan salah satu yang terkena dampak, peraturan untuk menjaga jarak (*physical distancing*) yang mengakibatkan pembelajaran menjadi *daring* (Ikhsan, 2021). Berbagai jenis perangkat lunak banyak digunakan sebagai alat untuk mendukung pembelajaran *daring* (Natuzzuhriyyah & Nafisah, 2021), namun banyak guru maupun siswa yang mengeluhkan pembelajaran secara *daring*, karena merasa terjadi penurunan kualitas pembelajaran (Mamluah & Maulidi, 2021), untuk itu perlu pengukuran terhadap keputusan pembelajaran *daring* dengan algoritma data mining.

Algoritma C4.5 merupakan metode datamining yang banyak digunakan untuk melakukan pengklasifikasian data, dikarenakan algoritma C4.5 akan didapatkan sebuah pohon keputusan yang mudah dipahami dan mudah dimengerti (Mardi, 2017). Salah satu algoritma yang populer dan modern untuk penggunaan di data mining adalah Algoritma C4.5 bisa disebut sebagai pohon keputusan (decision tree), yang menggunakan struktur pohon yang memiliki node dan setiap node digunakan dalam merepresentasikan atribut, kemudian cabang yang dimiliki mempresentasikan nilai dari atribut, sedangkan untuk daunnya mempresentasikan kelas. Sehingga konsep decision tree ini adalah dengan mengumpulkan data berikutnya akan terbentuk rule-rule solusi permasalahan (Elisa, 2017). Algoritma C4.5 merupakan algoritma terbaik untuk mendapatkan hasil akurasi yang lebih tinggi, dalam penelitian Masripah dikatakan hasil akurasi algoritma C4.5 sebesar 88,90% (Siti, 2016).

Hasil Sebuah klasifikasi menggunakan metode Information Gain banyak digunakan dalam penelitian, salah satunya adalah penelitian Rizal dalam penelitiannya ia menerapkan algoritma C4.5 dengan split kriteria information gain sehingga mendapatkan hasil klasifikasi yang sangat baik dengan nilai AUC sebesar 0,986 (Saputra, Wasiyanti, & Pribadi,

2021). Dalam penelitian yang dilakukan Hoiriyah Algoritma C4.5 berbasis seleksi atribut memiliki nilai akurasi sebesar 96.81%. dengan menggunakan atribut chi squared, Information Gain, Relief, dengan Criterion : Information Gain yang memiliki nilai akurasi yang tertinggi. Sehingga dapat dikatakan klasifikasi algoritma C4.5 menggunakan kriteria atribut Information gain adalah yang tertinggi untuk nilai akurasinya (Hoiriyah, 2018).

Ada beberapa penelitian yang telah menggunakan teknik Optimasi guna meningkatkan nilai akurasi sebuah penelitian, diantaranya adalah penelitian yang dilakukan oleh Lala dan Yuni dengan melakukan optimasi pada algoritma Support Vector Machine (SVM) menggunakan optimasi PSO, dengan hasil PSO merupakan metode yang cukup baik untuk mengklasifikasi data (Nilawati & Achyani, 2019). Particle Swarm Optimization (PSO) merupakan salah satu teknik optimasi untuk meningkatkan akurasi, PSO terdiri dari sekumpulan partikel yang mencari posisi yang terbaik (Junianto & Riana, 2017).

Pada penelitian sebelumnya yang membahas tentang Pencarian Criteria Splitting Terbaik pada algoritma C4.5 untuk mengukur pemilihan pembelajaran pada Era Pandemi Covid-19 terlihat hasil untuk Criteria yang paling tinggi hasil akurasinya adalah Gain Ratio dan Information Ratio sebesar 85.88% dengan nilai AUC yang paling tinggi yaitu 0,80 pada criteria Gain ratio dengan hasil klasifikasi Baik (Masripah, Nurwulandari, & Saputra, 2022).

Pada penelitian ini Penulis melakukan perbandingan antara penelitian sebelumnya yang menggunakan algoritma C4.5 dengan menambahkan teknik optimasi pada algoritma C4.5. Optimasi yang akan dilakukan adalah menggunakan algoritma PSO diharapkan dengan menggunakan optimasi PSO dapat meningkatkan nilai akurasi pada information gain algoritma C4.5 dengan tujuan untuk meningkatkan nilai akurasi pada information gain algoritma C4.5 dalam mengukur Keputusan Pembelajaran Daring.

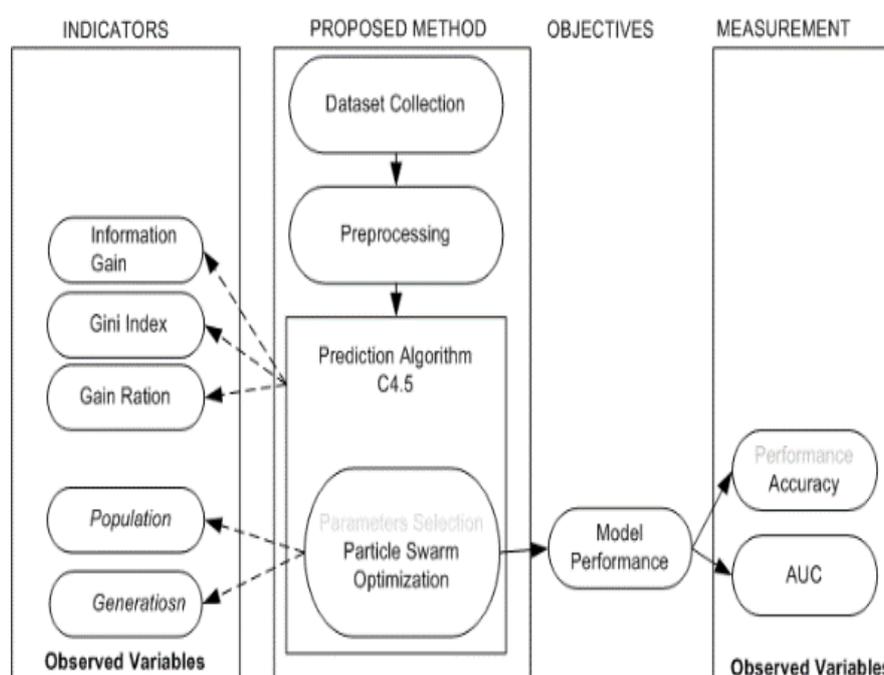
2. METODE PENELITIAN

Metode penelitian yang digunakan untuk menyelesaikan masalah penelitian adalah

dengan menggunakan Teknik Data Mining dengan menggunakan Algoritma C4.5 dengan teknik Particle Swarm Optimization (PSO). Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (Iterative Dichotomiser 3) yang dikembangkan oleh J. Ross Quinlan yang merupakan salah satu algoritma induksi pohon keputusan.

Penelitian ini dilakukan dalam beberapa tahapan, seperti terlihat pada kerangka pemikiran Gambar 1. Permasalahan pada penelitian ini adalah belum diketahui algoritma yang akurat untuk pengukuran

keputusan pembelajaran daring. Untuk itu dibuat approach (model) yaitu algoritma C4.5 dengan teknik optimasi PSO, untuk memecahkan permasalahan kemudian dilakukan pengujian dari masing-masing kriteria splitting algoritma C4.5 seperti penelitian Masripah (Masripah et al., 2022) dan pengujian diteruskan menggunakan teknik PSO untuk meningkatkan performa dari akurasi yang sudah ada. Pada tahap akhir akan dibandingkan hasil akurasi sebelum dioptimasi dan sesudah dioptimasi.

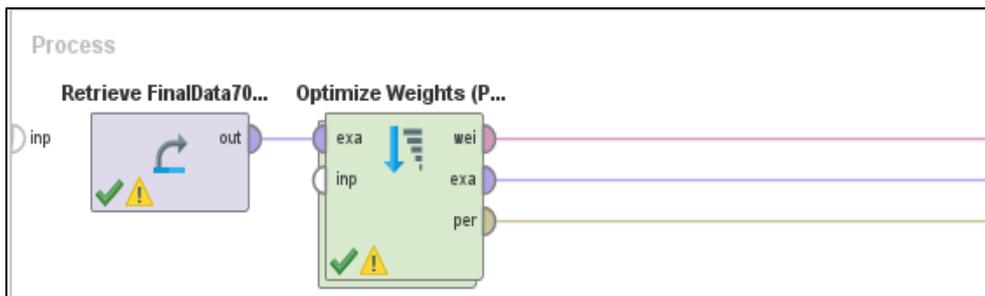


Gambar 1. Kerangka Pemikiran

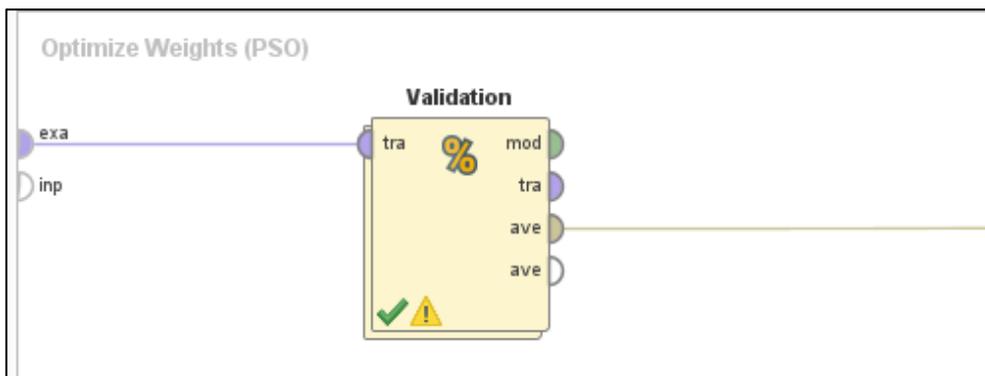
3. HASIL DAN PEMBAHASAN

Untuk pengolahan data, penulis menggunakan perbandingan nilai data 70 dan 30 seperti penelitian sebelumnya, pada penelitian sebelumnya yaitu membahas Pencarian Criteria Splinting Terbaik pada algoritma C4.5 didapat nilai Kriteria yang paling tinggi hasil akurasinya adalah Gain Ratio dan Information Ratio sebesar 85.88% dengan nilai AUC yang paling tinggi yaitu 0,80 pada criteria Gain ratio dengan hasil klasifikasi Baik (Masripah et al., 2022). Pada penelitian berikutnya peneliti menambahkan optimasi PSO pada pemilihan

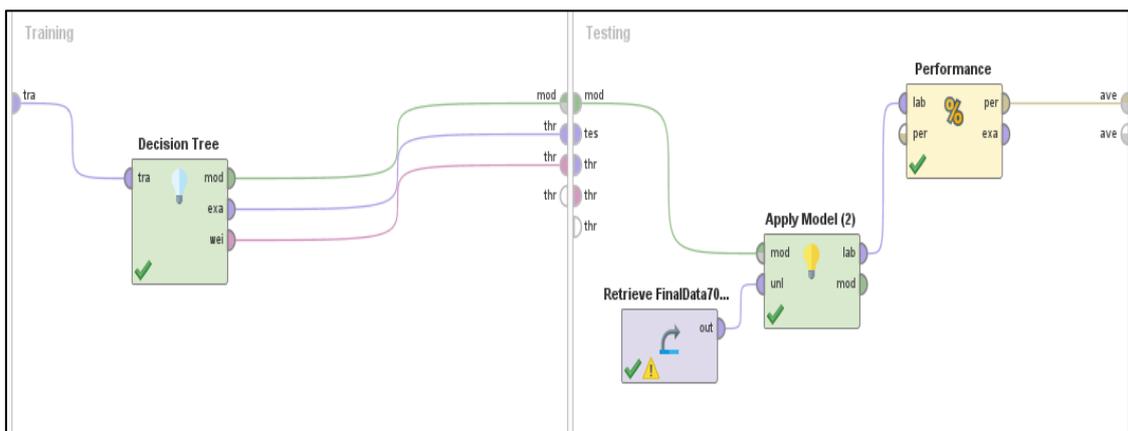
kriteria dengan hasil akurasi dan kappa. Berikut proses Algoritma C4.5 dengan PSO, pada gambar 2 proses pengolahan data yang pertama dilakukan adalah melakukan proses Optimasi Weights (PSO). Proses kedua menggunakan validasi yang nantinya akan dihubungkan dengan Decesion Tree yang terlihat pada gambar 3. Proses terakhir adalah melakukan testing pada data dengan melihat performance dari algoritma yang digunakan, hal ini terlihat pada gambar 4.



Gambar 2. Proses Algoritma C4.5 dengan PSO



Gambar 3. Proses Validasi



Gambar 4. Proses Performance Algoritma

3.1. Akurasi Algoritma C4.5 Berbasis PSO

Pada penelitian ini metode eksperimen dilakukan dengan pengujian model dengan 80:20, didapat nilai akurasi algoritma C4.5 dengan berbasis PSO dan di dapat juga nilai kappa sebagai pengukuran range klasifikasi (Altman DG, 1997), berikut adalah hasil masing-masing dari criteria splitting. Tabel 1 adalah tabel view untuk criteria splitting Information Ratio, sedangkan pada tabel 2 merupakan tabel view dari criteria splitting Gain Ratio dan pada tabel 3 adalah tabel view untuk akurasi criteria splitting Gini Index.

Tabel 1. Tabel View Criteria Information Ration

Accuracy : 91.76%

	True Tidak Setuju	True Setuju	Class precision
Pred. Tidak Setuju	36	3	92.31%
Pred. Setuju	4	42	91.30%
Class recall	90.00%	93.33%	

Tabel 2. Tabel View Criteria Splitting Gain Ratio

Accuracy : 89.41%			
	True Tidak Setuju	True Setuju	Class precision
Pred. Tidak Setuju	36	5	87.80%
Pred. Setuju	4	40	90.91%
Class recall	90.00%	88.89%	

Tabel 3. Tabel View Criteria Splitting Gini Index

Accuracy : 90.59%			
	True Tidak Setuju	True Setuju	Class precision
Pred. Tidak Setuju	35	3	92.11%
Pred. Setuju	5	42	89.36%
Class recall	87.50%	93.33%	

Dari hasil pada tabel 4 dapat disimpulkan bahwa criteria splitting information

gain berbasis PSO memiliki akurasi yang paling tinggi yaitu sebesar 91,76% dan nilai kappa sebesar 0,834.

Tabel 4. Komparasi Nilai Akurasi

	Information Gain + PSO	Gain Ratio + PSO	Gini Index + PSO
Akurasi	91,76%	89,41 %	90,59%
Kappa	0,834	0,788	0,811

3.2. Hasil Atribut Optimasi

Pada penelitian ini juga algoritma PSO digunakan untuk mengukur atribut yang memiliki value atau tidak. Tabel 5 menunjukkan atribut optimasi pada panelitian ini.

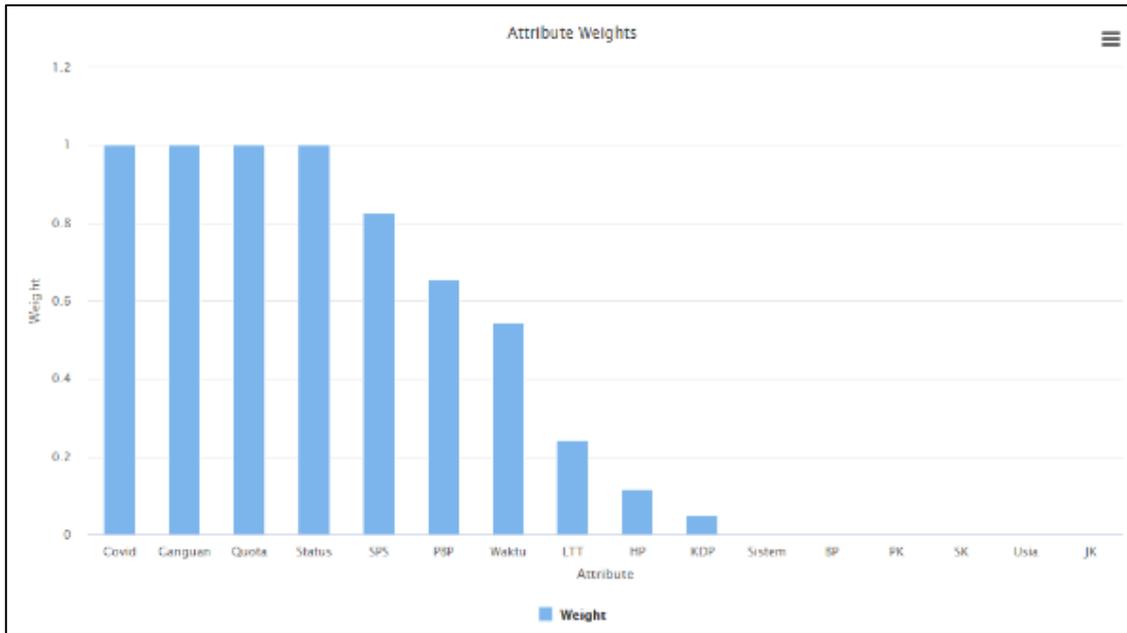
Tabel 5. Atribut Optimasi

No	Information Gain		Gain Ratio		Gini Index	
	Attribute	Weight	Attribute	Weight	Attribute	Weight
1	JK	0.0	JK	0.14232071251507028	JK	0.0
2	Usia	0.0	Usia	0.07736300953569464	Usia	0.3184263069593011
3	Status	1.0	Status	0.2652043324489136	Status	0.0
4	SK	0.0	SK	0.2802396559329807	SK	1.0
5	PK	0.0	PK	0.765188825590622	PK	1.0
6	LTT	0.24435877250718088	LTT	0.3991493078098214	LTT	0.9785461541018604
7	HP	0.11562967187870599	HP	0.5455566409578625	HP	0.0
8	BP	0.0	BP	0.5028577310431629	BP	0.0
9	PBP	0.6553125969280355	PBP	0.7688632659462901	PBP	1.0
10	Quota	1.0	Quota	0.900851348782737	Quota	1.0
11	Gangguan	1.0	Gangguan	0.5163865180448856	Gangguan	1.0
12	Covid	1.0	Covid	0.8083191918055612	Covid	1.0
13	Sistem	0.0	Sistem	0.36960084019513373	Sistem	0.4799220128523563
14	Waktu	0.5430252417113376	Waktu	0.42629839434304473	Waktu	1.0
15	SPS	0.8264765092938252	SPS	0.37073761504108216	SPS	1.0
16	KDP	0.050470178998571286	KDP	0.004687313260214765	KDP	0.00488937063567172

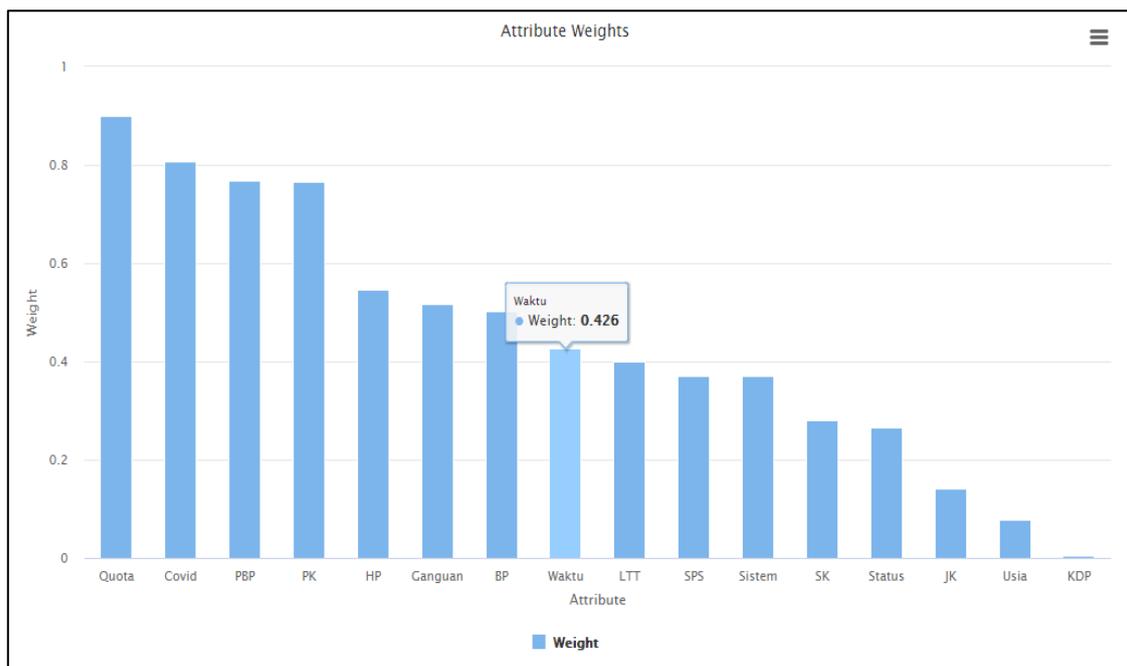
Dari tabel 5, dapat dilihat hasil optimasi atribut, terdapat beberapa atribut yang memiliki nilai 0.0 artinya atribut tersebut tidak berpengaruh terhadap proses klasifikasi.

Gambar 5, menampilkan hasil grafik criteria Splitting Information Gain, grafik attribute

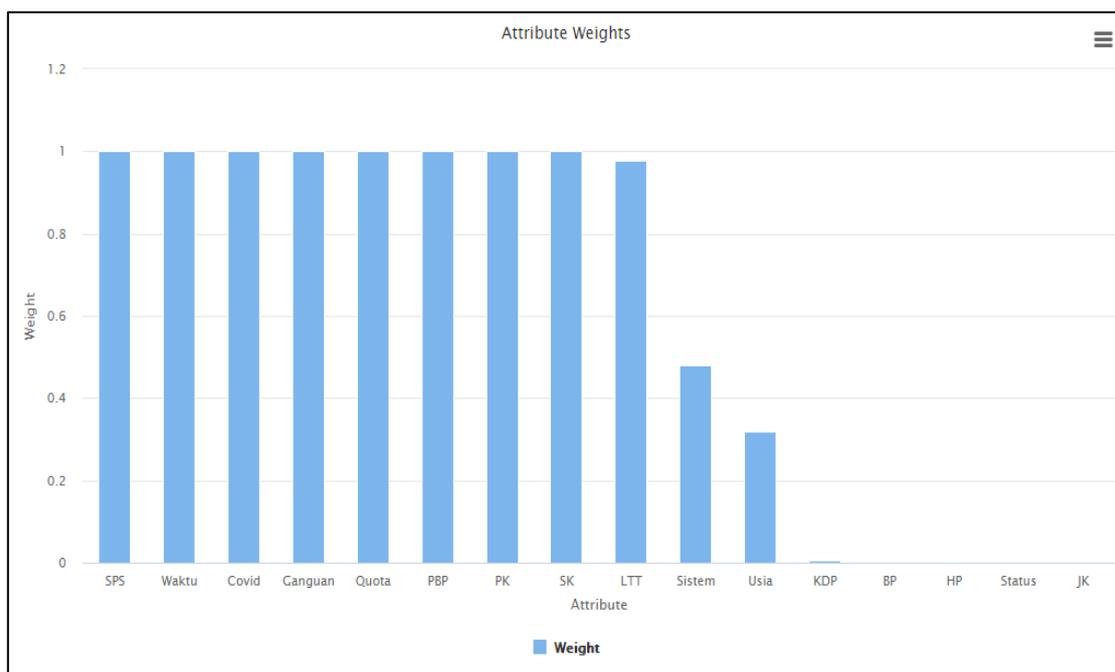
weights untuk X-Axis colom berdasarkan atribut dan value kolom adalah weight. Sedangkan pada gambar 6 merupakan hasil grafik dari Criteria Splitting Gain Ratio, kolom dan value juga berdasarkan Weight Atribut. Gambar 7 adalah gambar untuk grafik Criteria Splitting Gini Index.



Gambar 5. Grafik Criteria Splitting Information Gain



Gambar 6. Grafik Criteria Splitting Gain Ratio



Gambar 7. Grafik Criteria Splitting Gini Index

4. KESIMPULAN

Setelah dilakukan optimasi menggunakan PSO, maka didapatkan hasil akurasi dan Kappa pada nilai information gain yaitu sebesar 91,76 dan 0,834, akurasi dan Kappa Gain Ratio sebesar 89,41 dan 0,788, akurasi dan Kappa Gini Index sebesar 90,59 dan 0,811. Sehingga dapat disimpulkan bawah penerapan algoritma PSO dapat berpengaruh terhadap nilai akurasi pada setiap criteria splitting algoritma C4.5.

5. REFERENSI

- Altman DG. (1997). *PRACTICAL STATISTICS FOR MEDICAL The Analysis of Time Series*. London.
- Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2(1), 36. <https://doi.org/10.15575/join.v2i1.71>
- Hoiriyah, H. (2018). Algoritma C4.5 Berbasis Seleksi Atribut Untuk Menentukan Kemungkinan Pengunduran Diri Mahasiswa. *Technologia: Jurnal Ilmiah*, 9(1), 67. <https://doi.org/10.31602/tji.v9i1.1104>
- Ikhsan, E. (2021). Penerapan K-Means Clustering dari Log Data Moodle untuk Menentukan Perilaku Peserta pada Pembelajaran Daring. *Sistmasi*, 10(2), 414. <https://doi.org/10.32520/stmsi.v10i2.1285>
- Irhandayaningsih, A. (2020). Pengukuran Literasi Digital Pada Peserta Pembelajaran Daring di Masa Pandemi COVID-19. *Anuva*, 4(2), 231–240. <https://doi.org/10.14710/anuva.4.2.231-240>
- Junianto, E., & Riana, D. (2017). Penerapan PSO Untuk Seleksi Fitur Pada Klasifikasi Dokumen Berita Menggunakan NBC. *Jurnal Informatika*, 4(1), 38–45.
- Mamluah, S. K., & Maulidi, A. (2021). Pembelajaran Jarak Jauh (PJJ) di Masa Pandemi COVID-19 di Sekolah Dasar. *Jurnal Basicedu*, 5(2), 869–877. <https://doi.org/10.31004/basicedu.v5i2.800>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Martin, K., & Eliza, F. (2020). Pengembangan Assessment dalam Pembelajaran Daring untuk Mata Kuliah Pengukuran dan Instrumen. *Jurnal Pendidikan Teknik Elektro*, 1(1), 114–117. <https://doi.org/10.24036/jpte.v1i1.50>

- Masripah, S., Nurwulandari, D. A., & Saputra, R. A. (2022). Pencarian Criteria Splitting Terbaik Pada Algoritma C4.5 Untuk Mengukur Pemilihan Pembelajaran Pada Era Pandemi Covid-19. *Jurnal Larik*, 2(1), 1–7.
- Natuzzuhriyyah, A., & Nafisah, N. (2021). Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naïve Bayes. *Techno Xplore : Jurnal Ilmu Komputer Dan Teknologi Informasi*, 6(2), 61–67. <https://doi.org/10.36805/technoxplore.v6i2.1377>
- Nilawati, L., & Achyani, Y. E. (2019). Optimasi Metode Particle Swarm Optimization (PSO) Pada Prediksi Penilaian Apartemen. *Paradigma - Jurnal Komputer Dan Informatika*, 21(2), 227–234. <https://doi.org/10.31294/p.v21i2.6159>
- Saputra, R. A., Wasiyanti, S., & Pribadi, D. (2021). Information Gain Pada Algoritma C4.5 Untuk Klasifikasi Penerimaan Bantuan Pangan Non Tunai (Bpnt). *Indonesian Journal of Business Intelligence (IJUBI)*, 4(1), 25. <https://doi.org/10.21927/ijubi.v4i1.1757>
- Siti, M. (2016). Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit. *Bina Insani Ict Journal*, 3(1), 187–193. Retrieved from <http://ejournal-binainsani.ac.id/index.php/BIICT/article/view/815/658>