

SISTEM PENDUKUNG KEPUTUSAN DENGAN C4.5 DAN PARTICLE SWARM OPTIMIZATION UNTUK KLASIFIKASI SELEKSI MASUK CALON TARUNA PADA PENERIMAAN TARUNA BARU

Iwan Mahendro

Universitas Dian Nuswantoro

A. Zainul Fanani

Universitas Dian Nuswantoro

Farikh Al Zami

Universitas Dian Nuswantoro

Korespondensi penulis: imahendro@gmail.com

Abstract. *Currently, universities in Indonesia have reached more than 2000 universities. Universities can still exist, one of the factors is the presence of students who enter the college. In accepting new students, each university is certainly not arbitrary, they will select prospective students in order to get students with the quality they expect. There are several ways to select prospective students, some of which are based on report cards, some are through a written test, and there are independent pathways. Not all universities do the processing of incoming data optimally. Usually they still process data manually or in other words, they don't use machine learning. The problem that exists is that there are universities that have not been optimal in processing data on prospective student admissions. Because the data processed has not been optimally processed, there are universities that have not been able to make predictions in the acceptance of prospective students for the next academic year. One way to make predictions in the future is to use data mining. The method that will be used in data mining is Decision Tree classification optimized with Particle Swarm Optimization. The results of the Decision Tree with Particle Swarm Optimization obtained an accuracy of 99.94%.*

Keywords: *decision tree, classification, machine learning, particle swarm optimization, prediction*

Abstrak. Perguruan tinggi di Indonesia saat ini sudah mencapai 2000 lebih perguruan tinggi. Perguruan tinggi bisa tetap eksis salah satu faktornya adalah dengan adanya mahasiswa yang masuk ke perguruan tinggi tersebut. Dalam penerimaan mahasiswa baru setiap perguruan tinggi tentunya tidak sembarangan, mereka akan melakukan seleksi terhadap calon mahasiswa agar mendapatkan mahasiswa dengan kualitas yang mereka harapkan. Ada beberapa cara dalam melakukan seleksi calon mahasiswa, diantaranya ada yang berdasarkan nilai rapor, ada yang melalui tes tertulis, dan ada jalur mandiri. Tidak semua perguruan tinggi melakukan pengolahan data yang masuk secara maksimal. Biasanya mereka masih mengolah data secara manual atau dengan kata lain belum menggunakan machine learning. Masalah yang ada yaitu adanya perguruan

Received Maret, 2022; Revised Maret 2, 2022; Accepted Juni, 2022

* Iwan Mahendro, imahendro@gmail.com

tinggi yang belum secara maksimal dalam mengolah data penerimaan calon mahasiswa. Karena belum optimalnya data yang diolah maka ada perguruan tinggi yang belum dapat membuat prediksi dalam penerimaan calon mahasiswa untuk di tahun ajaran berikutnya. Salah satu untuk dapat membuat prediksi di masa mendatang adalah dengan menggunakan data mining. Metode yang akan digunakan dalam data mining adalah klasifikasi *Decision Tree* dioptimasi dengan *Particle Swarm Optimization*. Hasil dari *Decision Tree* dengan optimasi *Particle Swarm Optimization* didapatkan akurasi sebesar 99,94%.

Kata kunci: decision tree, klasifikasi, machine learning, particle swarm optimization, prediksi

LATAR BELAKANG

Penerimaan calon mahasiswa baru pada suatu perguruan tinggi merupakan suatu kegiatan yang sudah biasa dilakukan oleh tiap – tiap perguruan tinggi. Ada beberapa cara dalam melakukan seleksi calon mahasiswa, diantaranya ada yang berdasarkan nilai rapor, ada yang melalui tes tertulis, dan ada jalur mandiri. Tidak semua perguruan tinggi melakukan pengolahan data yang masuk secara maksimal. Biasanya mereka masih mengolah data secara manual atau dengan kata lain belum menggunakan machine learning. Machine learning merupakan suatu sistem yang nantinya dapat belajar sendiri setelah diberikan data yang diperlukan. Bagi perguruan tinggi swasta, penerimaan mahasiswa baru merupakan tulang punggung untuk jalannya sebuah roda perputaran ekonomi di suatu perguruan tinggi. Apabila dari tahun ke tahun mahasiswa yang diterima banyak maka perguruan tinggi akan lebih mengelola keuangannya yang nanti kesejahteraan dosen dan karyawannya akan meningkat. Namun sebaliknya apabila di tiap tahunnya mahasiswa yang diterima sedikit maka perguruan tinggipun akan mengalami kesulitan dalam mengelola keuangannya, akibatnya perkembangan perguruan tinggi itu juga akan terhambat. Setiap perguruan tinggi diberi keluasaan dalam mengelola keuangannya sendiri, mulai dari biaya pendaftaran, biaya kuliah, dan biaya lain yang dibutuhkan.

Universitas Maritim AMNI Semarang merupakan suatu perguruan tinggi yang khusus dibidang kemaritiman. Karena kekhususannya inilah maka tidak semua orang tahu akan keberadaan universitas ini, sehingga setiap kali waktu penerimaan mahasiswa baru perlu adanya sosialisasi di sekolah – sekolah. Dengan adanya sosialisasi ini, maka

akan membuat banyak orang lebih mengenal dan tertarik di dunia maritim. Di Semarang perguruan tinggi dibidang maritim tidak hanya Universitas Maritim AMNI Semarang, tetapi ada beberapa perguruan tinggi lain seperti PIP Semarang, POLIMARIN, Akpelni Semarang. Dengan adanya perguruan tinggi maritim di Semarang lebih dari satu, maka sudah dipastikan untuk mendapatkan mahasiswa baru akan mengalami persaingan yang ketat. Untuk itu perlu ada cara ataupun metode khusus yang digunakan agar dapat memenangkan persaingan ini. Pada umumnya setiap perguruan tinggi dalam melakukan seleksi mahasiswa menggunakan berbagai tes. Ada tes lewat jalur akademik, ada tes lewat nilai raport, ada tes lewat jalur undangan. Khusus untuk perguruan maritim selain tes administrasi juga diberlakukan berbagai tes yang lain, diantaranya tes fisik, tes kesehatan, dan tes wawancara.

Meskipun perguruan tinggi berkeinginan mendapatkan mahasiswa sebanyak mungkin namun kualitas dari mahasiswa yang masuk juga harus dipertimbangkan. Oleh karena itu penelitian ini juga akan berkontribusi pada prediksi kelulusan mahasiswa tersebut dan memberikan arahan kepada mahasiswa yang bersangkutan agar lulus tepat waktu.

KAJIAN TEORITIS

Penelitian yang menggunakan klasifikasi data mining dengan metode C4.5 sudah ada yang melakukannya. Penelitian yang menggunakan metode Decision Tree juga dilakukan oleh Mashael A. Al-Barrak and Muna Al-Razgan. Penelitian mereka menggunakan Decision Tree bertujuan untuk memprediksi nilai IPK di akhir semester (Barrak, 2016). Hasil yang diperoleh dari penelitian mereka adalah didapatkan nilai akurasi sebesar 87%. Data set yang mereka gunakan yaitu nama siswa, ID siswa, IPK akhir, semester kelulusan, jurusan, kebangsaan, kampus, dan program yang diambil siswa. Kesimpulan yang didapat dari penelitian ini adalah metode yang digunakan dapat digunakan untuk memprediksi IPK akhir siswa berdasarkan nilai yang sudah ada dan metode ini juga dapat mengevaluasi program studi apa yang paling mempengaruhi dari nilai IPK. Penelitian tentang penerimaan mahasiswa baru juga dilakukan oleh Petrus Dwi Ananto Pamungkas. Dalam penelitiannya teknik yang digunakan adalah dengan algoritma pohon keputusan dengan aplikasi Rapid Miner 5.1. Tujuan dari penelitian yang dilakukan adalah untuk menghasilkan informasi yang mudah dan cepat tentang

faktor – faktor yang menjadi penentuan dalam mendaftar di perguruan tinggi. Data set yang digunakan adalah informasi dari teman, informasi dari saudara kandung, informasi dari kerabat sebelum gelombang ke tiga, dan informasi dari kerabat setelah gelombang ke tiga. Kesimpulan dari penelitian ini adalah sumber informasi yang diperoleh calon mahasiswa menjadi faktor utama untuk melanjutkan pendaftaran di perguruan tinggi (Pamungkas, 2015).

Pada tahun 2017 juga ada yang melakukan penelitian dengan menggunakan Decision Tree yaitu Durrotul Mukhibah dan Ana Kurniawati, mereka menggunakan data mining untuk memprediksi performance dari software engineer di PT Emerico. Hasil yang didapatkan dari penelitian mereka adalah tingkat akurasi dengan model Decision Tree adalah sebesar 97,0297% dengan tingkat kesalahan sebesar 2,9703% (Mukhibah dan Kurniawati, 2017). Penelitian yang menggunakan metode Decision Tree juga dilakukan oleh Patmi Kasih di tahun 2019. Judul penelitiannya adalah Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara. Data training yang dia gunakan sebanyak 60 record, dan atribut yang digunakan sebanyak 6 atribut yaitu Gender (G), Teknik (T), Interpretasi (Ip), Penampilan (P), Komitmen (K), Oktaf (Ot) dan 1 (satu) atribut target yaitu Hasil dengan class target adalah "L" untuk lulus (diterima) dan "TL" untuk tidak lulus (tidak diterima) (Kasih, 2019). Hasil akhir yang diperoleh dari penelitiannya yaitu dari 60 data record yang dijadikan data training, diperoleh tingkat akurasi 100% karena telah berhasil mengcover seluruh data training yang ada. Metode Decision Tree juga digunakan oleh Triuli Novianti dan Iwan Santosa dalam penelitian mereka. Mereka menggunakan Decision Tree untuk meneliti penentuan jadwal kerja karyawan di Universitas Muhammadiyah Surabaya (Novianti, dkk, 2016). Jumlah data yang mereka gunakan sebanyak 103 data karyawan. Data training dan testing yang digunakan yakni umur, jenis kelamin dan unit kerja. Dari 103 data kotor setelah dilakukan pembersihan data tidak lengkap dan melalui proses seleksi sesuai yang dibutuhkan dalam analisa jadwal kinerja karyawan maka didapat seluruh data lengkap sehingga dapat digunakan sebanyak 103 data bersih. Dari data yang ada dilakukan split untuk digunakan 1/3(34) data training dan 2/3(69) data test. Pada pengujian ini menggunakan cross validation 5 fold. Hasil pengujian keseluruhan data dengan memakai cross validation 5 fold di dapat akurasi pengujian sebesar 70%.

Algoritma C4.5 mempunyai keunggulan dibandingkan dengan algoritma yang lain yaitu mempunyai kemampuan untuk mengolah data set kecepatan klasifikasi, setiap atribut bersifat diskrit, binari dan kontinue, serta transparansi pengetahuan atau klasifikasi. (Kotsiantis, 2028). Selain kelebihan tentunya algoritma C4.5 juga terdapat kelemahan yaitu terletak pada pembacaan data yang berjumlah besar. Sehingga untuk meningkatkan kinerja klasifikasi, algoritma C4.5 perlu ditambahkan algoritma lain. Dalam penelitian ini, algoritma Particle Swarm Optimization (PSO) sebagai fitur seleksi sangat efektif karena hanya membutuhkan parameter sedikit waktu komputasi. Penelitian yang dilakukan oleh Dwi Meylitasari Br. Tarigan dkk, yang melakukan penelitian berupa seleksi fitur pada klasifikasi penyakit gula darah menggunakan PSO pada algoritma C4.5. Hasil dari penelitian tersebut yaitu PSO dapat meningkatkan kinerja akurasi C4.5 dari 86% menjadi 95%. (Tarigan & Rini, 2020).

METODE PENELITIAN

Data merupakan fakta – fakta yang digambarkan lewat angka, simbol, kode, dan lainnya. Sebelum digunakan data perlu dikelompokkan terlebih dahulu. Bentuk data dikelompokkan menjadi beberapa jenis, yaitu berdasarkan sumbernya, berdasarkan sifatnya, berdasarkan cara memperolehnya, dan berdasarkan waktu pengumpulannya.

Jenis yang pertama adalah berdasarkan sumber pengambilan data yaitu data internal dan data eksternal. Data internal yaitu data yang didapatkan dari suatu instansi atau suatu organisasi di tempat penelitian. Contohnya : data jumlah karyawan. Sedangkan data eksternal adalah data yang didapatkan diluar suatu instansi atau organisasi dari tempat penelitian. Contohnya : data hasil penjualan dari sebuah instansi.

Jenis data yang kedua yaitu berdasarkan dari sifat datanya, yaitu data kualitatif dan data kuantitatif. Data kualitatif yaitu data yang tidak berbentuk bilangan. Contohnya : data agama, warna, atau jenis kelamin. Sedangkan data kuantitatif yaitu data yang berbentuk bilangan. Contohnya : umur, panjang, ataupun tinggi.

Jenis data yang ketiga yaitu berdasarkan cara memperolehnya, yaitu data primer dan data sekunder. Data primer adalah data yang didapat secara langsung di lapangan. Data primer juga biasa disebut dengan data asli. Data sekunder adalah data yang didapat oleh orang yang melakukan penelitian dari sumber – sumber yang sudah ada. Data

sekunder ini biasanya didapat dari perpustakaan atau dari berbagai laporan penelitian yang sudah dilakukan.

Jenis data yang keempat yaitu berdasarkan waktu pengumpulannya, terdapat dua kategori yaitu data *Cross Section* dan data *Time Series*. Data *Cross section* yaitu data yang menggambarkan suatu keadaan dalam waktu tertentu. Contohnya adalah data dari hasil kuesioner oleh responden dalam waktu yang sudah ditentukan. Sedangkan data *Time Series* yaitu data yang dikumpulkan dari waktu ke waktu dengan tujuan untuk dapat melihat perkembangan dari objek yang diteliti. Contohnya data sensus suatu penduduk di suatu daerah.

Data yang Digunakan

Data yang digunakan dalam penelitian ini berasal dari data privat yaitu dari Seleksi masuk taruna baru di Universitas Maritim AMNI Semarang berjumlah 1604 record yang terdiri dari 17 Atribut, dengan 9 atribut bertipe numeric dan 8 atribut bertipe kategorikal.

Data disimpan dalam format excel seperti terlihat pada gambar 1 berikut.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	tahun	lhr	jk	BB	TB	asal daerah	Pulau	provinsi	kategori s	jurusan	tahun lulu	tahun daftar	asal ortu	TKD	wawancara	SAMAP
2	1998	pria	kurang	lebih dari	luar semarai	Jawa	Jateng	SMA	IPA	0	2016	Pulau Jaw	5	2	5	
3	1998	pria	kurang	lebih dari	luar semarai	Jawa	Jateng	SMA	IPS	0	2016	Pulau Jaw	5	4	2	
4	1998	pria	kurang	lebih dari	luar semarai	Jawa	Jabar	SMK	Teknik Ke	0	2016	Pulau Jaw	4	4	2	
5	1997	pria	ideal	lebih dari	luar semarai	Jawa	Jabar	SMA	IPA	2020	2015	Pulau Jaw	5	5	4	
6	1998	pria	ideal	lebih dari	luar jawa	Sumatera	sumbar	SMK	Teknik Ke	2020	2016	Luar pulau	5	3	4	
7	1999	pria	ideal	lebih dari	luar semarai	Jawa	Jatim	SMK	Teknik Ke	2020	2016	Pulau Jaw	5	4	4	
8	1998	pria	ideal	lebih dari	luar jawa	Sumatera	lampung	SMA	IPS	2020	2016	Luar pulau	5	4	3	
9	1998	pria	ideal	lebih dari	luar jawa	Sumatera	Riau	SMA	IPS	2020	2016	Luar pulau	5	4	3	
10	1998	pria	ideal	lebih dari	luar jawa	Sulawesi	SULAWESI TENGAH	SMK	Teknik Pe	2020	2016	Luar pulau	5	3	4	
11	1998	pria	ideal	lebih dari	luar jawa	Sumatera	Lampung	SMA	IPS	2020	2016	Luar pulau	5	5	4	
12	1999	pria	kurang	kurang tin	luar semarai	Jawa	Jateng	SMK	Nautika k	0	2016	Pulau Jaw	5	2	3	
13	1997	pria	ideal	lebih dari	luar semarai	Jawa	Jateng	SMA	IPA	2020	2015	Pulau Jaw	5	4	4	
14	1998	pria	ideal	lebih dari	luar semarai	Jawa	Jatim	SMA	IPA	2020	2016	Pulau Jaw	4	4	5	
15	1998	pria	kurang	lebih dari	luar semarai	Jawa	Jateng	SMK	Teknik Ke	0	2016	Pulau Jaw	5	4	2	
16	2000	pria	kurang	lebih dari	luar semarai	Jawa	Jatim	SMA	IPA	0	2016	Pulau Jaw	4	3	2	
17	1997	pria	ideal	lebih dari	luar semarai	Jawa	Jatim	SMA	IPA	2020	2015	Pulau Jaw	5	5	4	
18	1998	pria	ideal	lebih dari	luar jawa	bangka	bangka belitung	SMA	IPA	2020	2016	Luar pulau	5	3	4	
19	1998	pria	ideal	kurang tin	luar jawa	Sumatera	Bengkulu	SMA	IPS	0	2016	Luar pulau	2	4	4	
20	1998	pria	kurang	lebih dari	luar semarai	Jawa	Banten	MA	IPA	0	2016	Pulau Jaw	2	4	3	
21	1998	pria	ideal	lebih dari	luar semarai	Jawa	Jatim	SMA	IPS	2020	2016	Pulau Jaw	5			

Gambar 1 Data Set dalam format excel

Metode yang digunakan dalam penelitian ini adalah eksperimen. Eksperimen dilakukan dengan tahapan training dan testing. Data yang digunakan untuk penelitian bersumber dari Universitas Maritim AMNI Semarang dari tahun 2015 sampai 2016. Jumlah data yang digunakan untuk training sebanyak 200 data, sedangkan jumlah data

yang digunakan untuk testing sebanyak 20 data. Metode yang digunakan dalam penelitian ini adalah Decision Tree C4.5 dan Particle Swarm Optimiztion (PSO).

Alat Penelitian

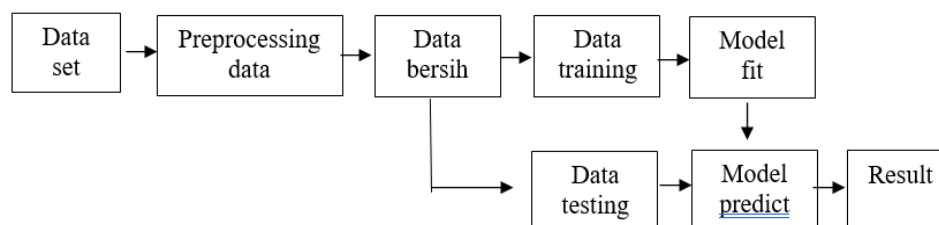
Dalam penelitian ini, peneliti menggunakan spesifikasi software dan hardware sebagai alat bantu seperti yang tercantum dalam Tabel 1.

Tabel1 Alat Penelitian

Software	Hardware
System operasi : Windows 10 Home	Processor : Intel(R) Core i3 @120 Ghz
Data mining : RapidMiner versi 5.3	RAM : 4.00 GB

Alur Eksperimen

Penelitian yang dilakukan agar dapat berjalan dengan baik maka perlu dibuat peencanaan terlebih dahulu. Perencanaan yang dimaksud adalah dengan membuat tahap – tahap penelitian



Gambar 2 Alur Eksperimen

Dari gambar 2 Alur eksperimen dapat dijelaskan bahwa awal mula eksperimen dimulai dari data mentah yang kemudian data mentah in akan dilakukan preprocessing terlebih dahulu. Setelah dilakukan preprocessing akan dihasilkan data yang bersih. Selanjutnya untuk dapat mengetahui seberapa bagus model yang kita buat, data bersih akan dibuat menjadi dua yaitu data training dan data testing.

Data training selanjutnya akan dilakukan model pembelajaran yaitu dengan model fit dulu atau di klasifikasikan setelah itu jika ingin mengetahui performa dari model pembelajaran maka dimasukkan data testing yang selanjutnya akan dilakukan prediksi. Dari hasil prediksi akan keluar hasilnya.

Hasil yang keluar akan dilihat dari akurasi prediksi, precision, dan recall. Jika hasilnya hasilnya baik maka akan dapat digunakan untuk sistem pendukung keputusan bagi perguruan tinggi untuk dimasa mendatang.

1. PENGUJIAN

Penelitian ini bertujuan untuk membangun model klasifikasi seleksi penerimaan taruna baru menggunakan metode C4.5 dengan *Particle Swarm Optimization* dengan performa yang baik. Dari metode C4.5 dengan PSO akan diperoleh tingkat akurasi yang baik sehingga dapat diperoleh model klasifikasi.

Parameter C4.5

Penggunaan parameter akan berpengaruh pada hasil akurasi yang akan dihasilkan oleh algoritma C4.5. Setting parameter yang akan digunakan seperti pada tabel 2.

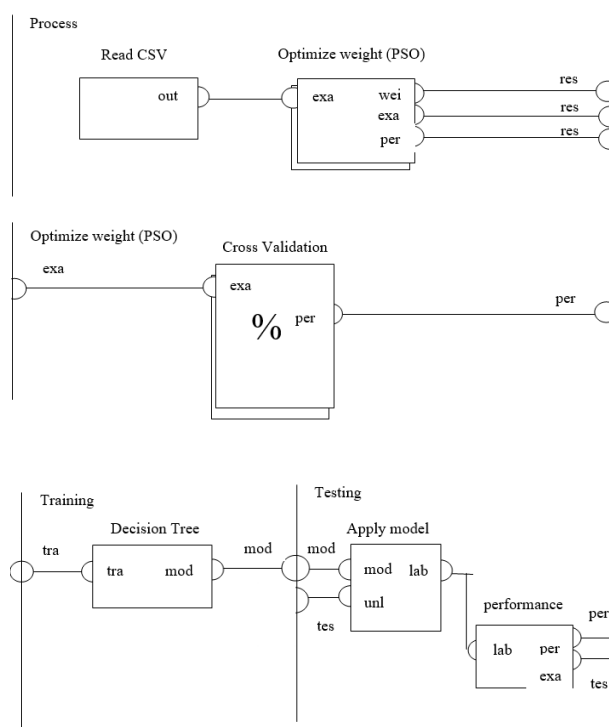
Tabel 2. Setting Parameter C4.5

Criterion	Gain ratio
Maximal depth	10
Apply pruning Confidence	0.1
Apply prepruning Minimal gain	0.01
Minimal leaf size	2
Minimal size for split	4
Number of pruning alternatives	3

Setelah dilakukan pen-settingan parameter C4.5 maka akan dilanjutkan dengan pemrosesan data.

Metode C4.5 + PSO

Pada desain model C4.5 berbasis PSO pada pembobotan atribut seperti pada gambar 3.



Gambar 3 Desain Model C4.5 berbasis PSO

Data yang masuk ke PSO berupa example set, kemudian untuk output dari PSO ada tiga yaitu weights, example set, dan performance. PSO sebelum dilakukan pemrosesan data maka perlu di setting parameter dahulu seperti yang terlihat pada tabel 3

Tabel 3 Parameter PSO

Population size	5
Maximum number of generation	30
Inertia weight	1.0
Local best weight	1.0
Global best weight	1.0
Min weight	0.0
Max weight	1.0

Di dalam proses PSO terdapat proses berupa Cross validation. Dimana Setting di dalam Cross validation yaitu Number of folds yaitu 10 dan sampling type yaitu automatic. Data yang masuk berupa example set dan hasil dari cross validation berupa performance. Proses dalam Cross Validation ada 2 proses, yaitu training dan testing. Pada bagian proses testing diisi dengan model Decision Tree.

Klasifikasi Decision Tree mendapat sebuah input berupa data training dan terdapat keluaran berupa model, dimana model ini akan dihubungkan ke bagian proses Testing. Pada bagian proses Testing terdapat dua proses lagi yaitu Apply model dan Performance. Pada Apply model mendapat dua masukan yaitu data model akan masuk ke model dan data tes akan masuk ke unlabel, yang dikeluarkan dari apply model yaitu data lab. Apply model berguna untuk menjalankan algoritma / algoritma C4.5. Pada Performance mendapat satu masukan dari apply model berupa data lab, yang kemudian dari performance akan dihasilkan dua luaran yaitu hasil akurasi dan hasil dari sampel. Performance berguna untuk mengukur performa dari model C4.5.

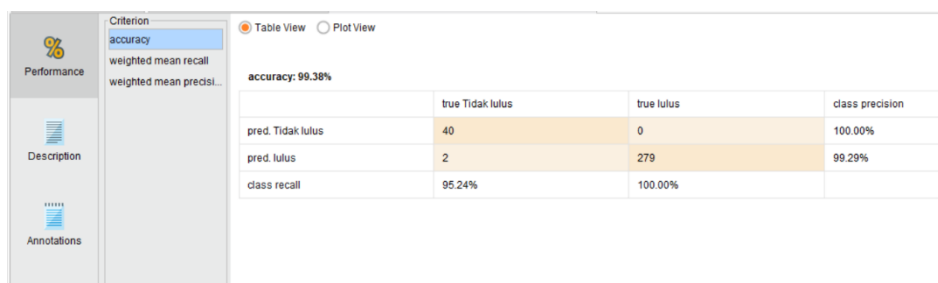
HASIL DAN PEMBAHASAN

Percobaan pada penelitian dilakukan dengan menggunakan tools rapidminer dengan menguji metode C4.5 dengan optimasi PSO.

Hasil pemodelan C4.5 dengan Optimasi PSO

C4.5

Hasil dari pemodelan C4.5 terlihat seperti pada gambar 4



	true Tidak lulus	true lulus	class precision
pred. Tidak lulus	40	0	100.00%
pred. lulus	2	279	99.29%
class recall	95.24%	100.00%	

Gambar 4 Hasil Pemodelan C4.5

Berdasarkan gambar 4.19 diketahui bahwa prediksi tidak lulus dengan true tidak lulus terdapat 40 item, sedangkan prediksi tidak lulus dengan true lulus terdapat 0 item. Sehingga prediksi tidak lulus memiliki class precision sebesar 100%.

Prediksi lulus dengan true tidak lulus terdapat 2 item dan prediksi lulus dengan true lulus terdapat 279 item, sehingga prediksi lulus pada class precision sebesar 99,29%. Dari hasil klasifikasi terlihat bahwa nilai akurasi dari model tersebut adalah sebesar 99,38% artinya keakuratan dalam klasifikasi sudah cukup baik.

C4.5 Optimasi PSO

Klasifikasi ini diawali dengan uji coba dengan memberi nilai pada parameter population size dan maximum number of generation bernilai default yaitu 5 dan 30.

Tabel 4. Hasil Percobaan dengan population size dan maximum number of generation secara berbeda

Akurasi	Lama Waktu Eksekusi	Parameter PSO
99,94%	24 s	default
99,94%	38 s	Posize=10, generate=30
99,94%	56 s	Posize=10, generate=40
99,94%	2m 4s	Posize=15, generate=40
99,94%	2m 14s	Posize=15, generate=50
99,94%	4m 42s	Posize=20, generate=50
99,94%	5m 20s	Posize=25, generate=50
99,94%	6m 34s	Posize=30, generate=50

Tabel 4 menunjukkan algoritma C4.5 + PSO menghasilkan nilai akurasi terbaik pada saat population size bernilai 5 dan maximum number of generation bernilai 30 yaitu akurasi bernilai 99,94% dengan waktu eksekusi 24 detik.

Tabel 4 juga menunjukkan pada saat population size bernilai 5 dan maximum number of generation bernilai 30 ternyata sudah konvergen artinya hasil sudah maksimal jika dilakukan percobaan lagi terjadi overfitting (kelebihan iterasi) yang dampaknya waktu eksekusi meningkat tetapi hasil akurasi tidak meningkat.

Hasil dari pemodelan C4.5 + PSO terlihat seperti pada gambar 5

accuracy: 99.94% +/- 0.20% (micro average: 99.94%)

	true Tidak lulus	true lulus	class precision
pred. Tidak lulus	210	0	100.00%
pred. lulus	1	1393	99.93%
class recall	99.53%	100.00%	

Gambar 5 Hasil Pemodelan C4.5 + PSO

Berdasarkan gambar 5 diketahui prediksi tidak lulus dengan true tidak lulus terdapat 210 item. Sedangkan prediksi tidak lulus dengan true lulus terdapat 0 item, sehingga prediksi tidak lulus mempunyai class precision sebesar 100%.

Prediksi lulus dengan true tidak lulus terdapat 1 item. Sedangkan prediksi lulus dengan true lulus terdapat 1393 item, sehingga prediksi lulus mempunyai class precision sebesar 99,93%. Dari hasil klasifikasi tersebut terlihat bahwa nilai akurasi adalah 99,94% artinya keakuratan dalam klasifikasi tersebut menjadi lebih baik dai decision tree yang tanpa optimasi PSO.

Setelah melakukan pemodelan dan perhitungan berdasar algoritma C4.5 dengan C4.5 + PSO kemudian dilakukan perbandingan hasil akurasi seperti terlihat pada tabel 4.

Pembobotan atribut sangat diperlukan karena tidak semua atribut mempunyai pengaruh terhadap hasil akurasi. Hasil dari pembobotan atribut oleh PSO pada saat akurasi terbaik dapat dilihat pada tabel 5

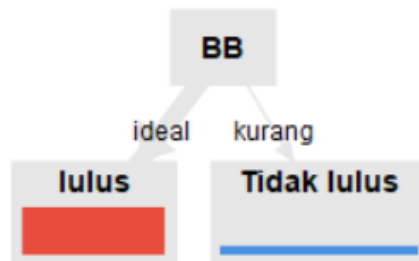
Tabel 5 Hasil Pembobotan Atribut

Atribut	Bobot
Tahun lhr	0.0766
jk	0.719
BB	1
TB	0.042
Asal daerah	0.867
Pulau	0.868
Kategori sekolah	0.035
Jurusan	0.124
Tahun lulus	0.211
Tahun daftar	0.098
Asal ortu	0
TKD	0
Wawancara	0.590
Samapta	0.0
Psikotes	0.247
Kesehatan	0
Spi	0.131

Tabel 5 memperlihatkan ada 17 atribut yang digunakan, dan atribut yang bobotnya bernilai 0 terdapat 16. Atribut yang bobotnya bernilai 0 tidak berpengaruh terhadap akurasi. 14 atribut yang mempunyai bobot 0 yaitu tahun lhr, jk, TB, Asal daerah, pulau, kategori sekolah, jurusan, tahun lulus, tahun daftar, asal ortu, TKD, wawancara, samapta, psikotes, kesehatan, spi.

Bobot dari atribut yang bernilai 1 dapat dimungkinkan mempengaruhi hasil akurasi secara signifikan. Terdapat 1 atribut yang bobotnya bernilai 1 yaitu BB.

Hasil Pemodelan Pohon Keputusan dan Rule



Gambar 6 Hasil Pemodelan Tree C4.5

Rule yang tercipta dari gambar 4.5 hasil pemodelan tree C4.5 adalah sebagai berikut.

- 1) Jika BB = ideal maka lulus
- 2) Jika BB = kurang ideal maka tidak lulus

KESIMPULAN DAN SARAN

Penelitian ini menggunakan pemodelan dengan algoritma C4.5 dengan C4.5 + PSO. Data yang digunakan adalah data privat dari Universitas Maritim AMNI Semarang. Penelitian ini difokuskan pada penerapan algoritma PSO dengan pembobotan atribut teknik klasifikasi data mining C4.5. Untuk validasi model menggunakan *10fold cross-validation* dan untuk evaluasi model menggunakan *confusion matrix*.

Hasil dari penelitian menunjukkan bahwa model C4.5 + PSO mempunyai akurasi sebesar 99.94% dan model C4.5 mempunyai akurasi sebesar 99.38%. sehingga terjadi peningkatan akurasi sebesar 0,60 %.

Dengan demikian terbukti bahwa PSO yang diterapkan pada pembobotan atribut C4.5 meningkatkan nilai akurasi. Hal ini dapat menjadikan C4.5 + PSO sebagai pendukung keputusan dalam menentukan seleksi penerimaan taruna baru di Universitas Maritim AMNI Semarang.

DAFTAR REFERENSI

- King Saud University, Saudi Arabia, M. A. Al-Barrak, M. Al-Razgan, And King Saud University, Saudi Arabia, 'Predicting Students Final GPA Using Decision Trees: A Case Study', *Int. J. Inf. Educ. Technol.*, Vol. 6, No. 7, Pp. 528–533, 2016, Doi: 10.7763/IJIE.T.2016.V6.745.
- P. D. A. Pamungkas, 'Menentukan Kemungkinan Masuknya Calon Mahasiswa Baru Pada Sebuah Perguruan Tinggi Swasta Menggunakan Teknik Klasifikasi Pohon Keputusan Dengan Aplikasi Rapidminer 5.', *J. Inform.*, P. 11, 2015.
- D. Mukhibah And A. Kurniawati, 'Implementasi Data Mining Dalam Prediksi Performance Software Engineer Pt. Emerio Menggunakan Decision Tree', Vol. 22, No. 1, P. 13, 2017.
- P. Kasih, 'Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara', Vol. 1, No. 2, P. 7, 2019.
- T. Novianti, I. Santosa, J. Sutorejono, And U. Trunojoyo, 'Penentuan Jadwal Kerja Berdasarkan Klasifikasi Data Karyawan Menggunakan Metode Decision Tree C4.5 (Studi Kasus Universitas Muhammadiyah Surabaya)', *J. Komun.*, Vol. 5, No. 1, P. 10, 2016.
- S. B. Kotsiantis, 'Supervised Machine Learning: A Review Of Classification Techniques', P. 20.
- D. M. B. Tarigan And D. P. Rini, 'Seleksi Fitur Pada Klasifikasi Penyakit Gula Darah Menggunakan Particle Swarm Optimization (PSO) Pada Algoritma C4.5', Vol. 4, No. 3, P. 7, 2020.
- A. Saifudin, 'Metode Data Mining Untuk Seleksi Calon Mahasiswa Pada Penerimaan Mahasiswa Baru Di Universitas Pamulang', *J. Teknol.*, Vol. 10, No. 1, P. 12, 2018.
- A. Suryadi And E. Harahap, 'Sistem Rekomendasi Penerimaan Mahasiswa Baru Menggunakan Naive Bayes Classifier Di Institut Pendidikan Indonesia', *Joutica*, Vol. 3, No. 2, P. 171, Sep. 2018, Doi: 10.30736/Jti.V3i2.231.