

PENERAPAN *PARTICLE SWARM OPTIMIZATION* PADA ALGORITMA C 4.5 UNTUK SELEKSI PENERIMAAN KARYAWAN

Agus Wiyatno

Program Studi Teknik Informatika
STMIK Nusa Mandiri Jakarta, Indonesia
Email : agus.agq@nusamandiri.ac.id

ABSTRACT

The Employees are the most vital element of the company as they had a big contribution and involved almost for all section on how the company will go up and down. Employees and the company affect the efficiency, effectiveness, designing, producing goods and services, oversee the quality, market products, allocating financial resources, and determines the overall goals and strategies of the organization. Therefore, organizations need accurate information and sustainable in order to get suitable candidates with the qualifications of the organization. Model algorithms are widely used in research related to the employee is C4.5 decision tree classification model. Advantages of using a decision tree classification models are the result of the decision tree is simple and easy to understand. Many studies using the method of decision tree and classification tree in predicting the employees selection but results the accuracy of the resulting value is less accurate. In this study created a C 4.5 Algorithm model and C 4.5 Algorithm model based on particle swarm optimization to get the rule in employees selection and provide a more accurate value of accuracy. After testing C 4.5 algorithm model based on particle swarm optimization, Implementation of particle swarm optimization can produce accuracy value of C 4.5 algorithm model from 80.80 % to 85.20 % and the AUC value from 0.878 to 0.891. By the formation the model selection of employees, the company can be helped for employee selection.

Keyword: Seleksi Karyawan, Algoritma C 4.5, Particle Swarm Optimization

1. PENDAHULUAN

Karyawan adalah unsur yang paling vital dalam organisasi yang berperan besar bagi kesuksesan organisasi. Karyawan dan perusahaan mempengaruhi efisiensi, efektivitas, merancang, memproduksi barang dan jasa, mengawasi kualitas, memasarkan produk, mengalokasikan sumber daya finansial, serta menentukan seluruh tujuan dan strategi organisasi (Ike Kusdyah Rachmawati, 2008). Karyawan yang berkualitas tentunya tidak didapat dengan mudah dan sederhana.

Karyawan sebuah organisasi merupakan sumber daya yang vital dan akan hanya diperoleh melalui upaya perekrutan yang efektif. Untuk itu, organisasi memerlukan informasi akurat dan berkelanjutan guna mendapatkan calon karyawan yang sesuai dengan kualifikasi organisasi (Ike Kusdyah Rachmawati, 2008).

Model seleksi karyawan sangat diperlukan karena sebagai tindakan dalam seleksi karyawan yang memenuhi kualifikasi dan standar kualitas demi tercapainya tujuan organisasi. Organisasi atau perusahaan dapat menseleksi penerimaan karyawan sebagai pola untuk mendapatkan karyawan-karyawan yang berkualitas dan memenuhi kualifikasi.

Model algoritma yang banyak digunakan dalam penelitian yang berhubungan dengan karyawan adalah model klasifikasi *decision tree C4.5*. Kelebihan menggunakan model klasifikasi *decision tree* ini adalah hasil pohon keputusannya sederhana dan mudah dimengerti. Menurut (Han & Kamber, 2006 : p292) proses learning dan klasifikasi pada algoritma *Decision Tree* sederhana dan cepat. Secara umum model algoritma klasifikasi *Decision Tree* mempunyai tingkat akurasi yang tinggi. Namun masalahnya menurut (Chawla, 2003) klasifikasi *Decision tree* biasanya dievaluasi dengan akurasi prediksi yang mempertimbangkan semua kesalahan yang sama akan tetapi akurasi prediksi mungkin tidak sesuai jika menggunakan *imbalance* dataset. Beberapa penggunaan metode klasifikasi *Decision Tree C4.5* telah dilakukan beberapa orang dalam penelitian seperti:

1. Algoritma C 4.5 untuk penilaian kinerja karyawan. Penelitian dilakukan oleh Windy Julianto, Rika Yunitarini, Mochamad Kautsar Sophan, tentang penilaian kinerja karyawan menggunakan Algoritma C 4.5
2. Penelitian yang dilakukan oleh Alao D. dan Adeyemo A. B. dengan topik *Analyzing Employee Attrition Using Decision Tree Algorithm*. Penelitian

ini membahas tentang analisis pengurangan karyawan menggunakan algoritma *decision tree*.

3. Penelitian yang dilakukan oleh N. Sivaram dan K. Ramar dengan topic *Applicability of clustering and classification algorithms for recruitment data mining*. Penelitian ini membahas tentang pengaplikasian klustering dan klasifikasi untuk proses rekrutmen menggunakan ID3, C 4.5, C 4.5 *unpruned*, Cart, dan Cart *unpruned*.

4. Penelitian yang dilakukan oleh Qasem A. Alradaideh dan Eman Al Nagi dengan topik *Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance*. Penelitian ini membahas tentang penggunaan teknik data mining untuk membangun model klasifikasi dalam memprediksi kinerja karyawan.

5. Penelitian yang dilakukan oleh Firmansyah dengan topik Penerapan Algoritma C 4.5 untuk penentuan kelayakan pemberian kredit koperasi.

Karena penelitian seleksi karyawan pernah dilakukan beberapa peneliti sebelumnya banyak menggunakan klasifikasi *decision tree* C 4.5 dan hasil akurasi masih kurang akurat, maka pada penelitian ini akan diukur akurasi untuk proses seleksi penerimaan karyawan menggunakan algoritma klasifikasi *decision tree* yang tepat dengan kriteria atribut data yang digunakan berbeda. Penerapan *Particle Swarm Optimization* diharapkan akan menghasilkan akurasi yang lebih akurat.

Adapun identifikasi masalah pada penelitian yang penulis rumuskan adalah :

1. Organisasi membutuhkan sebuah model yang dapat menyeleksi penerimaan karyawan. Algoritma C 4.5 memiliki hasil yang lebih baik dibandingkan algoritma lain. Akan tetapi *Particle Swarm Optimization* akan digunakan sebagai algoritma optimasi dalam seleksi penerimaan karyawan untuk mencapai tingkat akurasi yang lebih baik.

2. Akurasi dari penelitian ini adalah untuk seleksi penerimaan karyawan.

Tujuan dari penelitian ini adalah untuk menerapkan *Particle Swarm Optimization* untuk meningkatkan akurasi dari algoritma C 4.5 untuk seleksi penerimaan karyawan.

2. LANDASAN TEORI

Studi literatur mengenai pembahasan prediksi menggunakan algoritma C 4.5 telah dilakukan pada penelitian sebelumnya. Berikut penelitian terdahulu

yang terkait dengan topik dalam proposal tesis ini yaitu :

1. Algoritma C 4.5 untuk penilaian kinerja karyawan. Penelitian dilakukan oleh Windy Julianto, Rika Yunitarini, Mochamad Kautsar Sophan, tentang penilaian kinerja karyawan menggunakan Algoritma C 4.5 dihitung dengan menggunakan teknik *Confusion Matrix* dengan nilai *precision* sebesar 58,33 %, *Recall* 82,35 %, *Accuracy* 91,39 % dan *Error Rate* sebesar 8,61 %.

2. Penelitian yang dilakukan oleh Alao D. dan Adeyemo A. B. dengan topik *Analyzing Employee Attrition Using Decision Tree Algorithm*. Penelitian ini membahas tentang analisis pengurangan karyawan menggunakan algoritma *decision tree*. Pada penelitian ini penulis menggunakan perbandingan 3 algoritma, algoritma C 4.5, REPTree, dan CART(SimpleCart) dengan *accuracy* C 4.5 sebesar 67 %, REPTree 62 %, dan CART 64 %.

3. Penelitian yang dilakukan oleh N. Sivaram dan K. Ramar dengan topic *Applicability of clustering and classification algorithms for recruitment data mining*. Penelitian ini membahas tentang pengaplikasian klustering dan klasifikasi untuk proses rekrutmen menggunakan ID3, C 4.5, C 4.5 *unpruned*, Cart, dan Cart *unpruned* dengan masing-masing akurasi ID3 sebesar 45,12 %, C 4.5 sebesar 77,29 %, C 4.5 *unpruned* sebesar 76,73 %, Cart sebesar 72,12 %, dan Cart *unpruned* 72,75 %.

4. Penelitian yang dilakukan oleh Qasem A. Alradaideh dan Eman Al Nagi dengan topik *Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance*. Penelitian ini membahas tentang penggunaan teknik data mining untuk membangun model klasifikasi dalam memprediksi kinerja karyawan.

5. Penelitian yang dilakukan oleh Firmansyah dengan topik Penerapan Algoritma C 4.5 untuk penentuan kelayakan pemberian kredit koperasi dengan nilai *accuracy* 90,00 %.

Algoritma Klasifikasi

Klasifikasi merupakan salah satu tujuan yang banyak dihasilkan dalam *data mining*. Klasifikasi merupakan proses pengelompokan sebuah variabel ke dalam kelas yang sudah ditentukan (Larose, 2005). *Data mining* mampu mengolah data dalam jumlah besar, setiap data terdiri dari kelas tertentu bersama dengan variabel dan faktor-faktor penentu kelas variabel tersebut. Dengan data mining peneliti dapat menentukan suatu kelas dari variabel data yang dimiliki.

Teknik klasifikasi merupakan suatu pendekatan sistematis untuk membangun model klasifikasi dari suatu himpunan data masukan. Tiap

teknik menggunakan suatu algoritma pembelajaran (*learning algorithm*) untuk mendapatkan suatu model yang paling memenuhi hubungan antara himpunan atribut dan label kelas dalam data masukan (Fajar Astuti Hermawati, 2013).

Ada banyak algoritma klasifikasi yang dapat digunakan dalam *data mining*. Mulai dari *k-nearest neighbor*, *decision tree*, *neural network* dan lainnya. Dalam penelitian terkait disimpulkan bahwa algoritma pohon keputusan dengan C 4.5 memiliki hasil yang baik untuk proses klasifikasi penerimaan karyawan.

Algoritma C 4.5

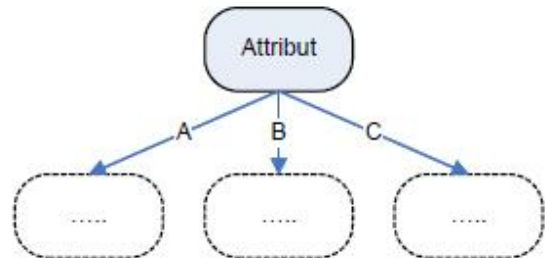
C 4.5 merupakan salah satu metode yang digunakan menginduksi pohon keputusan yang ditemukan J. Ross Quinlan. Algoritma ini diturunkan dari algoritma ID3 yang populer digunakan dalam membuat pohon keputusan. C 4.5 merupakan algoritma yang cocok digunakan untuk mengklasifikasi data dalam jumlah besar kedalam kelas-kelas tertentu berdasarkan pola data yang ada (Wu & Kumar, 2009). Di dalam *data mining* dan *machine learning* C 4.5 digunakan untuk mempelajari data dalam jumlah besar, membuat model pembelajaran berupa pohon keputusan yang dapat diterapkan untuk memprediksi data yang belum muncul.

Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda (Fajar Astuti Hermawati, 2013). Dan Pohon keputusan merupakan sekumpulan pertanyaan yang disusun secara sistematis, dimana setiap pertanyaan dibentuk berdasarkan nilai dari atribut yang sedang diuji. Jawaban dari pertanyaan tersebut akan berlanjut pada pertanyaan lain dan seterusnya hingga berhenti pada label daun yang berarti kelas dari variabel. Sekumpulan pertanyaan ini digambarkan dalam bentuk diagram pohon yang sangat mudah dimengerti. Dalam diagram pohon, akar pohon digambarkan sebagai pertanyaan pertama, dan setiap percabangan yang muncul akan disebut cabang pohon yang terdiri dari pengujian terhadap nilai dalam atribut yang diuji. Cabang yang ada akan bercabang hingga cabang terakhir disebut daun. Daun merupakan label jenis data yang sedang diuji, dapat dikatakan sebagai hasil klasifikasi ataupun hasil prediksi data.

Dalam proses pengujian atribut, cabang baru yang terbentuk akan diperhatikan dari tipe atribut

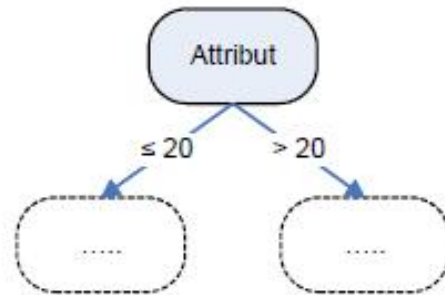
(Han & Kamber, 2006). Berikut 3 jenis cabang yang mungkin timbul dalam pohon keputusan :

1. Jika atribut bernilai diskrit, maka cabang yang terbentuk akan selalu sama dengan jumlah variasi nilai yang terdapat pada atribut tersebut.



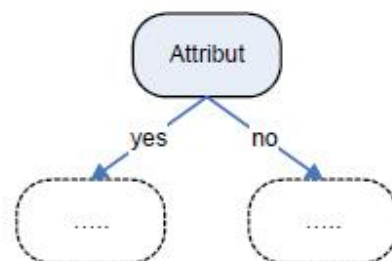
Gambar 2.1 Cabang pohon variasi nilai atribut

2. Jika cabang bernilai kontinu, maka akan dipecahkan menurut titik perpecahan, sedangkan titik perpecahan dikalkulasi dengan masing-masing algoritma penyusunan pohon keputusan. Cabang perpecahan yang terbentuk akan berpola seperti $<$ atribut, dan satu cabang lagi $>$ atribut.



Gambar 2.2 Cabang pohon nilai kontinu

3. Jika atribut yang diuji bernilai biner, maka cabang yang terbentuk pasti dua dan melibatkan ya atau tidak



Gambar 2.3 Cabang pohon nilai biner

Particle Swarm Optimization

Particle Swarm Optimization merupakan teknik optimisasi berbasis populasi yang dikembangkan oleh Eberhart dan Kennedy pada tahun 1995. PSO mensimulasikan perilaku dari sekelompok burung yang mencari makanan.

Sekelompok burung sedang mencari sepotong makanan didalam suatu area. Semua burung dalam kelompok tersebut tidak tahu dimana letak makanan tersebut, dalam proses pencarian makanan beberapa burung secara mendadak berpisah dari kawanan dan membentuk kawanan baru dan kembali berkelompok. Proses pengelompokkan burung ini bertujuan untuk menjaga jarak optimum antara makanan dengan burung tersebut dan kawanan burung lainnya. Cara yang paling efektif untuk menemukan potongan makanan tersebut adalah dengan mengikuti burung yang jaraknya paling dengan dengan makanan tersebut.

Perilaku yang sama juga digambarkan pada kelompok hewan lainnya seperti ikan. Dengan berkelompok, kawanan ikan mengalami persaingan dalam membagi makanan, namun kawanan tersebut juga menjadi lebih mudah untuk menemukan makanan baru didalam daerah yang persebaran makanannya tidak diketahui (Kennedy & Eberhart, 1995). Metode tersebut diadopsi dalam algoritma PSO sendiri yaitu melakukan perulangan untuk mengoptimasi sebuah masalah dengan memberikan nilai kualitas. Dengan begitu dalam data yang besar dapat diketahui posisi terbaik dalam perhitungan pengolahan data.

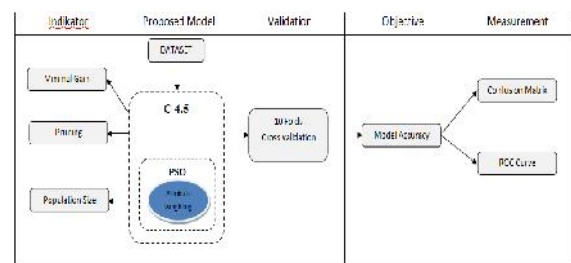
Particle Swarm Optimization, didasarkan pada perilaku sebuah kawanan burung atau ikan. Algoritma PSO meniru perilaku social organism ini. Perilaku social terdiri dari tindakan individu dan pengaruh dari individu-individu lain dalam suatu kelompok (Budi Santosa & Paul Willy, 2011)

KERANGKA PEMIKIRAN

penelitian ini bertujuan untuk meningkatkan akurasi algoritma C4.5 dengan PSO pada seleksi calon karyawan. Pengelompokkan ini bertujuan untuk melihat calon karyawan yang lolos dan tidak lolos seleksi berdasarkan lulusan, sertifikasi, kemampuan dan pengalaman. Dengan C4.5 dan PSO calon karyawan dibagi kedalam 2 kelompok yaitu kelompok lolos dan tidak lolos seleksi.

Untuk pengujian hasil prediksi digunakan *x-validation* dalam 10 langkah (*10 folds cross-validation*). Dengan *x-validation*, pengukuran hasil dapat lebih akurat karena data yang ada dibagi kedalam 10 data dengan jumlah yang sama, kemudian satu persatu diambil untuk test, dan 9 bagian lainnya digunakan untuk *training* (Witten, Frank, & Hall, 2011). Dengan *x-validation* akurasi dari hasil pengukuran data akan lebih terjamin karena mengurangi kemungkinan data yang tidak konsisten dalam tahap prediksi.

Untuk mengukur peningkatan akurasi dari hasil masing masing validasi, digunakan *confusion matrix*. *Confusion matrix* merupakan matrik 2 dimensi yang menggambarkan perbandingan antara hasil prediksi dengan kenyataan. Hasil yang baik dilihat dari nilai diagonal dari kiri atas ke kanan bawah yang tinggi dan diagonal dari kanan atas ke kiri bawah yang rendah. Sedangkan *ROC curve* akan digunakan untuk mengukur *AUC (Area Under Curve)*. *ROC curve* membagi hasil positif dalam sumbu y dan hasil negative dalam sumbu x (Witten, Frank, & Hall, 2011). Sehingga semakin besar area yang berada dibawah kurva, semakin baik pula hasil prediksi. Hasil prediksi C4.5 dan PSO akan dibandingkan dengan hasil C4.5 tanpa menggunakan PSO. Akurasi hasil pengolahan antara C4.5 dan PSO dengan C4.5 akan dibandingkan. Kerangka pemikiran terlihat pada gambar 2.6



Gambar 2.6 Kerangka Pemikiran

3. METODE PENELITIAN

Data pada penelitian ini dikumpulkan berdasarkan data calon karyawan perusahaan. Data untuk analisa diambil berdasarkan calon karyawan yang lolos menjadi karyawan dan tidak lolos sebagai populasi awal. Penelitian bertujuan untuk memprediksi penerimaan karyawan .

Data dikumpulkan dengan atribut berupa nama, usia, status, pendidikan_terakhir, jurusan, sertifikasi, pengalaman, dan kemampuan serta label yang menyatakan bahwa calon karyawan tersebut lolos atau tidak.

Metode penelitian pada penelitian ini adalah penelitian eksperimen dengan tahapan penelitian menggunakan pemodelan standar untuk *data mining* yang sesuai sebagai strategi pemecahan masalah secara umum dari unit penelitian yaitu *Cross-Industry Standard Process for Data Mining (CRISP-DM)* (Larose, 2005) adapun tahapan penelitian sebagai berikut :

1. Tahapan pemahaman penelitian
 Pada bagian ini dijelaskan tentang latar belakang penelitian, masalah penelitian, batasan masalah, dan tujuan dan manfaat penelitian.
2. Tahapan pemahaman data

Pada bagian ini dijelaskan tentang bagaimana dan darimana data dikumpulkan.

3. Tahapan pengolahan data

Bagian ini menjelaskan tentang dataset akhir yang digunakan untuk analisis dan memilih kasus dan variabel yang akan dianalisis.

4. Tahapan pemodelan

Pada bagian ini menjelaskan tentang pemilihan dan penerapan pemodelan yang sesuai, dan mengkalibrasi pengaturan model untuk mengoptimalkan hasil.

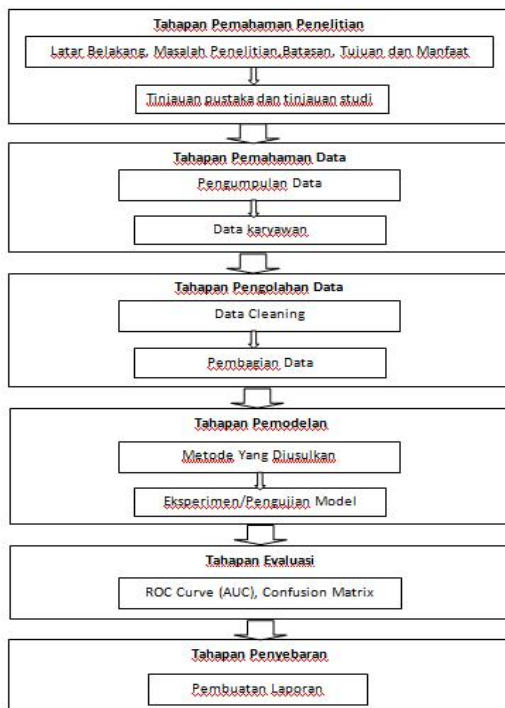
5. Tahapan evaluasi

Tahapan evaluasi yaitu menggambarkan tentang evaluasi terhadap model yang digunakan, seberapa keakuratan dengan penggunaan model tersebut.

6. Tahapan penyebaran

Tahap ini adalah tahap terakhir pada tahapan penelitian yaitu pembuatan laporan dari tahapan pemahaman penelitian sampai tahapan evaluasi.

Tahapan-tahapan penelitian menurut *Cross-Industry Standard Process for Data Mining (CRISP-DM)* data dilihat pada gambar 3.1 :

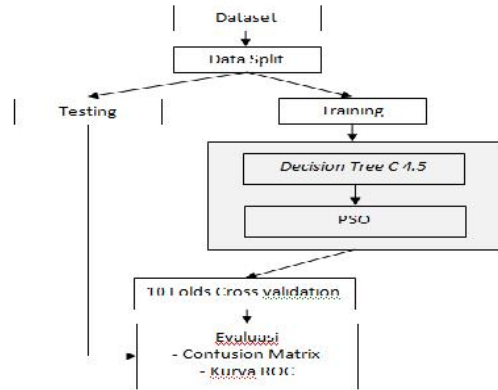


Gambar 3.1 Tahapan Penelitian

METODE YANG DIUSULKAN

Metode yang akan diusulkan pada penelitian ini adalah dengan menerapkan Algoritma C 4.5 berbasis *Particle Swarm Optimization (PSO)* untuk

klasifikasi calon karyawan. Atribut yang digunakan dalam klasifikasi antara lain umur, status, pendidikan terakhir, jurusan, sertifikasi, pengalaman, dan kemampuan dapat dilihat pada gambar 3.2 :



Gambar 3.2 Model yang Diusulkan

4. HASIL PENELITIAN DAN PEMBAHASAN

Pada tahap ini dilakukan eksperimen dan pengujian model yaitu menghitung dan mendapatkan *rule-rule* yang ada pada model algoritma yang diusulkan. Setelah itu diuji rule yang telah didapatkan ke dalam model *cross validation* untuk mendapatkan hasil yang lebih baik.

Jumlah awal data yang diperoleh dari pengumpulan data yaitu sebanyak 358 data. Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan sebagai berikut :

1. Data Cleaning

Data cleaning bekerja untuk membersihkan nilai yang kosong tidak konsisten atau mungkin *tupel* yang kosong (*missing values* dan *noisy*)

2. Split Data

Split Data atau pembagian data dalam penelitian ini akan dibagi menjadi dua yaitu data *training* dan data *testing*. Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 358 record dan 9 atribut. Pembagian data sendiri dibagi menjadi 70 % untuk data *training* dan data *testing* 30 % dari *dataset* yang telah dikumpulkan.

Untuk menetapkan model yang baik dalam penelitian, ada beberapa indikator yang akan disesuaikan untuk mencapai model yang maksimal. Untuk pembuatan pohon keputusan, indikator *minimal gain* dan *pruning* disesuaikan agar mendapat model dengan akurasi yang tinggi. Pengukuran akurasi dan nilai model yang terbentuk aka didasarkan dengan menggunakan *confusion matrix*

dan ROC Curve. Tabel indikator dan hasil pengujian untuk pohon keputusan dapat dilihat pada tabel 4.1.

Tabel 4.2 Indikator pohon keputusan

Minimal Gain	No pruning		Pruning	
	ACC	AUC	ACC	AUC
0,01	80,80 %	0,869	80,80 %	0,867
0,02	80,40 %	0,848	80,40 %	0,867
0,03	82,00 %	0,842	82,00 %	0,862
0,04	82,00 %	0,842	82,00 %	0,862
0,05	80,00 %	0,835	80,00 %	0,855
0,06	79,60 %	0,841	79,60 %	0,860
0,07	80,80 %	0,854	80,80 %	0,869
0,08	80,00 %	0,850	80,00 %	0,869
0,09	80,40 %	0,860	80,40 %	0,858
0,1	80,80 %	0,878	80,80 %	0,880
0,15	78,00 %	0,783	78,00 %	0,783
0,2	78,40 %	0,776	78,40 %	0,776

Setelah data didapatkan dari nilai *gain* tertinggi berdasarkan akurasi AUC, model akan dioptimasi menggunakan PSO. Indikator PSO juga akan disesuaikan agar dapat memberikan peningkatan yang besar.

Berikut adalah tabel indikator hasil pengujian pada nilai *accuracy* dan AUC tertinggi yang dikombinasi dengan indikator PSO.

Tabel 4.2 hasil pengujian nilai *gain* pada PSO

Minimal Gain	Population Size	ACC	AUC
0,03	5	81,60 %	0,836
0,03	10	82,80 %	0,872
0,03	15	80,80 %	0,875
0,03	20	82,00 %	0,882
0,03	25	85,20 %	0,891
0,03	30	80,40 %	0,850
0,04	5	80,40 %	0,840
0,04	10	82,40 %	0,886
0,04	15	80,80 %	0,875
0,04	20	81,20 %	0,880
0,04	25	83,60 %	0,893

Minimal Gain	Population Size	ACC	AUC
0,04	30	79,60 %	0,864
0,07	5	76,40 %	0,880
0,07	10	78,80 %	0,869
0,07	15	78,00 %	0,879
0,07	20	78,40 %	0,880
0,07	25	79,60 %	0,884
0,07	30	75,20 %	0,858
0,1	5	78,80 %	0,876
0,1	10	80,40 %	0,877
0,1	15	78,80 %	0,854
0,1	20	78,00 %	0,877
0,1	25	80,80 %	0,872
0,1	30	78,00 %	0,870

Hasil dari pengujian model yang telah dilakukan adalah untuk mengukur tingkat akurasi dan AUC (*Area Under Curve*) pada algoritma C 4.5.

accuracy: 80.80% ± 0.67% (mikro: 80.80%)			
	True Label	True Class	Class Precision
pred_label	10	26	77.26%
pred_class	19	11	54.55%
class_recall	84.11%	71.52%	

Gambar 4.1 Confusion Matrix C 4.5

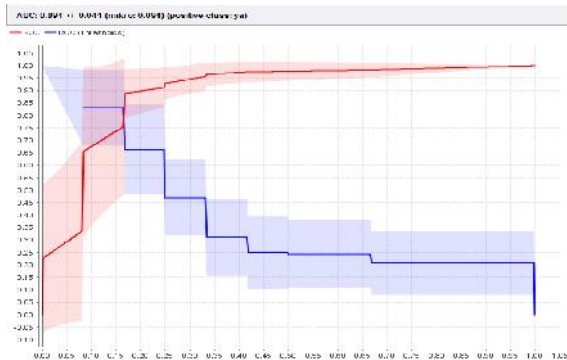


Gambar 4.2 Nilai AUC pada ROC Curve Algoritma C 4.5

Hasil dari pengujian model yang telah dilakukan adalah untuk mengukur tingkat akurasi dan AUC (*Area Under Curve*) pada algoritma C 4.5 dengan nilai *minimal gain* 0,03.

accuracy: 85.20% ± 1.40% (mikro: 85.20%)			
	True Label	True Class	Class Precision
pred_label	104	21	83.20%
pred_class	16	109	87.70%
class_recall	86.67%	80.85%	

Gambar 4.3 Confusion Matrix C 4.5 berbasis PSO



Gambar 4.4 Nilai AUC pada ROC Curve Algoritma C 4.5 berbasis PSO

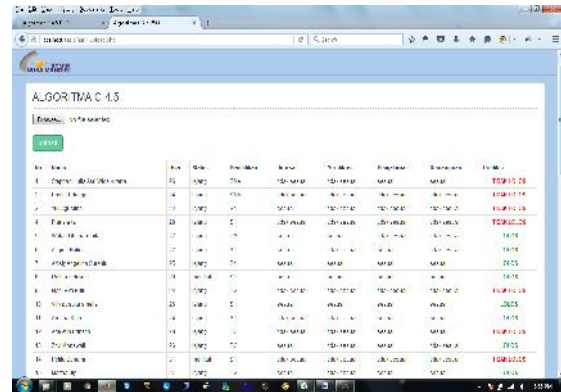
Dari hasil pengujian diatas baik evaluasi menggunakan *confusion matrix* maupun *ROC Curve* terbukti bahwa hasil pengujian algoritma C 4.5 berbasis PSO memiliki nilai akurasi lebih tinggi dibandingkan dengan algoritma C 4.5. nilai akurasi untuk model C 4.5 sebesar 80,80% dan nilai akurasi untuk model C 4.5 berbasis PSO sebesar 85,20% dengan selisih akurasi antara model C 4.5 dengan model C 4.5 berbasis PSO sebesar 4.4, berikut table perbandingan antara model C 4.5 dan C 4.5 berbasis PSO.

Tabel 4.3 Pengujian algoritma C 4.5 dan C 4.5 berbasis PSO

	Accuracy	AUC
C 4.5	80,80 %	0,878
C 4.5 PSO	85,20 %	0,891

Graphical User Interface

Dari hasil eksperimen menggunakan model algoritmaa C 4.5 dan model algoritma C 4.5 berbasis *particle swarm optimization* (PSO) maka akan dihasilkan *rule*, dimana *rule* tersebut akan diimplementasikan kedalam *graphical user interface* (GUI). Berikut adalah *graphical user interface* seleksi karyawan dari rule-rule yang dihasilkan dari tahap eksperimen :



Gambar 4.5 Graphical User Interface dari rule yang dihasilkan

5. KESIMPULAN

Dari hasil eksperimen penerapan *particle swarm optimization* pada model algoritma C 4.5 memberikan nilai akurasi yang lebih tinggi yaitu 85,20 % dibandingkan dengan algoritma C 4.5 80,80 %. dari hasil tersebut didapatkan selisih antara kedua model tersebut adalah 4,4 %. Sementara untuk evaluasi menggunakan *ROC Curve* untuk kedua model yaitu, untuk algoritma C 4.5 nilai AUCnya adalah 0,878 dengan tingkat diagnosa *Good Classification* dan untuk model algoritma C 4.5 berbasis *particle swarm optimization* nilai AUCnya adalah 0,891 dengan tingkat diagnosa *Good Classification*. Dari hasil evaluasi *ROC Curve* tersebut terlihat bahwa model algoritma C 4.5 berbasis *particle swarm optimization* lebih tinggi jika dibandingkan dengan algoritma C 4.5. dari hasil tersebut didapatkan selisih antara kedua model tersebut adalah 0,13. sehingga dapat disimpulkan bahwa algoritma C 4.5 berbasis *particle swarm optimization* lebih akurat untuk memprediksi penerimaan karyawan.

Saran

Penerapan optimasi menggunakan *particle swarm optimization* terbukti dapat meningkatkan akurasi dari algoritma C 4.5. namun ada beberapa faktor yang dapat dicoba untuk penelitian selanjutnya, agar algoritma C 4.5 dapat menghasilkan model yang lebih baik lagi, berikut adalah saran-saran yang diusulkan :

1. Atribut seperti nama calon karyawan tidak terpengaruh terhadap bobot atribut, sebaiknya digunakan penyeleksian atribut untuk menyeleksi atribut yang digunakan.

2. Penelitian ini diharapkan dapat digunakan oleh bagian HRD atau SDM suatu institusi sebagai bahan pertimbangan dalam proses perekrutan karyawan.
3. Penelitian ini dapat dioptimasi dengan algoritma optimasi lainnya, seperti *Ant Colony Optimization* (ACO) atau *Geneted Agregation* (GA).
4. Metode ini dapat dikembangkan dengan metode klasifikasi data mining yang lainnya, seperti Naïve Bayes, *Support Vector Machine* untuk melakukan perbandingan.
5. Memperbanyak jumlah atribut dan memanfaatkan metode penyeleksian.

6. DAFTAR PUSTAKA

- Aprilla, D., Baskoro, Donny Aji, Ambarwati, Lia, & Wicaksana, I Wayan Simri. (2013). *Belajar Data Mining dengan Rapid Miner*. Jakarta
- Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *A Guide for Students in Computer Science and Information Systems*. London: Springer.
- Chawla, N.,V. (2003). *C4.5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure*. In: *ICML workshop on learning from imbalanced data sets II*. Washington, DC, USA
- Dawson, C. W. (2009). *Projects in Computing and Information Systems a student's guide* (Second Edition ed.). Harlow, UK: Addison-Wesley.
- Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*. Springer-Verlag.
- Handoko, T. Hani,. (1996). *Manajemen Perencanaan dan Sumber Daya Manusia*. Yogyakarta : PT. BPFE.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (Second Edition ed.). San Francisco: Elsevier Inc.
- Hasibuan , Malayu S.P. (2001). *Manajemen Sumber Daya Manusia*. Jakarta: Bumi Aksara.
- Hasibuan , Malayu S.P. (2002). *Manajemen Sumber Daya Manusia*. Jakarta: Bumi Aksara.
- Hermawati, Fajar Astuti (2009). *Data Mining*. Yogyakarta:Andi.
- Kamus Besar Bahasa Indonesia (2008) Jakarta.
- Kennedy, J., & Eberhart, R. (1995, November-December). *Particle Swarm Optimization. Proceedings of IEEE International Conference on Neural Networks , 1942-1948*.
- Larose, D. T. (2005). *Discovering Knowledge in Data*. Canada: Wiley Interscience.
- Manullang, M., & Manullang, Marihot (2004). *Manajemen Personalia*. (Third Edition ed.)Yogyakarta: Gadjah Mada University Press.
- Santosa, Budi, & Willy, Paul. (2011). *Metoda Metaheuristik Konsep dan Implementasi*. Surabaya: Guna Widya.
- Santosa, Budi. (2007) *Data Mining: Teknik Pemanfaatan Dataa Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- UU RI No, 13 tahun 2013
- Witten, H. I., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Mechine Learning Tools And Technique*. Burlington: Elsevier Inc.
- Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. Taylor & Francis Group, LLC.