

# Optimasi Prediksi Keputusan Calon Nasabah Potensial Berbasis Particle Swarm Optimization

Saruni Dwiasnati<sup>1</sup>, Yudo Devianto<sup>2</sup>

<sup>1,2</sup>Fakultas Ilmu Komputer, Universitas Mercubuana Jakarta

<sup>1</sup>e-mail: saruni.dwiasnati@mercubuana.ac.id

<sup>2</sup>e-mail: yudo.devianto@mercubuana.ac.id

## Abstrak

Data mining merupakan proses pencarian data secara otomatis pada database berukuran besar yang menghasilkan informasi. Dalam era globalisasi, persaingan bisnis menjadi sangat tajam dan berwarna, baik dipasar nasional maupun di pasar internasional. Meningkatnya permintaan pasar mengakibatkan daya saing dan jumlah pesaing pelaku bisnis juga meningkat, setiap perusahaan di tuntut untuk selalu memperhatikan kebutuhan dan keinginan konsumen serta berusaha memenuhi apa yang mereka harapkan dengan cara yang lebih memiliki nilai jual atau variatif dari pada yang dilakukan oleh para pesaing lainnya. Salah satu teknik klasifikasi yang digunakan pada penulisan ini yaitu decision tree khususnya algoritma C4.5. Pada penelitian sebelumnya yang dilakukan oleh penulis bahwa algoritma C4.5 memiliki hasil accuracy yang kurang maksimal karena algoritma C4.5 merupakan algoritma yang lemah, maka di perlukanlah peningkatan accuracy. Penulisan ini menggunakan tools Rapidminer versi 8 sebagai media untuk menguji data yang akan diolah untuk mendapatkan hasil accuracy dari variabel yang digunakan. Penelitian ini memperoleh hasil dari accuracy yang lebih akurat pada prediksi Keputusan yang tanpa menggunakan algoritma optimasi.

Kata kunci : Calon Nasabah Asuransi, Algoritma C4.5, Particel Swarm Optimization.

## Abstract

*Data mining is the process of automatically searching data on a large database that generates information. In the era of globalization, business competition has become very sharp and colorful, both in the national market and in the international market. Increasing market demand results in competitiveness and the number of business competitors is also increasing, each company is required to always pay attention to the needs and desires of consumers and try to fulfill what they expect in a way that has a more selling or varied value than that done by other competitors. One of the classification techniques used in this paper is decision tree, especially the C4.5 algorithm. In previous studies conducted by the author that the C4.5 algorithm has a less than optimal accuracy because the C4.5 algorithm is a weak algorithm, it is necessary to increase accuracy. This writing uses the Rapidminer version 8 tool as a medium to test the data to be processed to get the accuracy of the variables used. This study obtained results from a more accurate accuracy on the prediction of decisions without using an optimization algorithm.*

*Keywords: Prospective Insurance Customers, Algoritma C4.5, Particel Swarm Optimization.*

## 1. Pendahuluan

Nasabah merupakan asset terpenting bagi setiap pelaku bisnis, yang menjadikan nasabah menjadi sebuah asset adalah dapat meningkatnya keuntungan perusahaan jika seorang nasabah loyal dalam melakukan transaksi atau melakukan pembelian terhadap sebuah produk yang di tawarkan oleh perusahaan. Keuntungan besar yang di peroleh oleh sebuah perusahaan jika dimana nasabah nya loyal adalah peningkatan laba

perusahaan dan perusahaan tersebut semakin di percaya oleh semua lapisan masyarakat, keuntungan yang di peroleh oleh nasabah jika menjadi nasabah loyal akan selalu di berikan prioritas untuk setiap semua transaksi yang akan dilakukan oleh nasabah tersebut dan akan selalu diberikan gift dari perusahaan tersebut.

Ketersediaan data yang banyak dalam sebuah bidang dan kebutuhan akan informasi atau pengetahuan dalam bidang

tertentu dapat menjadi acuan sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis dan dukungan infrastruktur di segala bidang merupakan cikal-bakal dari lahirnya teknologi *data mining*. Penggunaan teknik *data mining* diharapkan dapat membantu percepatan proses pengambilan keputusan, memungkinkan perusahaan untuk mengelola informasi yang terkandung didalam data nasabah dan menjadi sebuah pengetahuan (*knowledge*) yang baru untuk menentukan keputusan yang dibutuhkan.

Banyak perusahaan yang ingin menawarkan sebuah produk kepada nasabah nya namun tidak memiliki kepastian apakah nasabah tersebut masih aktif bertransaksi atau ada perubahan data pribadi dari nasabah tersebut yang memudahkan seorang agent untuk menghubungi nasabah tersebut. Seorang nasabah memiliki tingkatan variabel yang menandakan dirinya, seperti tingkat umur yang berbeda akan berbeda pula kebutuhan dari nasabah tersebut, jenis kelamin juga dapat mempengaruhi perbedaan kebutuhan dari sebuah produk, pekerjaan, wilayah tempat tinggal nasabah juga akan berbeda produk apa yang dibutuhkan. Hal ini yang menimbulkan kesulitan seorang agent untuk menawarkan produk apa yang cocok atau tidak pada seorang nasabah. Nasabah yang berpotensi memiliki beberapa kriteria untuk dapat disebut dengan nasabah berpotensi.

Setiap metode yang dapat dilakukan menggunakan data mining dapat dijelaskan dari beberapa aspek yang dapat dihasilkan.. Data mining merupakan suatu langkah dalam melakukan Knowledge Discovery in Databases (KDD). Knowledge discovery sebagai suatu proses terdiri atas pembersihan data (*data cleaning*), integrasi data (*data integration*), pemilihan data (*data selection*), transformasi data (*data transformation*), data mining, evaluasi pola (*pattern evaluation*) dan penyajian pengetahuan (*knowledge presentation*). Data mining mengacu pada proses untuk menambang (*mining*) pengetahuan dari sekumpulan data yang sangat besar untuk menghasilkan sebuah pengetahuan baru dari bidang tertentu (Firdaus, 2017).

Pada Data Mining terdapat 3 bagian yaitu: Association, Klasifikasi dan Clustering. Association memiliki definisi sebuah proses yang digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut

dari sekumpulan data yang dimiliki, sedangkan Klasifikasi adalah teknik yang dilakukan untuk memprediksi class atau properti dari setiap instance data, dan Clustering sendiri memiliki makna pengelompokan data tanpa berdasarkan kelas data tertentu ke dalam kelas objek yang sama sesuai dengan topik yang diangkat.

Prediksi memiliki kemiripan definisi dengan klasifikasi, akan tetapi data yang digunakan pada klasifikasi berdasarkan perilaku atau nilai yang diperkirakan pada masa yang akan datang yang dilihat dari pola/nilai pada masa lalu. Contoh dari tugas prediksi misalnya untuk memprediksikan adanya pengurangan jumlah pelanggan dalam beberapa waktu yang lalu dan prediksi harga saham dalam empat bulan ke depan.

Salah satu metode utama data mining adalah Supervised Learning (Pembelajaran dengan Guru), algoritma ini melakukan proses belajar berdasarkan nilai dari indikator target yang terhubung dengan indikator prediktor. Sedangkan metode analisis yang terdapat pada data mining salah satu nya adalah Klasifikasi. Klasifikasi sendiri memiliki definisi sebuah proses yang digunakan untuk menemukan model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas yang bertujuan untuk menemukan pola yang lebih memiliki nilai dari data yang berukuran relatif besar hingga sangat besar. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefinisikan sebelumnya. Dalam klasifikasi hanya terdapat satu atribut dari sekian banyaknya atribut yang bisa menjadi kemungkinan yang disebut atribut target, sedangkan atribut lainnya yang terdapat disebut atribut prediktor. Tiap kemungkinan nilai yang dimiliki oleh atribut target menunjukkan class yang diprediksi berdasarkan nilai-nilai dari atribut prediktor.

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang termasuk salah satu algoritma yang masuk dalam bagian klasifikasi yang dapat digunakan untuk melakukan klasifikasi atau pengelompokan data yang bersifat prediktif untuk membentuk pohon keputusan (*Decision Tree*). Pohon keputusan merupakan metode klasifikasi dan prediksi yang cukup familiar di kalangan penggiat Data Mining. Pohon keputusan (*Decision*

Tree) berguna untuk mengeksplorasi data, menemukan hubungan yang tidak saling terlihat antara sebuah calon variabel input dengan sebuah variabel target lainnya. Proses pada decision tree yaitu mengubah bentuk data dalam bentuk tabel menjadi model pohon(tree), mengubah model pohon menjadi rule yang dapat disederhanakan.

Saat mencari hasil dari alur sebuah decision tree langkah awal yang harus dilakukan adalah tentukan dahulu atribut mana yang akan menjadi simpul parents dan atribut mana yang akan menjadi simpul child. Pemilihan atribut yang baik adalah atribut yang memungkinkan untuk mendapatkan decision tree yang paling kecil ukurannya dan unik atau atribut yang bisa memisahkan objek menurut kelasnya masing-masing menjadi lebih detail. Untuk dapat menghasilkan sebuah pohon keputusan, maka nilai entropinya harus di ketahui dahulu hasilnya dengan menggunakan persamaan:

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

Keterangan:

S : himpunan kasus

A : atribut

N : jumlah partisi S

Pi : proporsi dari Si terhadap

Setelah langkah pertama selesai dan mendapat hasil himpunan kasusnya, maka selanjutnya bagaimana mencari nilai information gain. Information Gain merupakan sebuah kriteria yang paling populer untuk pemilihan atribut yang ingin di cari. Information gain adalah salah satu attribute selection measure yang digunakan untuk memilih *test* attribute tiap node pada tree. Atribut dengan informasi gain tertinggi dipilih sebagai testatribut dari suatu node [Sunjana, 2010]. Dalam prosesnya perhitungan gain bisa terjadi atau tidak suatu missing value. Information gain dapat dihitung dari output data atau variabel dependen y yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (y,A). Gain (y,A) dari atribut A relatif terhadap output data y adalah:

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

Keterangan:

S : himpunan kasus

A : atribut

N : jumlah partisi atribut A

[Si] : jumlah kasus pada partisi ke-i

[S] : jumlah kasus pada S

Particle Swarm Optimization pertama kali dicetuskan oleh Kennedy dan Eberhart pada tahun 1995 yang membahas mengenai perilaku kerumunan makhluk hidup. PSO terdiri dari sekumpulan partikel yang mencari posisi terbaik, yang merupakan posisi terbaik untuk masalah optimasi dalam ruang fitur. Penerapan PSO sebagai penentu parameter regulasi akan memberikan pengaruh pada accuracy klasifikasi yang telah ada [Widiasri, Justitia, & Arifin, 2011]. Persamaan untuk menghitung perpindahan posisi dan kecepatan partikel yaitu:

$$V_i(t) = V_i(t-1) + c_1 r_1 [X_{pbest\ i} - X_i(t)] + c_2 r_2 [X_{Gbest} - X_i(t)]$$

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

Keterangan :

Vi (t) : kecepatan partikel i saat iterasi t

Xi (t) : posisi partikel i saat iterasi t

c1 dan c2 : learning rate untuk kemampuan individu (cognitive) dan pengaruh hasil.

R1 dan r2 : bilangan random yang berdistribusi uniformal dalam interval 0 dan 1

X<sub>pbest i</sub> : posisi terbaik partikel i

X<sub>Gbest</sub> : posisi terbaik global

Data yang akan digunakan pada penulisan ini adalah 150 dataset. Penulisan ini dilakukan atas kemajuan dari penelitian sebelumnya yang membandingkan 2 algoritma yaitu C4.5 dan Naïve Bayes, setelah salah satu algoritma tersebut di dapatkan hasil accuracy tertingginya yaitu C4.5 maka hasil accuracy di optimasikan agar mendapatkan nilai accuracy yang jauh lebih baik. Hasil yang didapatkan dari algoritma C4.5 yang belum di optimasikan sebesar 76,21%, sedangkan menggunakan algoritma C4.5 yang sudah dioptimasikan menggunakan Particle Swarm Optimization sebesar 82,92% masih mungkin ditingkatkan lagi jika menggunakan algoritma optimasi lainnya, seperti WMA.

## 2. Metode Penelitian

Penelitian ini akan mengusulkan hasil accuracy baru untuk sebuah permasalahan memprediksi keputusan Calon Nasabah Potensial pada sebuah perusahaan asuransi

Swasta ternama. Data yang digunakan dalam penelitian ini di ambil pada tahun 2018 bulan Januari-Desember, data yang di ambil kisaran 100.000 an namun untuk data testing untuk penelitian ini penulis hanya mengambil sampel data sebanyak 150. Pada tahap Klasifikasi akan dilakukan validasi silang (Cross Validation) dengan 10 kali validasi untuk mendapatkan hasil accuracy yang jauh lebih optimal dibandingkan dengan penelitian sebelumnya yang hanya menggunakan algoritma saja bukan algoritma yang optimasi.

Tahapan metodologi dalam penelitian ini dijelaskan pada gambar di bawah ini:



Gambar 1. Tahapan Metodologi Penelitian

Keterangan:

1. Identifikasi Masalah  
Berisikan pernyataan yang bersifat umum terhadap permasalahan yang diamati.
2. Perumusan Masalah  
Bagaimana sebuah masalah dan cari beberapa contoh pertanyaan yang memerlukan jawaban secara ilmiah.
3. Penelusuran Pustaka  
Mencari semua teori atau model untuk mencari solusi permasalahan yang biasanya didapat dari buku teks ilmiah, tesis, karangan ilmiah, dsb.
4. Rancangan Penelitian  
Rancangan penelitian berisikan pengetahuan, algoritma, metode,

- produk(sistem) atau model untuk menunjang penelitian.
5. Pengumpulan Data  
Setiap penelitian harus ada data kuantitatif, data kualitatif serta data kombinasi.
6. Pengolahan atau Analisa Data  
Analisis data dipilih sesuai dengan data yang kita dapatkan. Bila sampelnya representatif bisa digunakan uji-uji tertentu.
7. Penyimpulan Hasil  
Dalam menulis laporan penelitian atau laporan akhir, kita harus berani mengemukakan dan menuliskan apa yang kita dapatkan selama melakukan penelitian tersebut (objektif).

Model desain ini akan melakukan pemrosesan data training dan data testing untuk menguji metode algoritma yang digunakan. Tahapan yang akan dilalui dibagi menjadi 3 bagian, yaitu preprocessing, seleksi fitur (Feature Selection) dan validation yang didalamnya berisi sub proses training dan testing.

**a. Pengumpulan Data**

Data yang digunakan pada penelitian kali ini merupakan data sekunder, karena sumber data diperoleh melalui media perantara atau secara tidak langsung yang berupa buku, catatan, bukti yang telah tervalidasi, atau arsip baik yang dipublikasikan maupun yang tidak dipublikasikan secara umum yaitu 100 ribu data yang akan dipecah menjadi 80 persen data training dan 20 persen data testing. Data testing yang digunakan sekitar 150 data yang telah diklasifikasikan berdasarkan variable yang ada. Masalah yang harus dipecahkan pada penelitian kali ini adalah, bagaimana menghasilkan nilai accuracy yang jauh lebih optimal untuk mengklasifikasikan prediksi Calon Nasabah Potensial untuk ditawarkan produk lainnya. Dataset yang digunakan pada penelitian ini ada pada Gambar di bawah ini:

id	name	lastname	Sex	Age	Marital	tanggungan	Occupation	Wilyayah	Keterangan	Waktu	Target
2	MIX_SUPER_NYENNI	HARDIANTO	P	46	kawin	1	Pegawai Swasta	Barat	Siang	Potensial	
3	MIX_SUPER_NLUMKAN	HAKIM	L	47	kawin	3	Pegawai Swasta	Barat	Malam	Potensial	
4	MIX_SUPER_NLAMBARIAL		L	47	kawin	1	Wiraswasta	Barat	Sore	Potensial	
5	MIX_SUPER_NWAGINO		L	46	kawin	2	Polisi	Barat	Sore	Potensial	
6	MIX_SUPER_NWINARTO		L	62	belum kawin	0	Wiraswasta	Barat	Siang	Potensial	
7	MIX_SUPER_NSHERI		L	59	kawin	0	Pegawai Swasta	Barat	Pagi	Potensial	
8	MIX_SUPER_NANDEMIAM	SUHARIYANTO	L	57	kawin	1	Wiraswasta	Barat	Sore	Potensial	
9	MIX_SUPER_NANDRI	PERMA	L	62	kawin	1	Pegawai Swasta	Barat	Pagi	Potensial	
10	MIX_SUPER_NEFFY	ZULFIYUE	P	47	kawin	1	Wiraswasta	Barat	Sore	Potensial	
11	MIX_SUPER_NNURBANNAH		P	57	kawin	3	Pegawai Swasta	Barat	Sore	Potensial	
12	MIX_SUPER_NENDANG	SRI SOESLOWATI	P	44	kawin	0	Pegawai Swasta	Barat	Siang	Tidak Potensial	
13	MIX_SUPER_NHARY	AGUS WIBOWO	L	38	belum kawin	0	Pegawai Swasta	Barat	Sore	Tidak Potensial	
14	MIX_SUPER_NSOEPRAHADI		L	40	belum kawin	0	Wiraswasta	Barat	Sore	Tidak Potensial	
15	MIX_SUPER_NHAMONANGAN	SIRAIT	L	45	kawin	2	Wiraswasta	Barat	Siang	Tidak Potensial	
16	MIX_SUPER_NCHEPL	ASMARADEWI	L	38	kawin	2	Pegawai Swasta	Barat	Sore	Tidak Potensial	
17	MIX_SUPER_NHARUN	RASYID	L	40	belum kawin	0	Pendidik	Barat	Siang	Tidak Potensial	
18	MIX_SUPER_NHADI	SUNAWAR	L	47	belum kawin	0	Pegawai Swasta	Barat	Sore	Potensial	
19	MIX_SUPER_NZAINAL	ARIFIN S	L	47	belum kawin	0	Pegawai Swasta	Barat	Sore	Potensial	

Gambar 2. Dataset

**b. Pengolahan Data**

Dataset ini dalam tahap preprocessing harus melalui 3 proses, yaitu:

Tahap ini akan melibatkan sub proses antara lain:

- a) Membuang duplikasi data
- b) Memeriksa data yang inkonsisten
- c) Memperbaiki kesalahan pada data.

Dan Hasil akhir dari data ini berupa kumpulan data yang sudah bersih atau tidak ada missing valuenya.

**c. Evaluasi dan Validasi Data**

Validasi dilakukan menggunakan 10 fold cross validation. Dimana dengan menggunakan teknik ini dengan membagi secara acak ke dalam tiap bagian dimana akan dilakukan proses klasifikasi terlebih dahulu, Sedangkan pengukuran akurasi diukur dengan confusion matrix dan kurva ROC (Receiver Operating Characteristics) untuk mengukur nilai AUC.

AUC digunakan untuk mengukur kinerja diskriminatif dengan memperkirakan probabilitas output yang sudah di dapatkan hasilnya dari sampel yang dipilih secara acak dari populasi positif atau negatif, semakin besar nilai AUC, semakin kuat klasifikasi yang dihasilkan. Karena AUC merupakan bagian dari daerah unit persegi, nilainya yang dihasilkan akan selalu sama yang dihasilkannya, antara 0,0 dan 1,0.

Tabel 1. Confusion Matrix

Classification	Predicted Class	
	Class= Yes	Class=No
Observed Class	Class = Yes	a (True Positive - TP) b (False Negative - FN)
	Class = No	c (False Positive - FP) d (True Negative -TFN)

Sumber : (Gorunescy, 2011)

**3. Hasil dan Pembahasan**

Data yang digunakan dalam pengklasifikasian Calon Nasabah Potensial terdiri dari 100.000 dataset,150 data yang digunakan untuk data testing berdasarkan variabel yang tersedia. Dalam penentuan rekomendasi Calon Nasabah Potensial menggunakan atribut yang terdiri dari 5 atribut dan 1 atribut yang digunakan untuk class keterangan. Atribut yang digunakan yaitu Sex, Marital, Age, Wilayah,

Occupation dan 1 atribut keterangan yaitu Keterangan.

Klasifikasi dilakukan dengan menggunakan Software Rapidminer dengan

Occupation				
Pegawai Negeri	9	5	0.0676	0.0375
Pegawai Swasta	40	37	0.3007	0.2781
Guru/Dosen	5	8	0.03759	0.0601
Wiraswasta	17	2	0.1278	0.015
Konsultan	2	0	0.015	0
Profesional	4	2	0.3007	0.015
Militer	0	2	0	0.015
Keterangan				
Pagi	24	2	0.1804	0.015
Siang	21	13	0.1578	0.0977
Sore	31	41	0.233	0.3082
Malam	1	0	0.00751	0

rule berbeda sesuai dengan model yang digunakan (Jungermann, 2009).

**3.1. Algoritma C4.5**

Algoritma C4.5 merupakan sebuah model yang dihasilkan dengan membentuk pohon keputusan. Untuk membuat pohon keputusan, ada beberapa langkah-langkah yang harus dilalui terlebih dahulu, yaitu:

1. Tentukan akar dari pohon keputusan. Akar tersebut diambil dari atribut terpilih dengan cara menghitung nilai entropy masing-masing atribut dengan persamaan sebagai berikut:

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

Tabel 2. Jumlah Kasus

	Jumlah Data	Buy	No	Entropy	Gain
	(S)	(S1)	(S2)		
Total	21	10	11	0.998	
Sex					0.1461070
L	12	4	8	0.918	
P	9	7	2	0.764	
Marital					0.1606079
Kawin	16	10	6	0.954	
tidak kawin	5	1	5	0.464	
age					1.0169522
produktif 1	0	0	0	0.000	
produktif 2	10	5	5	1.000	
produktif 3	3	1	2	0.918	
produktif 4	8	5	3	0.954	
Wilayah					0.4819546
Barat	13	7	6	0.996	
Tengah	3	2	1	0.918	
Timur	5	2	3	0.971	
Occupation	Jumlah Data	Buy	No	Entropy	Gain
Pegawai negeri	21	11	10	0.998	0.0000000
Pegawai swasta	0	0	0	0.000	
Guru/dosen	0	0	0	0.000	
Militer	0	0	0	0	
Keterangan	Jumlah Data	Buy	No	Entropy	Gain
Pagi	5	2	3	0.971	1.0402963
siang	6	5	1	0.650	
sore	8	3	5	0.954	
Malam	2	1	1	1.000	

2. Setelah mendapat hasil nilai entropy dari masing-masing atribut yang ingin diketahui, selanjutnya mencari nilai gain

dengan persamaan terdapat pada Tabel 3 sebagai berikut ini:

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

Tabel 3. Perhitungan Probabilitas Prior

P	Buy	No	P(X C <sub>i</sub> )		
			Buy	No	
Total	133	77	56	0.578947	0.421053
Sex					
M	62	34	0.466165	0.2556	
F	15	22	0.1127	0.1654	
Marital					
Kawin	50	13	0.3759	0.3759	
Duda	0	2	0	0.015	
Tidak Kawin	27	41	0.203	0.3082	
age					
Produktif 1	0	0	0	0	
Produktif 2	51	19	0.3834	0.1428	
Produktif 3	4	8	0.03	0.0601	
Produktif 4	22	29	0.1654	0.218	
Wilayah					
Barat	35	37	0.2631	0.2781	
Tengah	0	0	0	0	
Timur	42	19	0.3157	0.1428	

- Ulangi langkah 1 dan 2 untuk mengisi entropy dan nilai gain sampai semua atribut terisi
- Proses partisi pohon keputusan akan berhenti saat:
  - Semua tupel dalam node N mendapat kelas yang sama
  - Tidak ada atribut di dalam tupel yang dipartisi lagi
  - Tidak ada tupel di dalam cabang yang kosong

Gambar 3 merupakan perhitungan accuracy data *training* menggunakan algoritma C4.5 yang menghasilkan accuracy 76,21%. Diketahui data *training* terdiri dari 150 record data, 92 data diklasifikasikan POTENSIAL dan 59 data diprediksi TIDAK POTENSIAL.

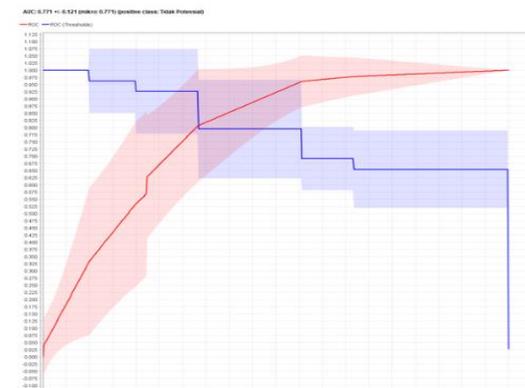
accuracy: 76.21% +/- 0.85% (mikro: 76.16%)

	true Potensial	true Tidak Potensial	class precision
pred. Potensial	82	6	91.10%
pred. Tidak Potensial	30	53	63.86%
class recall	87.39%	89.83%	

Gambar 3. Model Confusion Matrix algoritma C4.5

Sedangkan hasil yang di peroleh dari pengolahan kurva ROC dapat dilihat pada Gambar 4 menghasilkan nilai AUC sebesar

0,771 yang masuk pada kategori Fair Classification.



Gambar 4. Kurva ROC algoritma C4.5

### 3.2. Algoritma C4.5 berbasis Particle Swarm Optimization

Particle Swarm Optimization (PSO) adalah salah satu algoritma optimasi Swarm Intelligence (SI). Algoritma dari SI kebanyakan terinspirasi hewan berkelompok yang saling bekerja sama tanpa pemimpin. Selain PSO, ada algoritma SI lainnya seperti Bee Colony, Bat Algorithm, Cuckoo Search dan lainnya yang tidak di bahas di postingan ini.

Agar nilai accuracy data yang menggunakan algoritma C4.5 dapat meningkat, maka diperlukan algoritma optimasi agar dapat menghasilkan nilai accuracy yang lebih optimal, yang salah satunya dapat menggunakan algoritma PSO.

Pemodelan menggunakan algoritma optimasi PSO terdapat pada gambar 4 berikut ini

accuracy: 82.92% +/- 11.15% (mikro: 82.78%)

	true Potensial	true Tidak Potensial	class precision
pred. Potensial	82	16	83.67%
pred. Tidak Potensial	10	43	81.13%
class recall	89.13%	72.88%	

Gambar 5. Model Confusion Matrix Algoritma C4.5 berbasis PSO

### 4. Kesimpulan

Kesimpulan yang diperoleh setelah melakukan eksperimen tersebut diperoleh hasil dari accuracy yang lebih akurat pada prediksi Keputusan Calon

Nasabah Potensial menggunakan algoritma C4.5 berbasis Particle Swarm Optimization memiliki nilai accuracy sebesar 82.92%, sedangkan prediksi Keputusan Calon Nasabah Potensial yang hanya menggunakan algoritma C4.5 saja menghasilkan nilai accuracy sebesar 76,51%. Dapat dilihat dari hasil eksperimen tersebut maka nilai accuracy yang dihasilkan dari kedua algoritma tersebut Algoritma C4.5 berbasis Particle Swarm Optimization terjadi peningkatan yang cukup meningkat, meningkatkan accuracy dari kedua algoritma tersebut mencapai 6.41%.

### Referensi

- Abadi. (2016). "Penentuan Penerima Bantuan Dana untuk Sekolah Menengah Di Kab. Banjar Menggunakan Metode AHP- TOPSIS". Jurnal Speed- Sentra Penelitian Engeneering dan Edukasi, Vol. 8 No. 1.
- Darmanto. (2014). "Penerapan Metode AHP (*Analytic Hierarchy Process*) untuk Menentukan Kualitas Gula Tumbuh". Jurnal SIMETRIS, Vol. 5 No. 1, April 2014, ISSN : 2252-4983.
- Firdaus, Diky (2017). "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer". Jurnal FORMAT, Vol 6 Nomor 2, ISSN: 2089-5615
- O'Brien. 2008. *Introduction to Information System, edition 12. (Pengantar Sistem Informasi Perspektif Bisnis dan Manajerial*. Jakarta: PT. Salemba Empat (Emban Patria).
- Montgomery, D.C. 2009. *Introduction to Statistical Quality Control*. Mc Graw Hill.
- Prayoga. (2017). Sistem Pendukung Keputusan Pemilihan Supplier Botol Galon menggunakan Metode AHP". Jurnal Teknologi dan Manajemen Industri, Vol. 2 No. 2. Agustus 2017.
- Saaty. 2018. *The Analytic Hierarchy Process*. McGraw Hill International.
- Saragih. (2013). "Penerapan Metode AHP (*Analytic Hierarchy Process*) Pada Sistem Pendukung Keputusan Pemilihan Laptop". Pelita Informatika Budi darma, Vol. IV No. 2, Agustus 2013, ISSN : 2301-9425.
- Sestri. (2013). "Penilaian Kinerja Dosen Dengan Menggunakan Metode AHP (Studi kasus : STIE Ahmad Dahlan Jakarta)". Jurnal Liquidity, Vol. 2 No. 1. Januari - juni 2013, Hal. 100-109.
- Ratniasih, Ni Luh, 2019. Optimasi Data Mining Menggunakan Algoritma Naïve Bayes dan C4.5 Untuk Klasifikasi Kelulusan Mahasiswa. Jurnal Teknologi Informasi dan Komputer, Volume 5, Nomor 1, Januari 2019.
- Sunjana. (2010). "Klasifikasi Data Nasabah Sebuah Asuransi menggunakan Algoritma C4.5". Seminar Nasional Aplikasi Teknologi Informasi, 19-20 Juni 2019, ISSN: 1907-5022.
- Widiasri, M., Justitia, A., & Arifin, A. Z. (2011). Penerapan Particle Swarm Optimization untuk Penentuan Parameter Regularisasi pada Kernel Regularized Discriminant Analysis. Industrial Electronics Seminar, 61-66, 26 Oktober 2011