

# PREDIKSI KELAHIRAN BAYI SECARA PREMATUR DENGAN MENGGUNAKAN ALGORITMA C.45 BERBASIS PARTICLE SWARM OPTIMIZATION

Ari Puspita

**Abstract** - Preterm labor can happen when pregnancy has not entered the 37th week, or three weeks or more before the birth day forecast (HPL). Babies who are born prematurely usually weigh less than 2.5 kilograms it is that causes the body's organs are not functioning properly infants can suffer from a variety of more serious health problems than babies who were born on schedule birth. Until now, there are some cases of premature labor of unknown cause. There are several factors and health problems that can trigger preterm labor that mothers who do not exercise, smoking, history of pregnancy, fetal condition, psychological condition. To the authors intend to make a research on how to predict a patient who will give birth prematurely. After testing the algorithm C4.5 models that the results obtained are C4.5 algorithm produces a value of 93.60% accuracy and AUC value of 0.946 to the level of diagnostics Excellent Classification. But after namely, the addition of C4.5 algorithm based particle swarm optimization accuracy rate of 96.00% and the AUC value of 0.967 with a diagnosis rate Excellent Classification. So that both methods have tinker difference in the amount of 2.4% accuracy.

**Intisari** – Persalinan prematur bisa terjadi ketika kehamilan belum memasuki minggu ke-37, atau tiga minggu atau lebih sebelum perkiraan hari lahir (HPL). Bayi yang lahir prematur biasanya kurang dari 2,5 kilogram itu yang menyebabkan organ tubuh tidak berfungsi dengan baik bayi bisa menderita berbagai masalah kesehatan yang lebih serius daripada bayi yang lahir pada jadwal kelahiran. Hingga saat ini, ada beberapa kasus persalinan prematur penyebabnya tidak diketahui. Ada beberapa faktor dan masalah kesehatan yang dapat memicu persalinan prematur bahwa ibu yang tidak berolahraga, merokok, riwayat kehamilan, kondisi janin, kondisi psikologis. Untuk itu penulis bermaksud untuk melakukan penelitian tentang bagaimana untuk memprediksi pasien yang akan melahirkan prematur. Setelah menguji model algoritma C4.5 bahwa hasil yang didapatkan algoritma C4.5 menghasilkan nilai akurasi 93.60% dan nilai AUC 0,946 dengan tingkat diagnosa Klasifikasi sangat baik. Tapi setelah yaitu penambahan partikel swarm tingkat algoritma C4.5 berdasarkan optimasi akurasi 96.00% dan nilai AUC 0,967 dengan tingkat diagnosis yang sangat baik Klasifikasi. Sehingga kedua metode memiliki tinker perbedaan jumlah akurasi 2,4%.

**Keywords:** Preterm, C4.5 Algorithm, Particle Swarm Optimization, Pregnancy

Program Studi Teknik Informatika STMIK Antar Bangsa,  
Kawasan Bisnis CBD Ciledug. Blok A5 No.22-28 Jl.HOS  
Cokroaminoto Karang Tengah. Tangerang (telp:021-73453000; e-  
mail : [aripuspita@gmail.com](mailto:aripuspita@gmail.com))

## I. PENDAHULUAN

Persalinan prematur dapat terjadi ketika kehamilan belum memasuki minggu ke 37, atau tiga minggu atau lebih sebelum hari perkiraan lahir (hpl). Bayi yang dilahirkan secara prematur biasanya memiliki berat badan kurang dari 2,5 kilogram hal ini yang menyebabkan organ tubuh bayi tidak berfungsi dengan baik dapat menderita berbagai masalah kesehatan yang lebih serius dibandingkan bayi-bayi yang dilahirkan sesuai jadwal kelahirannya. Hingga saat ini, ada beberapa kasus persalinan prematur yang belum diketahui penyebabnya. Ada beberapa faktor dan masalah kesehatan yang dapat memicu persalinan prematur yaitu ibu yang tidak sehat, merokok, riwayat kehamilan, kondisi janin, kondisi psikologis [2]. Kelahiran bayi prematur di Indonesia masih menjadi salah satu kasus yang sangat tinggi. Data dari beberapa rumah sakit menunjukkan presentasi antara 14-20% dari seluruh bayi yang dirawat. Keadaan ini terutama disebabkan masalah sosial-ekonomi yang dialami sebagian besar masyarakat Indonesia. Insidensi prematur dengan berat bayi lahir rendah (BBLR) di Indonesia menurut *World Health Organization (WHO)* tahun 1990 adalah 14% sedangkan menurut *United Nation Children's Emergency Fund (Unicef)* tahun 1995 insidensinya 11%. Laporan dari instalasi maternal perinatal RSUP Dr. Sarjito Yogyakarta tahun 2002, dari 930 bayi yang dirawat terdapat 20,8% bayi prematur, 33% BBLR, dan 20% retardasi pertumbuhan intra uteri [2]

Masalah yang dihadapi oleh bayi prematur adalah masih lemahnya sistem organ dan fungsinya untuk beradaptasi dengan lingkungan di luar rahim. Imaturnya sistem imunitas atau kekebalan, paru-paru atau respirasi, termoregulasi, kardiovaskuler, gastrointestinal dan lainnya [1]

## II. KAJIAN LITERATUR

### 1. Prematur

Prematur merupakan salah satu penyebab morbiditas dan mortalitas utama selama periode neonatus. Permasalahan yang timbul akibat kelahiran prematur pada periode perinatal meliputi sistem serebrospinal yang disebabkan asfiksia atau iskemik, perdarahan intraventrikuler, sindrom distres pernafasan, displasia bronkopulmoner, fibroplasia retrolental, infeksi atau sepsis dan masalah metabolik(asidosis,hipoglikemia,hiperbilirubinemia, dll).

Sedangkan menurut [7] Persalinan prematur adalah persalinan yang dimulai setiap saat setelah awal minggu gestasi ke-20 sampai akhir minggu gestasi ke-37. Persalinan premature merupakan persalinan yang terjadi pada kehamilan kurang dari 37 minggu (antara 20-37 minggu) atau dengan berat janin kurang dari 2500 gram. Masalah utama dalam persalinan prematur adalah perawatan bayinya, semakin muda usia kehamilannya semakin besar morbiditas dan mortalitasnya. [9]

## 2. Data Mining

*Data mining* didefinisikan sebagai proses menemukan dan menggambarkan pola struktural dalam data sebagai alat untuk membantu menjelaskan data dan membuat prediksi dari data tersebut [8]. *Data mining* telah menarik banyak perhatian dalam dunia sistem informasi dan dalam masyarakat secara keseluruhan dalam beberapa tahun terakhir, karena ketersediaan luas dalam jumlah besar data dan kebutuhan segera untuk mengubah data tersebut menjadi informasi yang berguna dan pengetahuan. Informasi dan pengetahuan yang diperoleh dapat digunakan untuk aplikasi mulai dari pasar analisis, deteksi penipuan, dan retensi pelanggan, untuk pengendalian produksi dan ilmu pengetahuan eksplorasi [7]. Adanya ketersediaan data yang melimpah, kebutuhan akan informasi atau pengetahuan sebagai sarana pendukung dalam pengambilan keputusan baik bagi individu, organisasi, perusahaan dan pemerintahan.

## 3. Klasifikasi

Klasifikasi merupakan bagian dari prediksi, dimana nilai yang diprediksi berupa label. Klasifikasi menentukan *class* atau grup untuk tiap contoh data, *input* dari model klasifikasi adalah atribut dari contoh data (*data samples*) dan *output*nya adalah *class* dari data *samples* itu sendiri, dalam *machine learning* untuk membangun model klasifikasi digunakan metode *supervised learning* [6]. Metode *supervised learning* yaitu metode yang mencoba untuk menemukan hubungan antara atribut masukan dan atribut target, hubungan yang ditemukan diwakili dalam struktur yang disebut model.

## 4. Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan yang merupakan pengembangan dari algoritma ID3, algoritma tersebut digunakan untuk membentuk pohon keputusan. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer, dalam klasifikasi pohon keputusan terdiri dari sebuah node yang membentuk akar, node akar tidak memiliki inputan. Node lain yang bukan sebagai akar tetapi memiliki tepat satu inputan disebut *node internal* atau *test node*, sedangkan node lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu *class*[4]. Pohon keputusan dibuat dengan membagi nilai-nilai atribut menjadi cabang untuk setiap

kemungkinan. Cara kerja pohon keputusan yaitu dengan melakukan penelusuran dari akar hingga ke cabang sampai *class* suatu objek ditemukan, *instance* diklasifikasikan dengan mengarahkan dari akar pohon sampai ke daun sesuai dengan hasil tes melalui *node internal*.

Algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal [7]. Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5 [6] yaitu:

1. Mempersiapkan data *training*, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai *gain* yang tertinggi dari masing-masing atribut atau berdasarkan nilai *index entropy* terendah. Sebelumnya dihitung terlebih dahulu nilai *index entropy*, dengan rumus:

$$Entropy(i) = -\sum_{j=1}^m f(i,j) \cdot \log_2 f[(i,j)] \dots\dots (2.1)$$

Keterangan:

- i = himpunan kasus
- m = jumlah partisi i
- f(i,j) = proposi j terhadap i

3. Hitung nilai *gain* dengan rumus:

$$Entropy\ split = -\sum_{i=1}^p \frac{n_i}{n} \cdot IE(i) \dots\dots\dots (2.2)$$

Keterangan:

- p = jumlah partisi atribut
- n<sub>i</sub> = proporsi n<sub>i</sub> terhadap i
- n = jumlah kasus dalam n

## 5. Particle Swarm Optimization

*Particle Swarm Optimization* (PSO) adalah algoritma pencarian berbasis populasi dan diinisialisasi dengan populasi solusi acak yang disebut partikel (Abraham, 2006). PSO merupakan metode pencarian penduduk yang berasal dari penelitian untuk gerakan sekelompok burung atau ikan. Serupa dengan algoritma genetik (GA), PSO melakukan pencarian menggunakan populasi (*swarm*) dari individu (partikel) yang akan diperbaharui dari iterasi ke iterasi.

Untuk PSO dapat diasumsikan sebagai kelompok burung secara acak mencari makanan disuatu daerah. Burung tersebut tidak tahu dimana makanan tersebut berada, tapi mereka tahu seberapa jauh makanan itu berada, jadi strategi terbaik untuk menemukan makanan tersebut adalah dengan mengikuti burung yang terdekat dari makanan tersebut (Abraham, 2006). PSO digunakan untuk memecahkan masalah optimasi.

## 6. Evaluasi dan Validasi Hasil

*Cross Validation* adalah teknik validasi dengan membagi data secara acak kedalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi [7]. Dengan menggunakan *cross validation* akan dilakukan percobaan sebanyak k. Data yang digunakan dalam percobaan ini adalah data *training* untuk mencari nilai *error rate* secara keseluruhan. Secara umum pengujian nilai k dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi

### a. Confusion Matrix

*Confusion matrix* merupakan tabel *matrix* yang terdiri dari dua kelas, yaitu kelas yang dianggap sebagai positif dan kelas yang dianggap sebagai negatif [7]. *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada system klasifikasi.

### b. ROC Curve

Kurva ROC (*Receiver Operating Characteristic*) adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual [7]. Sebuah grafik ROC adalah plot dua dimensi dengan proporsi positif salah (*fp*) pada sumbu X dan proporsi positif benar (*tp*) pada sumbu Y. Titik (0,1) merupakan klasifikasi yang sempurna terhadap semua kasus positif dan kasus negatif. Nilai positif salah adalah tidak ada ( $fp = 0$ ) dan nilai positif benar adalah tinggi ( $tp = 1$ ). Titik (0,0) adalah klasifikasi yang memprediksi setiap kasus menjadi negatif {-1}, dan titik (1,1) adalah klasifikasi yang memprediksi setiap kasus menjadi positif {1}. Grafik ROC menggambarkan *trade-off* antara manfaat (*'true positives'*) dan biaya (*'false positives'*). Berikut tampilan dua jenis kurva ROC (*discrete* dan *continuous*). biaya (*'false positives'*). urva ROC (*discrete* dan *continuous*).

## 7. Tinjauan Studi

Ada beberapa penelitian yang menggunakan data mining *decision tree*, secara garis besar tinjauan pustaka dalam tesis ini meliputi:

a. Predictors of Preterm Birth in Birth Certificate Data, Memprediksi kelahiran bayi secara premature berdasarkan akte kelahiran dilakukan dengan mengkomparasi beberapa model algoritma yaitu logistic regression, neural network, SVM, Bayesian classifier, dan cart. Dari ke lima model yang peneliti gunakan logistic regression yang memiliki nilai AUC tertinggi yaitu 0,605[5]

b. Prediction of preterm birth in symptomatic women using decision tree modeling for biomarkers, Dalam penelitian ini peneliti hanya menggunakan model *decision tree* dan peneliti melihat berdasarkan umur dari janin yang berada dalam rahim ibu.

Wanita 30 tahun berisiko untuk melahirkan bayi prematur. Dan perkembangan IQ pada janin pun akan berpengaruh. Ibu yang bergejala untuk melahirkan melahirkan prematur dapat di pengaruhi oleh tingkat

tekanan darah Usia janin 28-31 minggu memiliki nilai sensitivity tertinggi yaitu 100% , Specivity 71% ,PPV 20% ,NPV ,100% [8]

- c. Analysis of Cardiotocogram Data for Fetal Distress Determination by Decision TreeBased Adaptive Boosting Approach, Dalam penelitian ini peneliti mengoptimasi bobot nilai atribut dengan menggunakan *Adaboost*. Beberapa model algoritma yang dikomparasi dalam penelitian ini yaitu Naive Bayes nilai akurasi sebesar 81.562%, *Radial Basis Function Network* 85,98%, *Bayesian Network* nilai akurasi sebesar 85,983%, *Support Vector Machine* nilai akurasi sebesar 88,75%, *Neural Network* nilai akurasi sebesar 92,098%, *Decision Tree* nilai akurasi sebesar 92,427% .
- d. *Comparing Decision Tree Method Over Three Data Mining Software* dalam penelitian ini peneliti mengkomparasi model *Decision Tree* didalam ketiga *software data mining* Dalam penelitian ini, kami memperkenalkan data mining dan konsep utama dari metode pohon keputusan yang paling Metode klasifikasi yang efektif dan banyak digunakan. Selain itu, penjelasan singkat dari perangkat lunak pertambahan tiga data, yaitu *SPSS-Clementine*, *RapidMiner* dan *Weka* juga disediakan. Setelah itu, perbandingan dilakukan pada 3515 dataset nyata dalam hal akurasi klasifikasi antara tiga algoritma pohon keputusan yang berbeda dalam memesan untuk menggambarkan prosedur penelitian ini. Algoritma pohon keputusan yang paling akurat adalah *Decision Tree* 92.49% dengan *Rapidminer*. [6]

## III PEMBAHASAN

### 1. Tahap Pengumpulan Data

Teknik pengumpulan data adalah teknik atau cara-cara yang dapat digunakan untuk menggunakan data Dalam pengumpulan data terdapat sumber data, sumber data yang terhimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder

Data yang diperoleh merupakan data primer karena data tersebut diambil dari sebuah klinik persalinan yang terdapat di daerah Tangerang. Atribut yang terdapat didalamnya merupakan faktor- faktor penyebab untuk kelahiran bayi secara prematur. Faktor penyebab bayi yang dilahirkan secara prematur biasanya didominasi dari kebiasaan ibu dari janin tersebut. Faktor yang sering terjadi adalah apabila ibu dari janin tersebut mengkonsumsi rokok, sering mengalami keputihan.

### 2. Tahap Pengolahan Awal Data

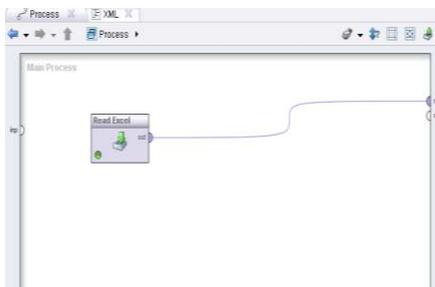
Ditahapan ini dijelaskan tentang tahapan awal data mining. Pengolahan awal data meliputi proses input data ke format yang dibutuhkan, pengelompokan dan penentuan

atribut data. Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 250 record, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*). Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan sebagai berikut [9]:

TABEL 1. ATRIBUT YANG DIGUNAKAN

No	Atribut	Nilai
1	Usia	Usia
2.	Sistol	Tekanan Darah Sistolic
3.	Diastol	Tekanan Darah Diastlic
4.	Riwayat darah tinggi	Tidak dan Ya
5.	Riwayat keguguran	Tidak dan Ya
6.	Riwayat prematur	Tidak dan Ya
7.	Trauma	Tidak dan Ya
8.	Konsumsi Rokok	Tidak dan Ya
9	Kehamilan ganda	Tidak dan Ya
10	Keputihan	Tidak dan Ya
11	Hasil	Prematur, Tidak Prematur

a. Data validation, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*). Data pasien yang ada akan diidentifikasi apakah ada data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*) lihat pada gambar berikut:



Gbr 1. Proses *Missing value* dengan Rapid Miner

b. Data *integration and transformationi*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan kedalam *software Rapidminer*.

c. Data *size reduction and discritization*, untuk memperoleh data set dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat informatif. Pada tahap ini data yang masih berupa nilai numerik dan juga *continue* ditransformasikan kedalam bentuk kategorikal dan dibuat skala atau interval sehingga menghasilkan *range* yang lebih kecil sebagai bahan pembelajaran algoritma C4.5

3. Model Algoritma C4.5

Algoritma C4.5 untuk model yang pertama dilakukan. Berikut langkah-langkah yang akan dilakukan sebagai berikut:

1. Menghitung jumlah kasus class YA dan class TIDAK serta nilai *Entropy* dari semua kasus. Kasus dibagi berdasarkan atribut pada Tabel 3.3 dengan jumlah kasus 250 *record*, kelas YA ada 93 *record* dan kelas TIDAK sebanyak 157 *record* sehingga didapat *entropy* Keseluruhan

$$Entropy(i) = - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j)$$

$$= (-93/250 \cdot \log_2 (93/250)) + (-157/250 \cdot \log_2 (157/250)) = 0,9521$$

2. Hitung nilai *entropy* dan nilai *gain* masing-masing atribut. Nilai *gain* tertinggi adalah atribut yang menjadi *root* dari pohon keputusan yang akan dibuat. Misalkan menghitung *entropy* bagi atribut Riwayat keguguran *Eya* [90,50] = (-90/140 log<sub>2</sub> 90/140) + (-50/140 log<sub>2</sub> 50/140) = 0,9403  
*Etidak* [3,107] = (-3/110 log<sub>2</sub> 3/110) + (-107/110 log<sub>2</sub> 107/110) = 0,1805

Nilai Entropynya dapat dihitung sebagai berikut:

$$Entropy\ split = \sum_{i=1}^p \frac{n_i}{n} IE(i)$$

$$E\ split\ riwayat\ keguguran = ((140/250 (0,9403) + (110/250 (0.1805)) = 0,4471$$

$$Gain\ usia = 0,9521 - 0,4471 = 0.5050$$

Perhitungan *entropy* dan *gain* untuk semua atribut dilakukan, untuk mendapatkan nilai *gain* tertinggi. Hasil perhitungan seluruh atribut terlihat pada Tabel 2

TABEL 2. NILAI ENTROPY DAN GAIN UNTUK MENENTUKAN

Simpul	AKAR			Entropy	Gain
	Kasus	Prematur	Tidak Prematur		
Jumlah Kasus	250	93	157	0,9521	
Riwayat Keguguran					
Pemah	140	90	50	0,9403	0,5050
Tidak	110	3	107	0,1805	
Riwayat Darah Tinggi					

Sumber: Hasil Pengolahan Data 2014

Dari tabel 2 dapat dilihat nilai *gain* tertinggi ada pada atribut riwayat keguguran yakni 0,5050 sehingga dapat atribut riwayat keguguran adalah akar dari pohon keputusan. Kemudian dilakukan kembali perhitungan nilai entropy dan *gain* untuk menentukan simpul 1.1, Perhitungan *entropy* dan *gain* untuk atribut riwayat keguguran disajikan dalam tabel 3.

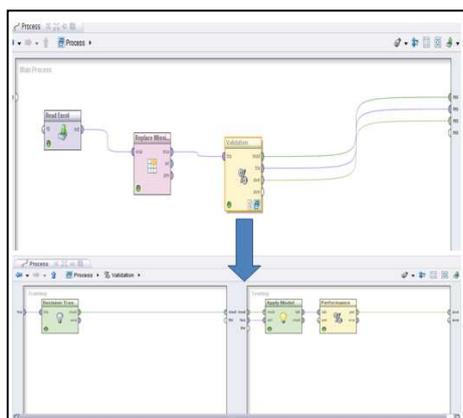
TABEL 3  
NILAI ENTROPY DAN GAIN UNTUK MENENTUKAN SIMPUL 1.1

Simpul	Kasus	Prematur	Tidak Prematur	Entropy	Gain
Jumlah	140	90	50	0,9403	
Kasus					
Riwayat Darah Tinggi					
Ya	110	88	22	0,7219	0,4606
Tidak	30	2	28	0,3534	
Riwayat Prematur					
Ya	123	80	43	0,9337	0,2505
Tidak	17	10	7	0,9774	
Trauma					

Sumber: Hasil Penelitian (2015)

#### 4. Evaluasi dan Validasi Hasil

Setelah diolah maka dilakukan teknik pengujian dengan metode *k-fold crossvalidation* pada tools *RapidMiner*, pegolahan pengujian untuk metode algoritma C4.5 terlihat seperti gambar 4.2 dibawah ini:



Gbr 2. Pengujian K-Fold Cross Validation Metode Algoritma C4.5

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

##### a. Confusion Matrix

Berikut Tabel yang didapat tersebut menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma C4.5 adalah sebesar 93,25%, dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* hasilnya pada Tabel 4 dibawah ini:

TABEL 4.

accuracy: 93.60% +/- 3.20% (mikro: 93.60%)			
	true Tidak Prematur	true Prematur	class precision
pred. Tidak Prematur	152	11	93.25%
pred. Prematur	5	82	94.25%
class recall	96.82%	88.17%	

MODEL CONFUSION MATRIX UNTUK ALGORITMA C4.5

TABEL 5

NILAI ACCURACY, SENSITIVITY, SPECIFICITY, PPV, DAN NPV

	%
Accuracy	93,60
Sensitivity	93,25
Specificity	94,25
PPV	96,82
NPV	88,17



Gbr 3. Nilai AUC dalam grafik ROC algoritma C4.5

##### b. Evaluasi ROC Curve

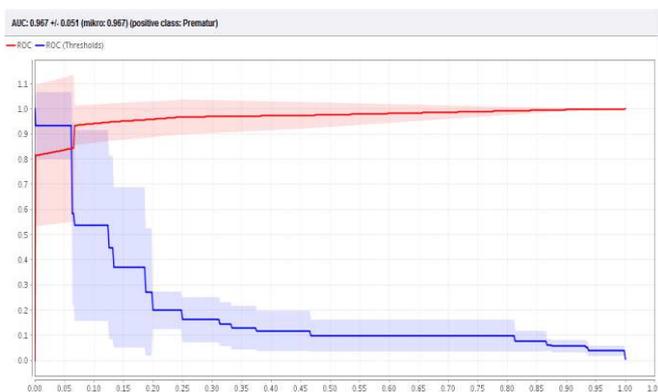
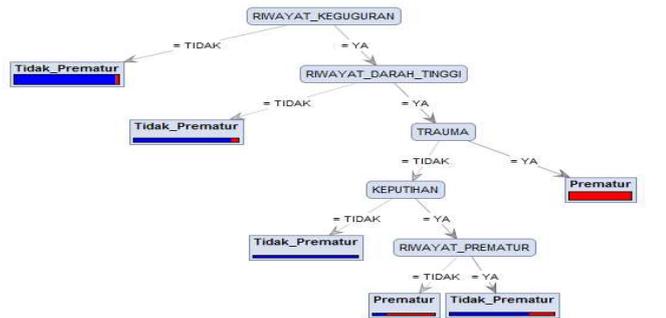
Dari Gambar 5 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.946. Tingkat akurasi dapat di diagnosa sebagai berikut (Gournescu, 2011):

- Akurasi 0.90 – 1.00 = *Excellent classification*
- Akurasi 0.80 – 0.90 = *Good classification*
- Akurasi 0.70 – 0.80 = *Fair classification*

- d. Akurasi 0.60 – 0.70 = *Poor classification*  
e. Akurasi 0.50 – 0.60 = *Failure*

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

Pada penelitian prediksi kelahiran bayi secara prematur menggunakan algoritma C4.5 berbasis *Particle Swarm Optimization*. *Particle Swarm Optimization* memiliki pencarian kinerja yang lebih baik untuk memecahkan banyak masalah optimasi dengan lebih cepat dan tingkat konvergensi yang stabil



Gbr 4 Pohon Keputusan Algoritma C4.5 Berbasis PSO

### 6. Hasil Pengujian Model Algoritma C4.5 Berbasis PSO

Gbr 5. Nilai AUC dalam grafik ROC algoritma C4.5 berbasis PSO

### 7. Analisis Evaluasi dan Validasi Model

Dari hasil pengujian diatas, baik evaluasi menggunakan *confusion matrix* maupun *ROC curve* terbukti bahwa hasil pengujian algoritma C4.5 berbasis PSO memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma C4.5. Nilai akurasi untuk model algoritma C4.5 sebesar 93.6% dan nilai akurasi untuk model algoritma C4.5 berbasis PSO sebesar 96,00% dengan selisih akurasi 2.4%, dapat dilihat pada Tabel 6 dibawah ini:

TABEL 6 PENGUJIAN ALGORITMA C4.5 DAN C4.5 BERBASIS PSO

Algoritma	Accuracy	AUC
C4.5	93.60%	0.946
C4.5 Berbasis PSO	96.00%	0.697

### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Allah SWT, Orang tua, rekan- rekan yang telah memberikan *support* dan bantuan hingga terselesaikan Penelitian ini

### KESIMPULAN

Dari hasil prediksi model algoritma C4.5 berbasis Decision Tree memberikan nilai akurasi yaitu 93.60%. Dengan nilai akurasi yang sudah mencapai *Excellent classification* hal ini menunjukkan bahwa Decision Tree merupakan model yang cukup baik.

### DAFTAR PUSTAKA

- [1] Asmara, F.Y. (2004). Pengaruh Stres Psikososial, Merokok Pasif, Dan Konsumsi Kopi Terhadap Kelahiran Prematur. Skripsi, Bagian Ilmu Kesehatan Anak FK UGM, Yogyakarta.
- [2] Courtney ,Karen L, dkk(2008) *Predictors of Preterm Birth in Birth Certificate Data. eHealth Beyond the Horizon – Get IT There S.K. Andersen et al. (Eds.)IOS Press*
- [3] Gorunescu, Florin. (2011). *Data Mining: Concepts and Techniques*. Verlag berlinHeidelberg: Springer
- [4] Han, J., & Kamber, M. (2006). *Data Mining Concepts and technique*. San Francisco: Diane Cerra
- [5] Hill, Jacquelyn L, dkk (2008). *Prediction Of Preterm Birth In Symptomatic Women Using Decision Tree Modeling For Biomarkers. American Journal of Obstetrics*
- [6] Moghimipour, Ida dan Ebrahimipou( 2014). *Comparing Decision Tree Method Over Three Data Mining Software. International Journal of Statistics and Probability; Vol. 3, No. 3*
- [7] Vercellis, C. (2009). *Business Intelligent: Data Mining and Optimization for Decision Making*. Southern Gate: John Wiley & Sons Inc.
- [8] Varney, Helen. (2007). *Buku Ajar Asuhan Kebidanan*. Edisi 4. Jakarta: EGC
- [9] Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques*. United States: Elsevier Inc
- [10] Saifuddin, A B. 2009. *Buku Acuan Pelayanan Kesehatan Maternal dan Neonatal*. Jakarta: YBPSP
- [11] Karabulut, Esra Mahsereci dan Ibricki, Turgay (2014). *Analysis Of Cardiotocogram Data for Fetal Distress Determination by Decision Tree Based Adaptive Boosting Approach. Journal of Computer and Communications. 32-37*



Ari Puspita. Lulus dari Magister Ilmu Komputer Program Pascasarjana STMIK Nusa Mandiri Tahun 2015. Saat ini aktif sebagai tenaga pengajar di STMIK Antar Bangsa.