

KOMPARASI METODE *DATA MINING* UNTUK PENENTUAN PROSES PERSALINAN IBU MELAHIRKAN

Hilda Amalia¹ dan Evicienna²

¹Manajemen Informatika, AMIK BSI JAKARTA, Jl. RS. Fatmawati No.24, Jakarta, 12450, Indonesia

²Komputerisasi Akuntansi, AMIK BSI JAKARTA, Jl. RS. Fatmawati No. 24, Jakarta, 12450, Indonesia

E-mail: hilda@bsi.ac.id, evicienna@bsi.ac.id

Abstract

Birth process is a thing awaited by every mother who was giving birth. In the process of labor there are two types of process of persalinaan that is normal or cesarean. In the labor process there is a risk of labor that is encountered maternal complication that can worsen the condition of the mother giving birth. The worst risk that can happen is the death of the mother and / or the newborn. *Data mining* is a popular method used to extract patterns or knowledge from large data piles. In this research, maternal data processing is done using C4.5, *naive bayes* and *neural network* algorithm. From the research, it is known that the accuracy value for maternal data processing for determination of delivery process with *naive bayes* method is 94%, 936% *neural network*, and C4.5 algorithm is 90%.

Keywords: *data mining, birth, comparison*

Abstrak

Proses Persalinan merupakan hal yang dinanti oleh setiap ibu yang sedang melahirkan. Dalam proses persalinan terdapat dua jenis proses persalinan yaitu secara normal atau sesar. Dalam proses persalinan terdapat resiko persalinan yang dihadapi yaitu komplikasi ibu melahirkan yang dapat memperburuk kondisi ibu melahirkan. Resiko terburuk yang dapat saja terjadi adalah kematian ibu dan/ atau bayi yang baru dilahirkan. *Data mining* merupakan metode yang populer digunakan untuk menggali pola atau ilmu pengetahuan dari tumpukan data yang besar. Dalam penelitian ini dilakukan komparasi metode *data mining* untuk pengolahan data ibu melahirkan dengan menggunakan algoritma C4.5, *naive bayes* dan *neural network*. Dari penelitian diketahui nilai akurasi untuk pengolahan data ibu melahirkan untuk penentuan proses persalinan dengan metode *naive bayes* yaitu 94%, *neural network* 936%, dan algoritma C4.5 yaitu 90%.

Kata Kunci: *data mining, persalinan, komparasi*

1. Pendahuluan

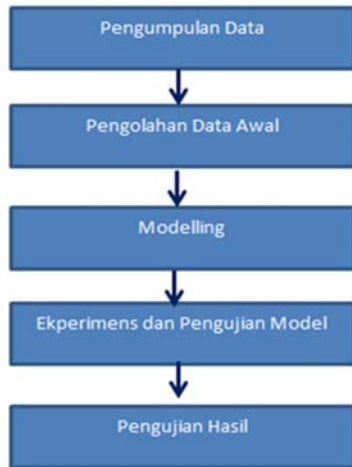
Resiko persalinan sering terjadi pada saat proses melahirkan. Resiko terburuk yaitu berupa kematian. Kematian ibu dan anak merupakan hal yang harus dihindari dalam proses persalinan. Tingginya angka kematian Ibu disebabkan oleh komplikasi pada proses kehamilan, persalinan dan nifas. Persalinan adalah suatu proses pengeluaran hasil konsepsi yang dapat hidup dari dalam uterus melalui vagina ke dunia luar [1]. Berdasarkan hasil Survei Kesehatan Rumah Tangga di Indonesia tahun 2001 diketahui kematian ibu yang dikarenakan komplikasi kehamilan, persalinan dan nifas sebanyak 89,5% dan sekitar 10,5% dikarenakan penyakit [2].

Berdasarkan survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2012, angka kematian ibu di Indonesia terbilang masih tinggi, yaitu 359 per 100.000 kelahiran hidup, sedangkan target gl-

obal MDGs (*Millenium Development Goals*) kelima yaitu menurunkan Angka Kematian Ibu (AKI) menjadi 102 per 100.000 kelahiran hidup [2].

Sementara itu, penyebab kematian pada bayi yang baru dilahirkan disebabkan oleh *premature* sebanyak 32%, asfiksia yaitu 22%, kelainan kongenital sebesar 7%, dan lain-lain 9% [3]. Untuk menghindari resiko-resiko yang terjadi pada proses persalinan dalam dunia kesehatan banyak digunakan prediksi klinis sistem penunjang keputusan.

Prediksi klinis berkembang sangat pesat dengan mengadopsi ilmu komputer dan teknologi informasi dalam pengolahan data kesehatan dan obat [4]. Prediksi klinik ini dapat dihasilkan dari pengolahan dengan menggunakan metode *data mining*. *Data mining* merupakan metode yang digunakan untuk menggali ilmu pengetahuan dari kumpulan data. Kumpulan atau tumpukan data yang hanya tersimpan diolah dengan *data mining* untuk meng-



Gambar 1. Metode Penelitian yang Digunakan

hasilkan ilmu pengetahuan yang dapat dimanfaatkan untuk kebutuhan prediksi klinis.

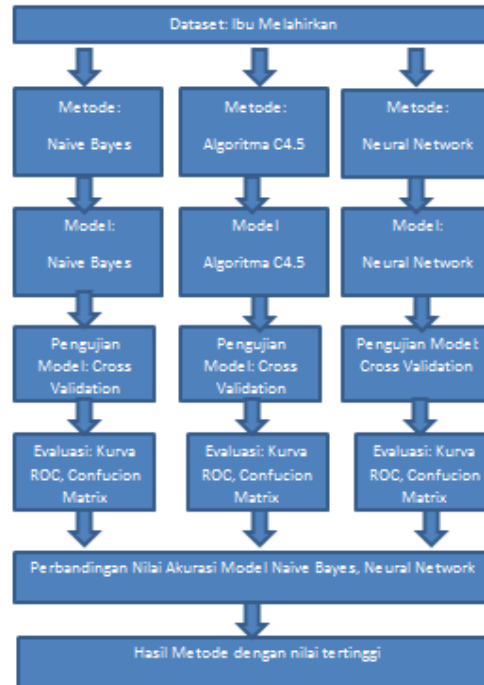
Penelitian terdahulu mengenai data-data kesehatan dibidang proses melahirkan dengan menggunakan metode *data mining* telah dilakukan oleh beberapa peneliti, diantaranya yaitu penggunaan *fuzzy* [5], dengan menggunakan metode *neural network* [5], serta komparasi *decision tree*, ANN dan *assosiate rule* [6]. Komparasi metode yaitu Algoritma J48, Simple CART dan *Naive bayes* untuk prediksi proses persalinan pada ibu melahirkan yang memiliki riwayat penyakit diabetes [7]. Dalam penelitian ini akan digunakan data ibu melahirkan dengan menggunakan metode *naive bayes*, algoritma C4.5 dan *neural network*.

2. Metode

Data mining adalah suatu proses logis yang digunakan untuk mencari dan menemukan pola melalui sejumlah besar data [8]. Tujuan dari teknik ini adalah untuk menemukan pola yang sebelumnya tidak diketahui. Dalam *data mining* tumpukan data masa lalu ini dianggap sebagai tambang yang dapat dioalah untuk menghasilkan sebuah pengetahuan yang berharga.

Algoritma C4.5

Algoritma C4.5 salah satu metode *data mining* yang digunakan untuk melakukan tugas *data mining* seperti klasifikasi dan estimasi [9]. Algoritma C4.5 merupakan model *tree* seperti sebuah pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas.



Gambar 2. Ilustrasi Penggunaan Metode Yang Dibutuhkan

Perhitungan metode algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal [10]. Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5 [11] 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 indeks entropi terendah. Sebelumnya dihitung terlebih dahulu nilai indeks entropi, dengan persamaan(1).

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

dimana i menunjukkan himpunan kasus, m menunjukkan jumlah partisi i , dan $f(i,j)$ merupakan proporsi j terhadap i .

Setelah itu, hitung nilai *gain* dengan persamaan(2). Pada persamaan tersebut, p adalah jumlah partisi atribut, n_i proporsi n_i terhadap i , dan n adalah jumlah kasus dalam n .

$$\text{Entropi split} = \sum_{i=1}^p \binom{n_i}{n} \cdot 1E \quad (2)$$

TABEL 1.
DATASET YANG DIGUNAKAN

| BB Janin | Kondisi Pasien | Hasil Pemeriksaan Pasien | Kondisi Janin | Hasil Pemeriksaan Janin | Hasil Akhir Proses Persalinan |
|----------|----------------|--------------------------|---------------|-------------------------|-------------------------------|
| 2.57 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 2.80 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 2.60 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 3.00 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 2.87 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 3.20 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 3.00 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 3.50 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 2.50 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 2.70 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 3.10 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |
| 2.50 kg | sehat | sehat dan kuat | normal | kuat dan normal | normal |

Setelah itu, ulangi langkah kedua hingga semua *record* terpartisi. Proses partisi pohon keputusan akan berhenti disaat: 1) Semua tupel dalam *record* dalam simpul *m* mendapat kelas yang sama; 2) Tidak ada atribut dalam *record* yang dipartisi lagi; 3) Tidak ada *record* didalam cabang yang kosong.

Neural network

Jaringan saraf adalah seperangkat unit *input/output* yang terhubung dan setiap sambungan memiliki bobot yang menyertainya. Selama tahap pembelajaran, jaringan belajar dengan menyesuaikan diri sehingga bisa memprediksi kelas yang benar label dari *input* tupel. Jaringan syaraf tiruan memiliki kemampuan yang luar biasa untuk mendapatkan makna dari rumit atau data yang tidak tepat dan dapat digunakan untuk mengekstrak pola dan mendeteksi tren yang terlalu rumit untuk diperhatikan baik manusia atau teknik komputer lainnya [8].

Naive bayes

Naïve bayes untuk menambang data mengenai donor darah. *Naive bayes* merupakan salah satu algoritma klasifikasi dalam *data mining* yang menggunakan teori probabilitas untuk menemukan kemungkinan yang paling tepat [12]. Dalam penelitian ini dilakukan beberapa langkah seperti pada Gambar 1. Berikut tahapan penelitian yang dilakukan.

Pengumpulan Data

Data dikumpulkan dari sebuah rumah sakit yang didaerah Jakarta. sampai saat ini data yang telah dikumpulkan adalah sebanyak 130 *record* dan terdiri dari lima atribut dengan satu atribut label.

Pengolahan Data Awal

Dalam pengolahan data awal dilakukan beberapa tahapan yaitu:

Data validation

Tahap ini digunakan untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*). Dalam penelitian ini dari total 130 *record* terdapat data noise, tidak lengkap sebanyak 30 *record*. Sehingga total *record* yang diperoleh yaitu 130 *record*.

Data integration and Transformation

Tahap ini digunakan untuk meningkatkan akurasi dan efisiensi algoritma. Dalam penelitian ini data diubah menjadi data kategorikal yang mampu mewakili data seperti atribut hasil pemeriksaan pasien.

Data size reduction and discretization

Tahap ini digunakan untuk memperoleh dataset dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat informatif. Penelitian ini menggunakan dataset yang dikumpulkan dari buku pasien. Kemudian data yang digunakan adalah data yang bisa mewakili dataset. Rincian dataset yang digunakan dapat dilihat pada Tabel 1.

Metode yang diusulkan

Peneliti membuat skema metode yang diusulkan seperti yang disajikan pada Gambar 2. Penelitian menggunakan tiga metode yang menghasilkan nilai akurasi. Nilai tersebut dibandingkan untuk mendapatkan metode yang terbaik.

3. Hasil dan Analisis

Dalam penelitian ini metode yang digunakan adalah metode algoritma C4.5, dan tool yang digunakan untuk pengolahan data yaitu RapidMiner.

Algoritma C4.5

Langkah-langkah pengolahan data dengan menggunakan RapidMiner untuk metode algoritma C4.5

TABEL 2
CONFUSION MATRIX ALGORITMA C4.5

| Accuracy:90% | | | |
|--------------|-------------|------------|-----------------|
| | True Normal | True Secar | Class Precision |
| Pred. Normal | 58 | 5 | 92,06% |
| Pred. Secar | 5 | 32 | 46,48% |

TABEL 3
CONFUSION MATRIX NAIVE BAYES

| Accuracy:94% | | | |
|--------------|-------------|------------|-----------------|
| | True Normal | True Secar | Class Precision |
| Pred. Normal | 64 | 4 | 93,85% |
| Pred. Secar | 2 | 33 | 94,46% |

menggunakan dua modul utama yaitu modul *read excel*, yang digunakan untuk menampung data awal dalam format excel kemudian dihubungkan dengan modul *validation*. Di dalam modul *validation* terdapat modul *decision tree*, *apply* model dan *performance*. Hasil *running* modul-modul tersebut menghasilkan model algoritma C4.5 yang dapat dilihat pada Lampiran Gambar I. nilai akurasi metode algoritma C4.5 seperti pada Tabel 2. Selain menghasilkan nilai akurasi dari pengolahan Rapid-Miner juga menghasilkan kurva ROC seperti pada Lampiran Gambar II.

Naive bayes

Langkah-langkah pengolahan dataset ibu melahirkan dengan menggunakan *Naive bayes* dengan menggunakan tools RapidMiner. Modul yang digunakan adalah modul *read excel* yang berisi dataset proses persalinan ibu melahirkan dihubungkan dengan modul *validation*, di dalam modul *validation* terdapat modul metode yang digunakan yaitu *naive bayes* dihubungkan dengan modul *performance* dan modul *apply* model.

Hasil pengolahan modul-modul tersebut menghasilkan nilai akurasi metode *naive bayes* seperti terlihat pada Tabel 3. Sementara tampilan kurva ROC metode *Naive bayes* tertera pada Lampiran Gambar III.

Neural network

Langkah-langkah pengolahan dataset ibu melahirkan dengan menggunakan *Neural network* dengan menggunakan tools RapidMiner. Modul yang digunakan adalah modul *read excel* yang berisi dataset proses persalinan ibu melahirkan dihubungkan dengan modul *validation*, di dalam modul *validation* terdapat modul metode yang digunakan yaitu *naive bayes* dihubungkan dengan modul *performance* dan modul *apply* model. Hasil pengolahan modul-modul diatas menghasilkan nilai akurasi metode *neural network* seperti tercantum pada Tabel 4.

TABEL 4
CONFUSION MATRIX NEURAL NETWORK

| Accuracy:93% | | | |
|--------------|-------------|------------|-----------------|
| | True Normal | True Secar | Class Precision |
| Pred. Normal | 62 | 6 | 91,10% |
| Pred. Secar | 1 | 21 | 96,86% |

TABEL 5
PERBANDINGAN AKURASI TIGA METODE

| | Algoritma C4.5 | Naive Bayes | Neural Network |
|---------|----------------|-------------|----------------|
| Akurasi | 90,00 | 94,00 | 93,00 |
| AUC | 0,815 | 0,971 | 0,962 |

Kurva ROC yang dihasilkan metode *Neural network* dapat dilihat pada Lampiran Gambar IV. Sedangkan Lampiran Gambar V menunjukkan model yang dihasilkan dari metode *Neural network*. Selanjutnya, perbandingan nilai akurasi untuk pengolahan dataset ibu melahirkan dari ketiga metode tersebut dapat dilihat pada Tabel 5.

4. Kesimpulan

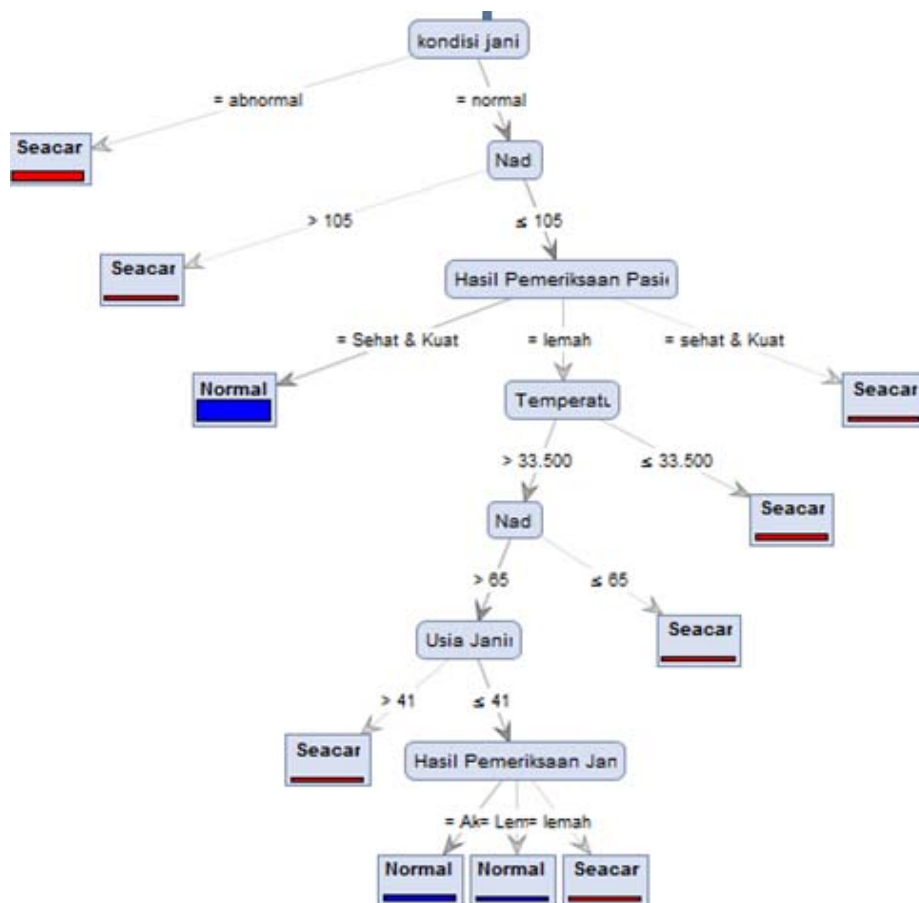
Dalam pengolahan data ibu melahirkan untuk penentuan proses ibu melahirkan menggunakan tiga metode *data mining* yaitu *neural network*, Dari tabel perbandingan nilai akurasi diatas diketahui metode dengan nilai akurasi tertinggi yaitu *naive bayes* yaitu 94% dan nilai AUC tertinggi juga yaitu 0.971. Performace AUC dikelompokan kedalam lima kelompok yaitu: 0.90–1.00 disebut Klasifikasi yang sangat baik; 0.80–0.90 disebut Klasifikasi Bagus; 0.70–0.80 disebut Klasifikasi Lepas; 0.60–0.70 disebut Klasifikasi Miskin [11]. Dalam penelitian untuk nilai AUC *naive bayes* dan *neural network* masuk ke dalam kategori sangat baik, sementara algoritma C4.5 termasuk ke dalam kategori bagus. Untuk penelitian berikutnya diperlukan penelitian lebih lanjut mengenai atribut yang mempengaruhi pengolahan data pengiriman. Untuk penelitian selanjutnya diterima untuk menggunakan lebih banyak data guna meningkatkan kualitas hasil penelitian.

Referensi

- [1] Mansjoer, A., et al., "Kapita Selekta Kedokteran," Jakarta: Media Aesculapius, 2000.
- [2] Tim Kajian AK. "Kajian kematian ibu dan anak di Indonesia," *Badan Penelitian dan Pengembangan Kesehatan*. Jakarta: Depkes RI, 2004.
- [3] H. Wiknjastro, "Ilmu Kebidanan", Jakarta: PT. Bina Pustaka Sarwono Prawirohardjo, 2000.

- [4] Krishnaiah, V., Net al., "A Study On Clinical Prediction Using *Data mining* Techniques," *International Journal of Computer Science Engineering and Information Technology Research (IJCEITR)*, vol. 3, no. 1, pp. 239-248, 2013.
- [5] J. J. Janghel and P. Tiwari, "Decision Support System for Fetal Delivery using Soft Computing Technique," *Fourth International Conference on Computer Science and Convergence Information Technonology*, 2009.
- [6] A. Sana, et al., "Automated Diagnosis and Cause Analysis of Cesarean Section Using Machine Learning Technique," *International Journal of Machine Learning and Computing*, vol. 2, no. 5, 2012.
- [7] Nagarajan, S., et al., "Comparison of Various *Data mining* Algorithms in the Prediction of Risk for Gestational Diabetes," *In Proceedi- ngs of the National Conference on Recent Trends in Mathematical Computing–NCRPMC*, vol. 13, pp. 192, 2013.
- [8] R. M. Bharati, "Data mining Technique and Applications," *Indian Journal of Computer Science and Engineering*, Vol.1 No.4 301-305, ISSN: 0976-5166, 2011.
- [9] J. Han and M. Kamber, "Data mining Concepts and Techniques", San Fransisco: Mofgan Kaufan Publisher, 2007.
- [10] Larose, D.T., "Discovering Knowledge in Database," New Jersey: Jon Willey & Sons Inc, 2005.
- [11] F. Gorunescu, "Data mining Concept and Technique", Verlag Berlin Heidelberg: Springer, 2011.
- [12] B. Max, *Principles of Data Mining, Undergraduate Topics in Computer Science* ISSN: 1863-7310, London: Springer, 2007.

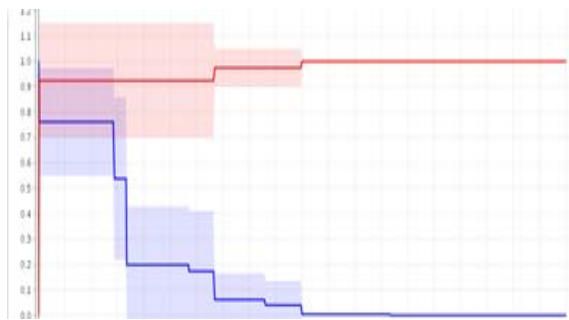
Lampiran



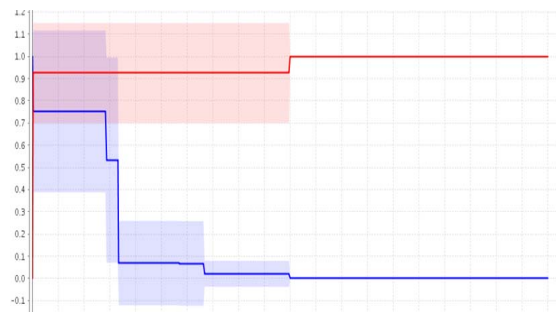
Lampiran Gambar I. Model Pohon Keputusan Algoritma C4.5



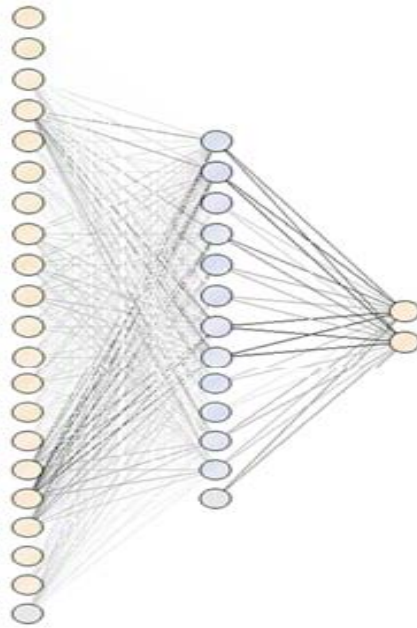
Lampiran Gambar II. Kurva ROC Algoritma C4.5. Sumber: hasil pengolahan data Rapidminer



Lampiran Gambar III. Kurva ROC Metode Naive Bayes. Sumber: hasil pengolahan data RapidMiner



Lampiran Gambar IV. Kurva ROC Metode Neural Network. Sumber: pengolahan data RapidMiner



Lampiran Gambar V. Model Neural Network. Sumber: Hasil Pengolahan RapidMiner