

# Kajian Komparasi Penerapan *Algoritma C4.5*, *Naïve Bayes*, dan *Neural Network* dalam Pemilihan Mitra Kerja Penyedia Jasa Transportasi: Studi Kasus CV. Viradi Global Pratama

Harry Dhika

Teknik Informatika, Universitas Indraprasta PGRI  
Jl. Nangka No. 58 Tanjung Barat  
email: dhikatr@yahoo.com

**Abstrak** – Dalam penelitian ini dilakukan komparasi *Algoritma C4.5*, *Naïve Bayes* dan *Neural Network* yang diaplikasikan terhadap data perusahaan yang bermitra dengan CV. Viradi Global Pratama untuk mencari mitra perusahaan dengan prediksi *very good*, *good* dan *bad*. Dari hasil pengujian dengan mengukur kinerja ketiga algoritma tersebut menggunakan metode pengujian *Confusion Matrix* dan *Kurva ROC (Receiver Operating Characteristic)*, diketahui bahwa algoritma *C4.5* memiliki nilai *accuracy* dan *AUC (Area Under Curve)* paling tinggi, diikuti oleh metode *Naïve Bayes*, dan yang paling rendah adalah metode *Neural Network*

**Kata Kunci:** *Algoritma C4.5*, *naïve bayes*, *neural network*, *Pemilihan Mitra Kerja*

## I. PENDAHULUAN

Dewasa ini semakin banyak berkembang perusahaan yang bergerak pada bidang jasa pengiriman barang. CV. Viradi Global Pratama merupakan sebuah perusahaan yang bergerak dibidang jasa pengiriman barang. Perusahaan ini selalu melayani pelanggannya ketika pelanggan tersebut membutuhkan jasa untuk pengiriman barang ke daerah atau wilayah lain. Dalam melakukan pengiriman barang, CV. Viradi Global Pratama membutuhkan alat transportasi yang dapat menjangkau wilayah-wilayah pengiriman tujuan barang. Oleh karena itu, perusahaan ini melakukan kerja sama dengan perusahaan-perusahaan penyedia jasa transportasi baik darat, laut maupun udara. Karena pada tahun-tahun sebelumnya, perusahaan tersebut selalu menggunakan cara konvensional dalam menggunakan alat transportasi terhadap semua pemesanan jasa pengiriman barang dan selalu menimbulkan masalah yaitu kendala kecelakaan atau kurang layaknya alat transportasi yang dapat menyebabkan barang hilang bahkan rusak. Setelah melihat penerapan dan penggunaan perbandingan mitra penyedia jasa pengiriman barang yang ada saat ini, ternyata masih ditemukan beberapa masalah misalnya kerusakan dalam pengiriman juga masih ditemukan khususnya ketika menerima atau mengirim barang bahkan hilangnya barang saat dikirim. Meskipun penyedia jasa pengiriman barang tersebut sudah memiliki banyak pengalaman, tetapi masih saja ada beberapa penyedia jasa pengiriman barang yang tidak konsisten dalam melakukan tugasnya tersebut secara baik dan benar. Pada umumnya alasan buruknya pelayanan adalah kurangnya pengetahuan terhadap koordinasi dalam pengiriman barang. Melihat masalah-masalah yang ada maka dalam penelitian ini akan digunakan metode pengambilan keputusan dengan algoritma *C4.5*, *Naïve*

*Bayes*, dan *Neural Network* untuk menentukan pemilihan mitra kerja penyedia jasa transportasi jasa pengiriman barang yang baik pada CV. Viradi Global Pratama. Pemilihan penyedia jasa transportasi barang yang baik berdasarkan pengalaman kerja, harga jasa pengiriman, legalitas perusahaan mitra, jumlah armada yang dimiliki, dan intensitas pelayanan kerja mitra dalam melayani jasa pengiriman barang.

## II. LANDASAN TEORI

Setiap lembaga atau perusahaan pasti bekerja dengan data, perusahaan yang sudah cukup lama tentunya juga memiliki data yang banyak, data dalam perusahaan tersebut semakin lama akan semakin menggenung atau semakin menumpuk dan disimpan tiap harinya, tiap bulannya dan tiap tahunnya.

Tentunya data perusahaan tersebut sangat sayang sekali jika di buang, karena data tersebut merupakan aset berharga dari perusahaan apalagi data tersebut tidak digunakan sama sekali, untuk itu data yang jumlahnya sudah cukup banyak dan menggenung dapat digunakan kembali dengan cara menambangnya.

Data mining atau menambang data didefinisikan sebagai metode yang digunakan untuk mengekstraksi informasi prediktif tersembunyi pada database, ini adalah teknologi yang sangat potensial bagi perusahaan dalam memberdayakan data warehousenya (Sulianta, 2010). Data mining disebut sebagai proses ekstraksi pengetahuan dari data yang besar, sesuai fungsinya data *mining* adalah proses pengambilan keputusan dari volume data yang besar yang disimpan dalam basis data, data warehouse, atau informasi yang disimpan dalam repositori (Han, 2006). Istilah data mining berasal dari kemiripan antara pencarian informasi yang bernilai dari database yang besar dengan menambang sebuah gunung untuk

memperoleh sesuatu yang bernilai (Sumathi & Sivanandam, 2006). Data *mining* juga dapat didefinisikan sebagai sebuah proses penemuan pola dalam data (Witten, 2011). Keduanya memerlukan penyaringan melalui sejumlah besar material, atau menyelidiki dengan cerdas untuk mencari keberadaan sesuatu yang bernilai. Berdasarkan polanya data mining dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, clustering, asosiasi (Larose, 2005).

#### Algoritma C4.5

Algoritma C4.5 merupakan algoritma klasifikasi dengan teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain. Keakuratan prediksi yaitu kemampuan model untuk dapat memprediksi label kelas terhadap data baru atau yang belum diketahui sebelumnya dengan baik. Dalam hal kecepatan atau efisiensi waktu komputasi yang diperlukan untuk membuat dan menggunakan model. Kemampuan model untuk memprediksi dengan benar walaupun data ada nilai dari atribut yang hilang. Dan juga skalabilitas yaitu kemampuan untuk membangun model secara efisien untuk data berjumlah besar (aspek ini akan mendapatkan penekanan). Terakhir interpretabilitas yaitu model yang dihasilkan mudah dipahami. Selain itu dapat juga dikatakan bahwa Algoritma merupakan kumpulan perintah yang tertulis secara sistematis guna menyelesaikan permasalahan logika dari matematika. Pengertian Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Sedang pohon keputusan dapat diartikan suatu cara untuk memprediksi atau mengklarifikasi yang sangat kuat. Pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Kusrini, 2009), yaitu:

1. Menyiapkan data *training*. Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus:
- 3.

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

Keterangan:

S = himpunan kasus

n = jumlah partisi S

p<sub>i</sub> = proporsi S<sub>i</sub> terhadap S

4. Kemudian hitung nilai *gain* menggunakan rumus:

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

Keterangan:

S = himpunan kasus

A = fitur

n = jumlah partisi atribut A

|S<sub>i</sub>| = proporsi S<sub>i</sub> terhadap S

|S| = jumlah kasus dalam S

5. Ulangi langkah ke-2 hingga semua *record* terpartisi.
6. Proses partisi pohon keputusan akan berhenti saat:
  - a. Semua *record* dalam simpul N mendapat kelas yang sama.
  - b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
  - c. Tidak ada *record* di dalam cabang yang kosong.

#### Naïve Bayes

Mudah diinterpretasikan sehingga pengguna yang tidak punya keahlian dalam bidang teknologi klasifikasi pun bisa mengerti. Dalam titerington et al. (1981), model independence ini menghasilkan pemecahan yang terbaik. Efektifitas metode Naïve Bayes juga terlihat pada contoh dalam Hand dan Yu (2001) dan perbandingan empiris lebih jauh, dengan hasil yang sama, terdapat pada Domingos dan Pazzani (1997) (Wu, 2009). Klasifikasi Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas (Kusrini, 2009). Klasifikasi Bayes didasarkan pada teorema Bayes, diambil dari nama seorang ahli matematika yang juga menteri *Prebysterian* Inggris, *Thomas Bayes* (1702-1761), yaitu (Bramer, 2007):

$$P(x|y) = \frac{P(y|x) P(x)}{P(y)}$$

keterangan :

y = data dengan kelas yang belum diketahui

x = hipotesis data y merupakan suatu kelas spesifik

P(x | y) = probabilitas hipotesis x berdasar kondisi y (*posteriori probability*)

P(x) = probabilitas hipotesis x (*prior probability*)

P(y | x) = probabilitas y berdasarkan kondisi pada hipotesis x

P(y) = probabilitas dari y

#### Neural Network

Pengertian lain *Neural network* adalah satu set unit *input/output* yang terhubung dimana tiap relasinya

memiliki bobot (Han, 2006). *Neural Network* dimaksudkan untuk mensimulasikan perilaku sistem biologi susunan syaraf manusia, yang terdiri dari sejumlah besar unit pemroses yang disebut *neuron*, yang beroperasi secara paralel (Alpaydin, 2010). *Neuron* mempunyai relasi dengan *synapse* yang mengelilingi *neuron-neuron* lainnya. Susunan syaraf tersebut dipresentasikan dalam *neural network* berupa graf yang terdiri dari simpul (*neuron*) yang dihubungkan dengan busur, yang berkorespondensi dengan *synapse*. Sejak tahun 1950-an, *neural network* telah digunakan untuk tujuan prediksi, bukan hanya klasifikasi tapi juga untuk regresi dengan atribut target kontinu (Vecellis, 2009). Langkah pembelajaran dalam algoritma *backpropagation* adalah sebagai berikut (Myatt, 2007):

1. Inisialisasi bobot jaringan secara acak (biasanya antara -0.1 sampai 1.0)
2. Untuk setiap data pada data *training*, hitung input untuk simpul berdasarkan nilai input dan bobot jaringan saat itu, menggunakan rumus:

$$Input_j = \sum_{i=1}^n O_i w_{ij} + \Theta_j$$

Keterangan:

$O_i$  = Output simpul  $i$  dari layer sebelumnya

$w_{ij}$  = bobot relasi dari simpul  $i$  pada layer sebelumnya ke simpul  $j$

$\Theta_j$  = bias (sebagai pembatas)

3. Berdasarkan input dari langkah dua, selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktivasi sigmoid:

$$Output = \frac{1}{1 + e^{-Input}}$$

4. Hitung nilai *Error* antara nilai yang diprediksi dengan nilai yang sesungguhnya menggunakan rumus:

$$Error_j = Output_j \cdot (1 - Output_j) \cdot (Target_j - Output_j)$$

Keterangan:

$Output_j$  = Output aktual dari simpul  $j$

$Target_j$  = Nilai target yang sudah diketahui pada data *training*

5. Setelah nilai *Error* dihitung, selanjutnya dibalik ke layer sebelumnya (*backpropagated*). Untuk menghitung nilai *Error* pada *hidden layer*, menggunakan rumus:

$$Error_j = Output_j (1 - Output_j) \sum_{k=1}^n Error_k w_{jk}$$

Keterangan:

$Output_j$  = Output aktual dari simpul  $j$

$Error_k$  = error simpul  $k$

$w_{jk}$  = Bobot relasi dari simpul  $j$  ke simpul  $k$  pada layer berikutnya

6. Nilai *Error* yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi menggunakan rumus

$$w_{ij} = w_{ij} + l \cdot Error_j \cdot Output_i$$

Keterangan:

$w_{ij}$  = bobot relasi dari unit  $i$  pada layer sebelumnya ke unit  $j$

$l$  = *learning rate* (konstanta, nilainya antara 0 sampai dengan 1)

$Error_j$  = *Error* pada output layer simpul  $j$

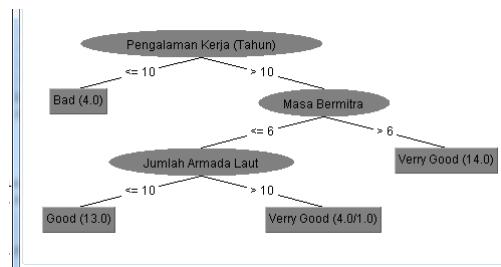
$Output_i$  = *Output* dari simpul  $i$

### III. PEMBAHASAN

Sample yang digunakan dalam mengolah sebanyak 35 perusahaan dan digunakan sebanyak 15 data perusahaan untuk menguji hasil data training.

#### Algoritma C4.5

Gambar 1 adalah pohon keputusan akhir yang dihasilkan dari perhitungan *entropy* dan *gain* untuk seluruh atribut. Terlihat atribut pengalaman kerja menjadi simpul akar karena mempunyai nilai *gain* yang paling besar. Dari simpul akar bercabang dua simpul sesuai dengan nilainya, yaitu pengalaman kerja  $\leq 10$  tahun dan  $> 10$  tahun. Kemudian pada cabang paling kanan, simpul 1.1 adalah Masa Bermitra, karena atribut tersebut mempunyai nilai *gain* tertinggi, di bawah simpul 1.1, yaitu simpul 1.1.1 merupakan atribut jumlah armada laut yang merupakan atribut dengan nilai *gain* tertinggi.



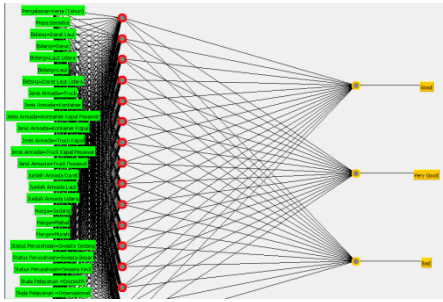
Gambar 1 Pohon Keputusan hasil perhitungan dengan metode C4.5

#### Naïve Bayes

Untuk menentukan kasus baru termasuk kelas mana, dilakukan perhitungan probabilitas *posterior* berdasarkan probabilitas *prior* yang telah dihitung sebelumnya. Misalkan diambil sebuah data testing  $X$ , untuk menentukan kelas mana, dilakukan perhitungan probabilitas *posterior*.

#### Neural Network

Gambar 2 adalah *neural net* yang dihasilkan dari pengolahan data *training* dengan metode *neural network* adalah *multilayer perceptron* yang dihasilkan dari data *training*.



Gambar 2 Neural net yang dihasilkan dengan metode neural network

Untuk setiap data pada data *training* dihitung input untuk simpul berdasarkan nilai input dan jaringan saat itu. Bobot awal untuk *input layer*, *hidden layer*, dan bias diinisialisasi secara acak. Simpul bias terdiri dari dua, yaitu pada *input layer* yang terhubung dengan simpul-simpul pada *hidden layer*, dan pada *hidden layer* yang terhubung pada *output layer*. Setelah semua nilai awal diinisialisasi, kemudian dihitung masukan, keluaran, dan *error*. Selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktivasi sigmoid. Setelah didapat nilai dari fungsi aktivasi, hitung nilai *error* antara nilai yang diprediksi dengan nilai yang sesungguhnya. Setelah nilai *error* dihitung, selanjutnya dibalik ke *layer* sebelumnya (*backpropagated*).

**Pengujian Model**

Model yang telah dibentuk diuji tingkat akurasi dengan memasukkan data uji yang berasal dari data *training*. Karena data yang didapat dalam penelitian ini setelah proses *preprocessing* hanya 50 data maka digunakan metode *full training set* untuk menguji tingkat akurasi. Untuk nilai akurasi model untuk metode C4.5 sebesar 97.1429%, metode *naïve bayes* sebesar 94.2857%, dan metode *neural network* sebesar 60%

1. Confusion Matrix

Tabel 1 adalah perhitungan berdasarkan data *training*, diketahui dari 35 data, 4 data diklasifikasikan *bad* sesuai dengan prediksi yang dilakukan dengan metode C4.5, lalu 13 data diprediksi *good* dan 18 data *very good*.

Tabel 1 Model Confusion Matrix untuk Metode C4.5

```

=== Confusion Matrix ===
  a  b  c  <-- classified as
17  0  0 | a = Verry Good
 1 13  0 | b = Good
 0  0  4 | c = Bad
    
```

Tabel 2 adalah *confusion matrix* untuk metode *naïve bayes*. Diketahui dari 35 data, 15 data diklasifikasikan *very good* sesuai dengan prediksi yang dilakukan dengan metode C4.5, lalu 16 data diprediksi *good*, 4 data diprediksi bad.

Tabel 2 Model Confusion Matrix untuk Metode Naïve Bayes

```

=== Confusion Matrix ===
  a  b  c  <-- classified as
15  2  0 | a = Verry Good
 0 14  0 | b = Good
 0  0  4 | c = Bad
    
```

Dengan metode *neural network*, menghasilkan kondisi seperti pada Tabel 3 Diketahui dari 35 data, 18 diklasifikasikan *Very Good* sesuai dengan prediksi yang dilakukan dengan metode C4.5, lalu 15 data diprediksi *good*, serta 2 data diprediksi *bad*.

Tabel 3 Model Confusion Matrix untuk Metode Neural Network

```

=== Confusion Matrix ===
  a  b  c  <-- classified as
12  4  1 | a = Verry Good
 6  8  0 | b = Good
 0  3  1 | c = Bad
    
```

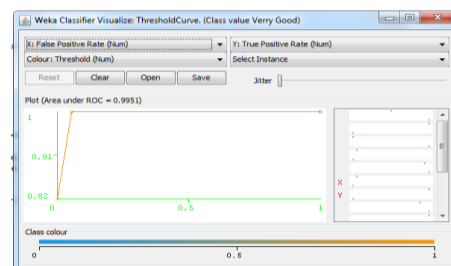
Dari tiga table *confusion matrix*, selanjutnya dilakukan perhitungan nilai *accuracy*, *precision*, *sensitivity*, dan *recall*. Perbandingan nilai *accuracy*, *precision*, *sensitivity*, dan *recall* yang telah dihitung untuk metode C4.5, *naïve bayes*, dan *neural network* dapat dilihat pada Tabel 4

Tabel 4 Komparasi Nilai Accuracy, Precision, dan Recall

	C4.5	Naïve Bayes	Neural network
Accuracy	97.1429%	94.2857%	60%
Precision	0.973%	0.95%	0.594%
Recall	0.971%	0.923%	0.6%

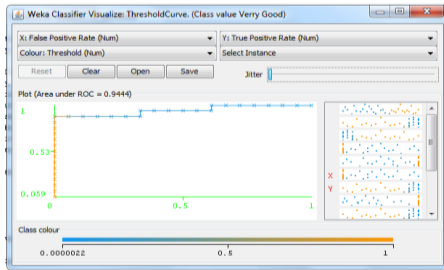
2. Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan ketiga metode komparasi bisa dilihat pada Gambar 3 yang merupakan kurva ROC untuk algoritma C4.5.



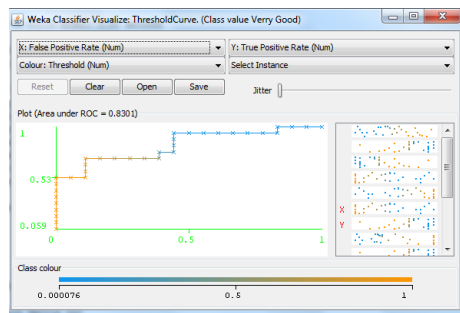
Gambar 3 Kurva ROC dengan algoritma C4.5

Kurva ROC pada gambar 3 mengekspresikan *confusion matrix* dari Tabel 1. Garis horizontal adalah *false positives* dan garis vertikal *true positives*.



Gambar 4 Kurva ROC dengan Metode *Naive Bayes*

Seperti terlihat pada Gambar 3, Gambar 4, dan Gambar 5, area di bawah kurva pada Gambar 3 paling luas diantara ketiga metode.



Gambar 5 Kurva ROC dengan Metode *Neural Network*

Pebandingan hasil perhitungan nilai AUC untuk metode *C4.5*, *naive bayes*, dan *neural network* dapat dilihat pada Tabel 5.

Tabel 5 Komparasi Nilai AUC

	C4.5	Naive Bayes	Neural network
AUC	0.9951	0,9444	0.8952

**Analisis Hasil Komparasi**

Model yang dihasilkan metode *C4.5*, *naive bayes*, dan *neural network* diuji menggunakan metode *Cross Validation*, terlihat perbandingan nilai *accuracy*, *precision*, *sensitivity*, dan *recall* pada Tabel 4, untuk metode *C4.5* memiliki nilai *accuracy*, *precision*, *sensitivity*, dan *recall* yang paling tinggi, diikuti dengan metode *naive bayes*, dan yang terendah adalah *neural network*.

Tabel 6 Komparasi Nilai *Accuracy* dan AUC

	C4.5	Naive Bayes	Neural network
Accuracy	97.1429%	94.2957%	60%
AUC	0.9951	0,9444	0.8952

Tabel 6 membandingkan *accuracy* dan AUC dari tiap metode. Terlihat bahwa nilai *accuracy* *C4.5* paling tinggi begitu pula dengan nilai AUC-nya. Untuk metode *neural network* dan *naive bayes* juga menunjukkan nilai yang sesuai. Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok (Gorunescu, 2011).

- a. 0.90-1.00 = klasifikasi sangat baik
- b. 0.80-0.90 = klasifikasi baik
- c. 0.70-0.80 = klasifikasi cukup
- d. 0.60-0.70 = klasifikasi buruk
- e. 0.50-0.60 = klasifikasi salah

Berdasarkan pengelompokan di atas dan Tabel 6 maka dapat disimpulkan bahwa metode *C4.5*, *naive bayes*, dan *neural network* termasuk klasifikasi sangat baik karena memiliki nilai AUC antara 0.90-1.00.

**Penerapan Algoritma Terpilih**

Dari algoritma terpilih yaitu algoritma *C4.5* terdapat data baru yang akan diinput yakni:

Tabel 7 Data Baru untuk Penerapan

Pengalam Masa Berr Bidang	Jenis Arm	Jumlah Ar	Jumlah Ar	Jumlah Ar	Harga	Status PerSkala	Pela	Jumlah Ca	Remark
27	8 Darat Laut	Kontainer	2	5	0 Mahal	Swasta Be	Internasio	<5	Verry Goo
23	8 Darat Laut	Kontainer	4	11	0 Mahal	Swasta Be	Domestik	<5	Verry Goo
12	2 Laut	Truck Kapi	15	12	0 Sedang	Swasta Se	Domestik	>5	Verry Goo
19	9 Darat Laut	Truck Kapi	18	5	0 Sedang	Swasta Se	Internasio	<5	Verry Goo
16	6 Darat	Truck	12	0	0 Sedang	Swasta Se	Domestik	>5	Good
19	10 Darat	Truck	13	0	0 Sedang	Swasta Se	Domestik	<5	Verry Goo
9	3 Darat	Truck	10	0	0 Murah	Swasta Ke	Domestik	<5	Bad
10	3 Darat	Truck	10	0	0 Sedang	Swasta Se	Domestik	<5	Bad
33	8 Darat Laut	Kontainer	4	5	0 Mahal	Swasta Be	Domestik	<5	Verry Goo
20	5 Darat Laut	Truck Kapi	9	12	4 Sedang	Swasta Se	Domestik	<5	Good

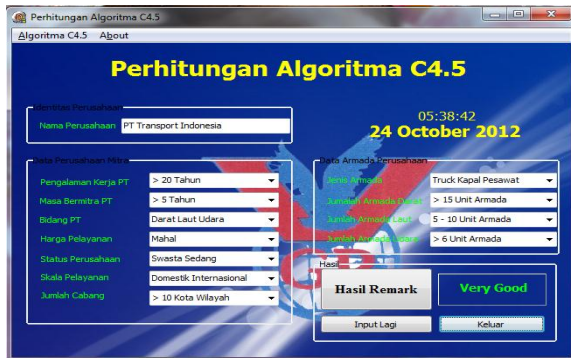
Dengan pengukuran tingkat akurasi sebesar 80% dan *confusion matrix* sebagai berikut:

```

=== Confusion Matrix ===
 a b c  <-- classified as
 6 0 0 | a = Verry Good
 2 0 0 | b = Good
 0 0 2 | c = Bad
    
```

Dari 10 data baru yang akan diinput pada tahapan *development* menggunakan aplikasi *delphi* dengan *interface* atau rancangan antar muka sebagai berikut:





Gambar 6 Perhitungan Algoritma Terpilih

Dengan menginput data perusahaan pada program tersebut sesuai dengan atribut yang dibutuhkan, kemudian klik tombol HASIL REMARK, maka secara otomatis tampil hasil klasifikasi perusahaan mitra *very good*. Untuk menginput kembali data baru klik tombol INPUT LAGI, dan tombol KELUAR digunakan untuk keluar dari aplikasi tersebut.

#### IV. KESIMPULAN

Untuk mengukur ketiga kinerja algoritma tersebut digunakan *Confusion Matrix* dan Kurva ROC dengan hasil bahwa algoritma yang memiliki nilai *accuracy* dan ROC paling tinggi adalah Algoritma C4.5 kemudian Naive Bayes dan terakhir diikuti oleh Neural Network. Dengan begitu dapat disimpulkan bahwa metode yang paling baik adalah *algoritma C4.5* dalam mengklasifikasikan data. Dengan demikian *algoritma C4.5* dapat memberikan solusi atau pemecahan masalah untuk permasalahan pemilihan mitra kerja penyedia jasa transportasi pada CV. Viradi Global Pratama.

#### DAFTAR REFERENSI

- [1] Contoh Kasus Prediksi layak tidaknya Indonesian menjadi Juara AFF, <http://dennistorio.com/DM/bola.php> diakses tanggal 29 juni 2012.
- [2] Defiyanti, Sofi, Crispina Pardede. Jurnal: Perbandingan Kinerja Algoritma ID3 Dan C4.5 Dalam Klasifikasi Spam-Mail, Gunadarma. Depok
- [3] Gorunescu, Florin, Data Mining: *Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer. 2011
- [4] Han, J.,& Kamber, M. Data Mining Concept and Tehniques.San Fransisco: Morgan Kauffman. 2006
- [5] Hanik, Umi. Jurnal: *Fuzzy Decision Tree* dengan Algoritma C4.5 pada Data Diabetes Indian Pima. 2011

#### Biodata Penulis

**Harry Dhika**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika Universitas Indraprasta PGRI Jakarta, lulus tahun 2009. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika STMIK Nusa Mandiri, lulus tahun 2012. Saat ini menjadi Dosen di UNIVERSITAS INDRAPRASTA PGRI Jakarta.