

SELEKSI MOBIL BERDASARKAN FITUR DENGAN KOMPARASI METODE KLASIFIKASI *NEURAL NETWORK*, *SUPPORT VECTOR MACHINE*, DAN ALGORITMA C4.5

Esty Purwaningsih

Program Studi Manajemen Informatika
AMIK BSI Tangerang

Jl. Letnan Sutopo BSD Serpong, Bumi Serpong Damai Sektor XIV Blok C1/1, Tangerang Selatan
esty.epw@bsi.ac.id

Abstract – *Car collection based features using the comparative method of Neural Networks, Support Vector Machine, and algorithm C4.5. Comparative method was chosen because there are advantages and privileges of each method, also because there are previous studies that have test vehicle classification using C4.5 Algorithm and Neural Network and make suggestions by using Support Vector Machine and there has been no research on classification of vehicles by using Support Vector Machine. The data used comes from PT. Tunas Mobilindo Perkasa. The data of this study is the data cars as well as car sales in PT. Tunas Mobilindo Perkasa with 2013. Untuk year period to apply the method Neural Networks, Support Vector Machine and C4.5 algorithms used software RapidMiner. Results of this application then uses compare Confusion Matrix and ROC curve. Based on this study proved that the method C4.5 algorithm has a higher accuracy level than the Neural Network and Support Vector Machine.*

Intisari – Seleksi mobil berdasarkan fitur menggunakan komparasi metode *Neural Network*, *Support Vector Machine*, dan Algoritma C4.5. Komparasi metode tersebut dipilih karena terdapat kelebihan dan keistimewaan dari masing-masing metode, juga karena terdapat penelitian terdahulu yang telah melakukan uji klasifikasi kendaraan dengan menggunakan metode Algoritma C4.5 dan *Neural Network*, dan memberikan usulan dengan menggunakan metode *Support Vector Machine* serta belum ada penelitian tentang klasifikasi kendaraan dengan menggunakan metode *Support Vector Machine*. Data yang digunakan bersumber dari PT. Tunas Mobilindo Perkasa. Data yang diteliti ini merupakan data mobil serta penjualan mobil di PT. Tunas Mobilindo Perkasa dengan periode tahun 2013. Untuk menerapkan metode *Neural Network*, *Support Vector Machine* dan Algoritma C4.5 ini digunakan perangkat lunak RapidMiner. Hasil penerapan ini kemudian dikomparasi menggunakan *Confusion Matrix* dan Kurva ROC. Berdasarkan penelitian ini terbukti bahwa

metode Algoritma C4.5 memiliki tingkat akurasi yang lebih tinggi dibandingkan *Neural Network* dan *Support Vector Machine*.

Kata Kunci : *Algoritma C4.5, Klasifikasi, Neural Network, Support Vector Machine.*

PENDAHULUAN

Meningkatnya pendapatan per kapita dan kurangnya ketersediaan armada kendaraan umum, membuat masyarakat kalangan menengah keatas lebih memilih memakai kendaraan mobil pribadi. Mobil merupakan alat transportasi yang banyak digemari karena efisiensinya memberikan keamanan dan kenyamanan bagi pemakainya. Jumlah mobil pribadi yang beredar pun semakin meningkat seiring bermunculannya merk dan tipe terbaru dari berbagai industri mobil, ditambah dengan semakin beredarnya iklan-iklan mobil, membuat sebagian konsumen tertarik dan terdorong untuk memilih mobil yang diminati (N.A Lazuardi dkk, 2013).

Pada dasarnya, mobil dapat dikategorikan dari berbagai jenis fitur yaitu mulai dari bentuk, ukuran, spesifikasi mekanik, roda, kinerja, sofa mobil dan sebagainya. Setiap individu dapat memilih fitur mobil dengan kepuasan tersendiri di semua bidang. Penelitian tentang klasifikasi kendaraan khususnya mobil telah dilakukan dengan menggunakan model *Artificial Neural Network Feedforward Multilayer* dengan algoritma *Backpropagation*. Dari hasil penelitian akhir yang diuji kesimpulannya adalah *Multilayer Perceptron* dengan *Algoritma Backpropagation* dapat digunakan sebagai penyelesaian untuk masalah klasifikasi dengan node penuh *training*, *learning* dan *transfer function* (Chauhan dan K.V Prema, 2012).

Pada klasifikasi kendaraan juga dilakukan Mishra dkk (2013), dengan menggunakan pendekatan *Probabilistik Neural Network* dan seleksi fitur. Tujuannya adalah untuk mengklasifikasikan kendaraan yang diminati

pelanggan dengan jumlah kendaraan berbeda yang tersedia di pasar berdasarkan survei pasar. Dari hasil penelitian akhir menyebutkan bahwa *decision tree* merupakan algoritma yang paling akurat, namun dalam uji T-Test tidak dominan terhadap algoritma lain.

Penelitian juga pernah dilakukan Wah, dkk (2013) dengan mengkomparasi beberapa algoritma klasifikasi data mining yaitu *Logistic Regression* (LR), *Decision Tree* (DT) dan *Neural Network* (NN). Hasil menunjukkan tidak ada model mengungguli karakteristik lain. Metode yang diusulkan untuk menutupi teknik klasifikasi lainnya adalah SVM.

Dari metode-metode yang pernah diteliti berkaitan dengan klasifikasi seleksi mobil maka pada penelitian ini digunakan metode dengan mengkomparasi beberapa metode yaitu *Neural Network*, *Support Vector Machine* dan Algoritma C4.5 dimana masing-masing metode mempunyai kelebihan.

Neural Network memiliki kemampuan sangat mudah diterima untuk data *noisy* dan akurasi tinggi serta lebih baik dalam data mining namun *Neural Network* mempunyai kelemahan dimana menggunakan data pelatihan cukup besar.

Kelebihan dari *Support Vector Machine* (SVM) adalah cukup populer dan baik untuk penggunaan klasifikasi karena tidak tergantung pada jumlah fitur dan bisa mengatasi masalah dimensi. Dari segi komputasi, SVM dapat melakukan proses *training* dengan cepat dan ini berguna dalam teknik *learning* ketika menghadapi masalah ketidaktegasan (Maimon, 2010).

Sedangkan kelebihan dari algoritma C4.5 atau dikenal dengan nama *Decision Tree* (Gorunescu, 2011) yaitu hasil analisa berupa diagram pohon yang sangat mudah dimengerti, mudah untuk dibangun, serta membutuhkan data percobaan yang lebih sedikit dibandingkan algoritma klasifikasi lainnya, mampu mengolah data nominal dan kontinyu, namun kelemahan dari algoritma Algoritma C4.5 ini adalah tidak bisa menggunakan sampel yang lebih besar.

Dari kelebihan dan kelemahan yang dimiliki oleh ketiga metode tersebut, maka pada penelitian ini dilakukan dengan membandingkan ketiga metode tersebut.

Tujuan dari penelitian ini adalah memperoleh model *Neural Network*, *Support Vector Machine* dan Algoritma C4.5 yang memiliki akurasi yang paling tinggi dan tepat untuk klasifikasi seleksi mobil berdasarkan fitur. Serta memiliki manfaat baik praktis maupun teoritis. Dimana penelitian ini memberikan manfaat praktis yaitu memberikan kontribusi keilmuan pada penelitian bidang klasifikasi data mining khususnya untuk klasifikasi seleksi mobil

berdasarkan fitur. Sedangkan manfaat teoritis yaitu dengan memberikan bukti *empiris* mengenai penerapan model *Neural Network*, *Support Vector Machine* dan Algoritma C4.5 pada klasifikasi seleksi mobil berdasarkan fitur di PT. Tunas Mobilindo Perkasa.

Masalah yang diangkat dalam penelitian ini dibatasi pada penerapan model *Neural Network*, *Support Vector Machine* dan Algoritma C4.5, yang akan ditetapkan untuk klasifikasi seleksi mobil berdasarkan fitur. Penelitian ini hanya berfokus pada mobil dengan merk Daihatsu di PT. Tunas Mobilindo Perkasa dengan acuan tahun amatan 2013.

Untuk menerapkan model *Neural Network*, *Support Vector Machine* dan Algoritma C4.5, maka digunakan perangkat lunak (*software*) yaitu RapidMiner dimana *software* tersebut memiliki sistem yang komprehensif untuk analisa data serta banyak digunakan karena kemampuan, fleksibilitas dan kemudahan dalam penggunaannya.

Hasil dari penerapan menggunakan model *Neural Network*, *Support Vector Machine* dan Algoritma C4.5, ini kemudian akan dikomparasi tingkat akurasi menggunakan *Confusion matrix* dan kurva ROC (*Receiver Operating Characteristic*) untuk mengetahui model yang memiliki tingkat akurasi yang paling tinggi, sehingga tujuan penerapan model tersebut untuk seleksi mobil berdasarkan fitur ini dapat tercapai.

BAHAN DAN METODE

Menurut (Dawson, 2009) ada empat metode penelitian yang umum digunakan yaitu tindakan penelitian, eksperimen, studi kasus dan survey. Dalam konteks penelitian, metode yang dilakukan mengacu kepada pemecahan masalah yang meliputi mengumpulkan data, merumuskan hipotesis atau proposisi, pengujian hipotesis, menafsirkan hasil, dan kesimpulan (Berndtsson, 2008).

Jenis penelitian yang dilakukan dalam penelitian ini adalah jenis penelitian eksperimen. Jenis penelitian eksperimen ini menguji kebenaran sebuah hipotesis dengan statistik dan menghubungkannya dengan masalah penelitian (Kothari, 2004).

Jenis penelitian eksperimen dibagi dua, yaitu eksperimen absolut dan eksperimen komparatif. Eksperimen absolut mengarah kepada dampak yang dihasilkan dari eksperimen, misalnya pengaruh honor dosen terhadap kinerja. Sedangkan eksperimen komparatif yaitu membandingkan dua objek yang berbeda, misalnya membandingkan dua algoritma yang berbeda dengan melihat hasil statistik masing-masing mana yang lebih baik (Kothari, 2004).

Pada penelitian ini, jenis penelitian yang diambil adalah Eksperimen Komparatif.

Penelitian ini dirancang dengan melakukan perbandingan atau komparasi dan evaluasi pada metode *Neural Network*, *Support Vector Machines* dan Algoritma C4.5 untuk mengetahui model manakah yang memiliki keakuratan lebih tinggi dalam seleksi mobil berdasarkan fitur. Penelitian ini menggunakan data mobil dan data penjualan mobil yang bersumber dari PT. Tunas Mobilindo Perkasa.

Metode pengumpulan data untuk mendapatkan sumber data dibagi menjadi dua, yaitu data primer dan data sekunder. Data primer yaitu data yang dikumpulkan pertama kali, dan untuk melihat apa yang sesungguhnya terjadi melalui observasi, interview, kuesioner, dll. Data sekunder adalah data yang telah dikumpulkan dan dianalisis oleh orang lain baik yang telah dipublikasikan maupun yang belum dipublikasikan, misalnya dari dokumentasi, literatur, buku, jurnal, dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti.

Teknik pengumpulan data yang digunakan pada penelitian ini adalah dengan menggunakan data primer. Data utama diperoleh dari data penjualan mobil Daihatsu dengan tahun amatan 2013 di PT. Tunas Mobilindo Perkasa sedangkan data pendukung didapatkan dari buku, jurnal dan publikasi lainnya.

Dalam melakukan eksperimen ini, penelitian menggunakan model *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang terdiri dari enam tahap, yaitu (Chapman, 2000):

- a. Tahap *Business understanding*
Dataset yang didapat berasal dari PT. Tunas Mobilindo Perkasa. Banyak penelitian yang telah melakukan klasifikasi kendaraan mobil dengan model *Neural Network* dan Algoritma C4.5 namun belum ada yang membandingkan metode *Neural Network*, *Support Vector Machine* dan Algoritma C4.5 dalam seleksi mobil berdasarkan fitur untuk mengetahui model mana yang paling akurat dimana masing-masing metode memiliki kelebihan.
Pengolahan data awal perlu dilakukan sebelum data digunakan untuk membandingkan metode *Neural Network*, *Support Vector Machine* dan Algoritma C4.5. Pengolahan awal data berupa pembentukan sumber data *random* (*set the random seed*) dan pembentukan variabel pemilihan (*partition variabel*).
- b. Tahap *data understanding*
Dataset diperoleh dari PT. Tunas Mobilindo Perkasa yang terdiri dari 719 record dengan 17 variabel bebas, diantaranya: nama cust,

gender, umur, *chasis number*, *engine number*, merk, tipe, mesin cc, sistem kemudi, kapasitas, rasio gigi akhir (rga), *anti-lock braking system* (abs), rem, bahan bakar, *air conditioner* (ac), *airbag* dan harga.

- c. Tahap *data preparation*

Data yang diperoleh sebanyak 17 variabel namun terdapat beberapa variabel yang tidak digunakan karena variabel tersebut tidak relevan untuk penelitian ini, kemudian dilakukan proses seleksi pemilihan variabel yang digunakan berdasarkan uji kesepakatan pada 30 responden yang kebanyakan pernah bertransaksi kendaraan mobil. Dari pemilihan variabel dengan uji kesepakatan maka dinyatakan bahwa variabel yang digunakan pada penelitian ini sebanyak 9 variabel yang terdiri dari 8 atribut *predictor* dan 1 atribut tujuan. Atribut *predictor* yang digunakan adalah merk, mesin cc, kapasitas, sistem kemudi, *anti-lock braking system* (abs), *air conditioner* (ac), *airbag* dan harga. Sedangkan atribut tujuan adalah sebagai *output* yang digunakan untuk membedakan hasil seleksi mobil berdasarkan fitur, yaitu: "Diminati" dan "Kurang" diminati. Untuk mendapatkan data yang valid dan berkualitas maka dilakukan uji data *validation* untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*).

- d. Tahap *modeling*

Menggunakan model pada data *training* yang sudah diseleksi variabelnya. Kemudian, data *training* tersebut diuji dengan model yang digunakan yaitu Algoritma C4.5, *Neural Network* dan *Support Vector Machine* akan menghasilkan akurasi dalam proses klasifikasi seleksi mobil berdasarkan fitur. Pengujian yang dilakukan terhadap data *training* dengan metode tersebut dilakukan dengan teknik *10 folds cross validation*. Hal ini dilakukan dengan tujuan menghasilkan akurasi yang paling tinggi pada hasil uji-t metode tersebut. Hasil akurasi tersebut dapat digunakan sebagai nilai yang paling baik setelah teknik *10 folds cross validation*. Semua *dataset* diuji dengan metode/model yang diusulkan pada aplikasi RapidMiner. Eksperimen dilakukan satu persatu terhadap *dataset* dan model yang digunakan.

- e. Tahap *evaluation*

Setelah tahap *modelling* dilakukan, hasil dari komparasi metode *Neural Network*, *Support Vector Machine* dan Algoritma C4.5 ini kemudian akan dikomparasi tingkat akurasinya menggunakan *Confusion matrix*

dan nilai AUC (*Area Under ROC Curve*) untuk mengetahui model/metode yang memiliki tingkat akurasi yang paling tinggi, sehingga tujuan penerapan metode *Neural Network*, *Support Vector Machines* dan Algoritma C4.5, untuk mengklasifikasikan seleksi mobil berdasarkan fitur dapat tercapai.

Dari hasil akurasi yang paling tinggi menandakan bahwa model tersebut yang diuji menghasilkan model baru yang paling baik. Model tersebut yang dianggap sementara menjadi model yang paling baik untuk menghasilkan klasifikasi dalam seleksi mobil berdasarkan fitur. Proses selanjutnya adalah proses pengujian model tersebut terhadap data *training* yang menggunakan variabel yang sudah diseleksi. Hasil dari proses ini akan menghasilkan sebuah nilai AUC (*Area Under ROC Curve*) dan angka yang berisi *False Positif*, *False Negatif*, *True Positif*, dan *True Negatif*. Hasil tersebut dikenal dengan *confusion matrix* dimana akan terdapat tabel dengan keadaan klasifikasi seleksi mobil berdasarkan fitur tersebut untuk mendapatkan persentase ketepatan prediksi/akurasi.

Setelah diperoleh nilai AUC, nilai tersebut digunakan sebagai penentuan metode yang menghasilkan akurasi yang paling tinggi. Semua nilai AUC pada *dataset* akan dibandingkan. Hasil AUC akan dimasukkan ke dalam sebuah tabel dan kemudian akan disimpulkan metode mana yang memiliki akurasi paling tinggi dalam klasifikasi seleksi mobil berdasarkan fitur.

Sedangkan nilai *confusion matrix* akan digunakan untuk menghitung performa prediksi. Performa prediksi seleksi mobil berdasarkan fitur menggunakan perhitungan nilai *precision*. *Precision* digunakan untuk mendapatkan persentase nilai *True* (terdapat diminati) terhadap nilai *True* dan nilai yang prediksi *True* tapi kenyataannya *False*.

Pada *dataset* yang memiliki kesalahan prediksi yang sedikit akan menghasilkan nilai *precision* yang besar. Setelah diperoleh semua nilai performa prediksi pada semua *dataset*, kemudian nilai-nilai tersebut dibandingkan. Nilai performa prediksi yang paling tinggi dianggap sebagai yang paling baik. Kesimpulan dari nilai performa prediksi yang paling tinggi menandakan bahwa model tersebut merupakan model yang paling cocok untuk klasifikasi seleksi mobil berdasarkan fitur.

f. Tahap *deployment*

Pada tahap ini diterapkan model yang memiliki akurasi yang paling baik pada PT.

Tunas Mobilindo Perkasa atau instansi yang relevan untuk mengklasifikasi seleksi mobil berdasarkan fitur dengan menggunakan data baru. Untuk memastikan bahwa perangkat lunak yang dibuat memiliki standar minimal kualitas, maka salah satu metode untuk pengukuran kualitas perangkat lunak secara kuantitatif adalah metode SQA (*Software Quality Assurance*).

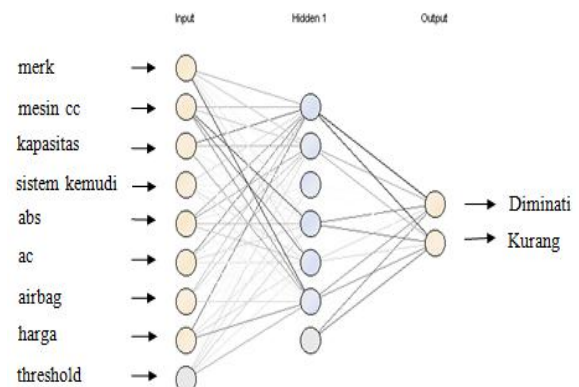
HASIL DAN PEMBAHASAN

Tujuan dari penelitian ini adalah untuk mengetahui metode *Neural Network*, *Support Vector Machine* dan Algoritma C4.5 yang memiliki akurasi klasifikasi seleksi mobil berdasarkan fitur yang paling tinggi. Data yang dianalisa adalah data hasil penjualan mobil dengan menggunakan metode *Neural Network*, *Support Vector Machine* dan Algoritma C4.5. Hasil dari analisa tersebut kemudian akan dibandingkan untuk mendapatkan model terpilih sesuai dengan kriteria pemilihan model terbaik, yaitu metode atau model yang memiliki akurasi paling tinggi.

a. *Neural Network*

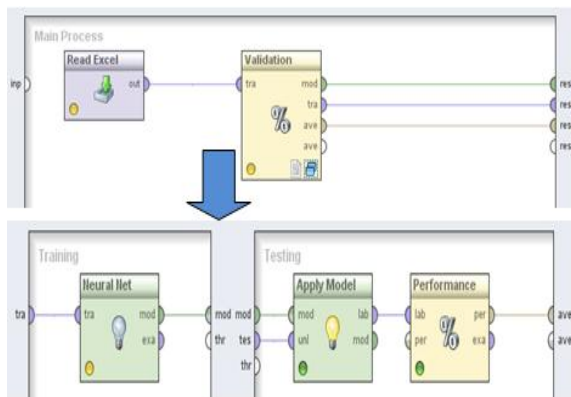
Neural net yang dihasilkan dari pengolahan data *training* seleksi mobil berdasarkan fitur dengan metode *neural network* adalah *multilayer perceptron*, sampel data *training* yang digunakan adalah data penjualan mobil pada PT. Tunas Mobilindo Perkasa dengan tahun amatan 2013.

Neural net yang dihasilkan pada gambar 1 dengan menggunakan tiga *layer* yang terdiri *input layer* terdiri dari delapan simpul, yaitu sama dengan jumlah atribut *predictor* yang digunakan yaitu merk, mesin_cc, kapasitas, sistem_kemudi, abs, ac, airbag, harga ditambah satu simpul bias. *Layer* yang kedua adalah *hidden layer* yang terdiri enam simpul dan satu simpul bias. *Layer* yang ketiga adalah *output layer* terdapat dua simpul yakni mewakili atribut kelas diminati dan kurang. Gambar 1 merupakan hasil percobaan *neural network* dengan satu *hidden layer* yang terdiri dari enam simpul.



Gambar 1. *Neural Net* Enam Simpul Hidden Layer
Sumber: Hasil Olah Data (2014)

Hasil analisis pengujian data *training* dengan model *neural network*, diketahui tingkat akurasi 82,11%, dari 710 data sebanyak 460 data diprediksi dengan tepat yaitu masuk dalam klasifikasi diminati, dan sebanyak 66 data diprediksi diminati tetapi ternyata masuk kedalam klasifikasi kurang, 61 data diprediksi kurang namun masuk kedalam klasifikasi diminati, 123 data diprediksi tepat yaitu masuk kedalam klasifikasi kurang seperti tabel *confusion matrix* yang disajikan pada Tabel 1. Hasil yang didapat dari pengolahan ROC untuk model *Neural Network* dengan menggunakan data *training* sebesar 0,884 dapat dilihat pada gambar 7.



Sumber: Uji Data (2014)

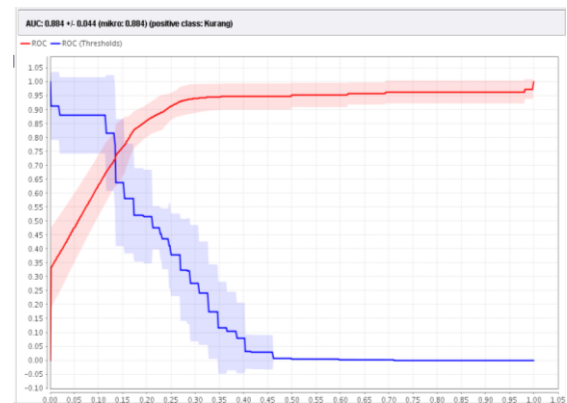
Gambar 2. Pengujian Model *Neural Network*

Tabel 1. Nilai Akurasi Data *Training* Model Algoritma C4.5

accuracy: 82.11% +/- 3.83% (mikroc: 82.11%)			
	true Diminati	true Kurang	class precision
pred. Diminati	460	66	87.45%
pred. Kurang	61	123	66.85%
class recall	88.29%	65.00%	

Sumber: Hasil Olah Data (2014)

$$\begin{aligned}
 \text{Akurasi} &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \\
 &= \frac{(123 + 460)}{(123 + 66 + 460 + 61)} \\
 &= 0,82112 \\
 &= 0,8211
 \end{aligned}$$



Gambar 3. Kurva ROC *Neural Network*

Sumber: Hasil Olah Data (2014)

b. *Support Vector Machine*

Nilai *training cycles* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C dan epsilon. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai *training cycles*:

Tabel 2. Eksperimen nilai *training cycle* SVM

C	Epsilon	SVM	
		Accuracy	AUC
0.0	0.0	76,20	0,79
1.0	1.0	73,88	0,50
1.0	0.0	74,79	0,74

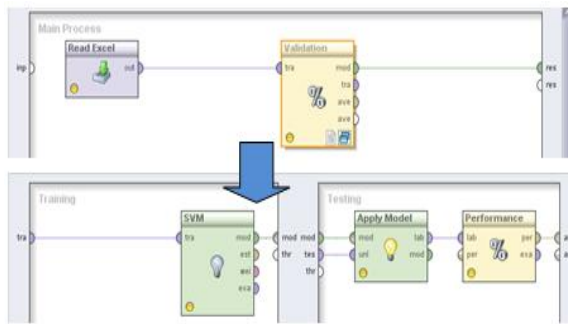
Sumber: Hasil Olah Data (2014)

Tabel 3. Kernel Model pada data *training* SVM

Atribute	Merk
Merk	0.467
Mesin cc	-1.153
Kapasitas	0.241
Sistem Kemudi	0.013
ABS	0.326
AC	0.187
Airbag	-0.537
Harga	0.805

Sumber: Hasil Olah Data (2014)

Pengujian data *training* dengan model *Support Vector Machine* diperoleh *accuracy* 76,20%, dari 710 data sebanyak 419 data diprediksi dengan tepat yaitu masuk dalam klasifikasi diminati, dan sebanyak 67 data diprediksi diminati tetapi ternyata masuk kedalam klasifikasi kurang, 102 data diprediksi kurang namun masuk kedalam klasifikasi diminati, 122 data diprediksi tepat yaitu masuk kedalam klasifikasi kurang seperti tabel *confusion matrix* yang disajikan pada Tabel.4. Hasil yang didapat dari pengolahan ROC untuk model *Support Vector Machine* dengan menggunakan data *training* sebesar 0,793 dapat dilihat pada gambar 5.



Sumber: Uji Data (2014)

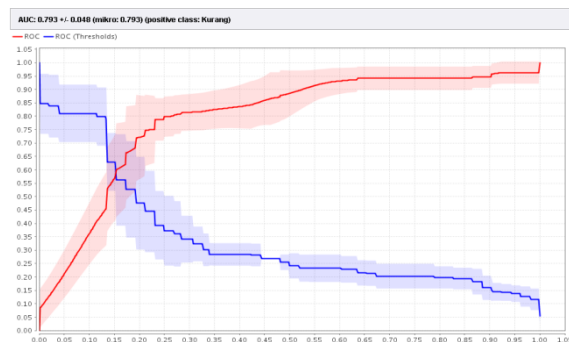
Gambar 4. Pengujian Model SVM

Tabel 4 Nilai Akurasi Data Training Model SVM

accuracy: 76.20% +/- 4.29% (mikroc: 76.20%)			
	true Diminati	true Kurang	class precision
pred. Diminati	419	67	86.21%
pred. Kurang	102	122	54.46%
class recall	80.42%	64.55%	

Sumber: Hasil Olah Data (2014)

$$\begin{aligned}
 \text{Akurasi} &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \\
 &= \frac{(122 + 419)}{(122 + 67 + 419 + 102)} \\
 &= 0,76197 \\
 &= 0,7620
 \end{aligned}$$



Sumber: Hasil Olah Data (2014)

Gambar 5. Kurva ROC SVM

c. Algoritma C4.5

Data *training* yang digunakan adalah untuk menentukan apakah fitur mobil yang ada pada mobil diminati atau kurang diminati oleh *customer*. Berikut akan dibahas langkah-langkah perhitungan klasifikasi seleksi mobil berdasarkan fitur yang diminati atau kurang diminati dengan menggunakan algoritma C4.5. Berikut langkah dalam pembuatan pohon keputusan, yaitu:

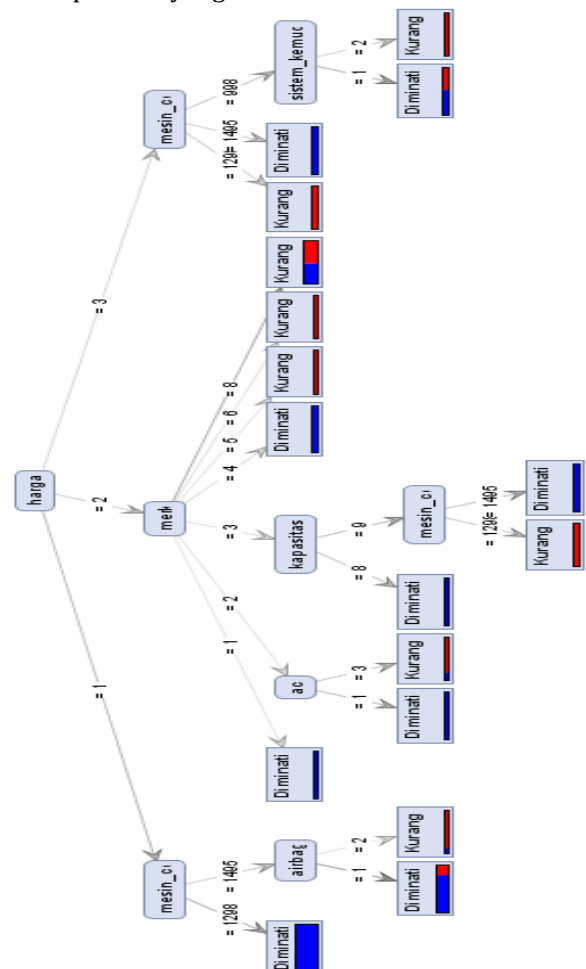
1) Menyiapkan data *training*

2) Hitung nilai *entropy* keseluruhan total kasus "Diminati" dan "Kurang" pada seleksi mobil berdasarkan fitur. Dari data *training* yang ada diketahui jumlah kasus yang fiturnya "Diminati" sebanyak 521 *record*, dan jumlah kasus yang fiturnya "Kurang" diminati adalah sebanyak 189 *record* total kasus keseluruhan adalah 710 kasus.

3) Hitung nilai entropi dan nilai *gain* masing-masing atribut. Nilai *gain* tertinggi adalah atribut yang menjadi *root* dari pohon keputusan yang akan dibuat. Misalkan menghitung entropi bagi atribut merk.

$$\begin{aligned}
 \text{Entropy}(S) &= \sum_{i=1}^n -p_i * \log_2 p_i \\
 &= -521/710 * \log_2 521/710 + (-189/710 * \\
 &\quad \log_2 189/710) \\
 &= 0,836
 \end{aligned}$$

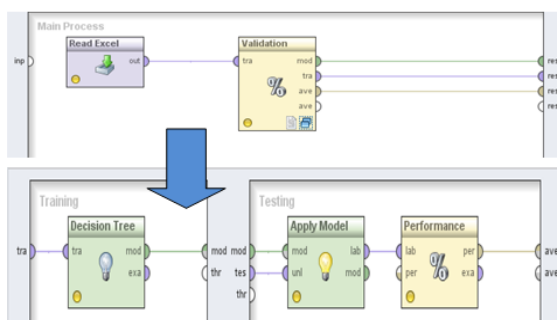
4) Hitung nilai entropi dan nilai *gain* masing-masing atribut. Nilai *gain* tertinggi adalah atribut yang menjadi *root* dari pohon keputusan yang akan dibuat.



Sumber: Hasil Olah Data (2014)

Gambar 6. Pohon Keputusan Data Seleksi mobil

Dari hasil analisis dengan menggunakan *tools* RapidMiner, diketahui tingkat akurasi 82,96%, dari 710 data sebanyak 440 data diprediksi dengan tepat yaitu masuk dalam klasifikasi diminati, dan sebanyak 40 data diprediksi diminati tetapi ternyata masuk kedalam klasifikasi kurang, 81 data diprediksi kurang namun masuk kedalam klasifikasi diminati, 149 data diprediksi tepat yaitu masuk kedalam klasifikasi kurang seperti tabel *confusion matrix* yang disajikan pada Tabel 5 Hasil yang didapat dari pengolahan ROC untuk algoritma C4.5 dengan menggunakan data *training* sebesar 0,888 dimana garis horizontal adalah *false* positif dan garis vertikal *true* positif, dapat dilihat pada gambar 8.



Sumber: Uji Data (2014)

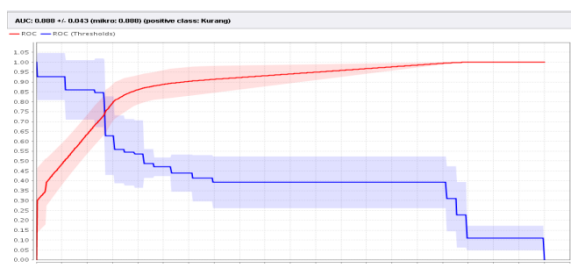
Gambar 7. Pengujian Model Algoritma C4.5

Tabel 5. Nilai Akurasi Data Training Model Algoritma C4.5

accuracy: 82.96% +/- 3.18% (mikro: 82.96%)			
	true Diminati	true Kurang	class precision
pred. Diminati	440	40	91.67%
pred. Kurang	81	149	64.78%
class recall	84.45%	78.84%	

Sumber: Hasil Olah Data (2014)

$$\begin{aligned}
 \text{Akurasi} &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \\
 &= \frac{(149 + 440)}{(149 + 40 + 440 + 81)} \\
 &= 0,82957 = 0,8296
 \end{aligned}$$



Sumber: Hasil Olah Data (2014)

Gambar 8. Kurva ROC Algoritma C4.5

Berdasarkan dari analisis perbandingan model, maka akurasi masing-masing model dapat dilihat pada tabel 6.

Tabel 6. *Confussion Matrix* dan AUC

Model	<i>Confussion Matrix</i>	AUC
Neural Network	82,11%	0,884
Support Vector Machine	76,20%	0,793
Algoritma C4.5	82,96%	0,888

Sumber: Hasil Olah Data (2014)

Dari table hasil komparasi di atas membandingkan tingkat akurasi *Confussion Matrix* dan AUC untuk model Algoritma C4.5, Neural Network dan Support Vector Machine, terlihat bahwa model Algoritma C4.5 memiliki nilai akurasi lebih tinggi yaitu 82,96% dibandingkan model Neural Network sebesar 82,11% dan Support Vector Machine sebesar 76,20%.

Dari hasil perhitungan yang tertera pada Tabel 6 diatas dengan menerapkan klasifikasi *performance* keakurasian AUC maka diperoleh hasil penelitian yaitu, terdapat dua metode yang merupakan kategori *Good Classification* yaitu untuk model Algoritma C4.5 dengan nilai AUC 0,888 dan model Neural Network dengan nilai AUC 0,884 sedangkan Support Vector Machine termasuk kategori *Fair Classification* dengan nilai AUC 0,793.

Rule hasil klasifikasi dari Algoritma C4.5 diterapkan dalam pembuatan aplikasi untuk klasifikasi seleksi mobil berdasarkan fitur agar dapat memudahkan pekerja di PT. Tunas Mobilindo Perkasa dalam membantu customer pada seleksi mobil berdasarkan fitur dan dapat membantu perusahaan dalam memutuskan untuk memproduksi fitur-fitur mana yang banyak diminati oleh customer.

Gambar 9. GUI Seleksi mobil berdasarkan fitur

Sumber: Uji Aplikasi (2014)

Pada gambar 9 menunjukkan program GUI seleksi mobil berdasarkan fitur dimana program tersebut akan menampilkan *output* berupa keterangan apakah masuk kedalam kategori diminati oleh *customer* atau kurang diminati. Program GUI seleksi mobil berdasarkan fitur ini dibangun dengan menggunakan *Microsoft Visual Basic*.

KESIMPULAN

Penelitian ini dibuat untuk mengklasifikasi seleksi mobil berdasarkan fitur dengan menggunakan perbandingan metode yaitu *Neural Network*, *Support Vector Machine* dan Algoritma C4.5, agar dapat diketahui metode atau model mana yang memiliki akurasi paling tinggi.

Untuk mengukur tingkat akurasi ketiga metode atau model tersebut digunakan metode pengujian *Confusion Matrix* dan *Kurva ROC*. Berdasarkan pengukuran tingkat akurasi ketiga metode atau model tersebut, diketahui bahwa model Algoritma C4.5 memiliki nilai akurasi paling tinggi yaitu 82,96% dibandingkan model *Neural Network* sebesar 82,11% dan *Support Vector Machine* sebesar 76,20%.

Dan dari hasil perhitungan, *performance* keakurasian AUC diperoleh hasil penelitian yaitu, terdapat dua metode yang masuk kedalam kategori *Good Classification* yaitu untuk model Algoritma C4.5 dengan nilai AUC 0,888 dan model *Neural Network* dengan nilai AUC 0,884 sedangkan *Support Vector Machine* termasuk kategori *Fair Classification* dengan nilai AUC 0,793. Sehingga dapat ditarik kesimpulan bahwa metode Algoritma C4.5 memiliki tingkat akurasi yang paling tinggi dibandingkan dengan *Neural Network* dan *Support Vector Machine*. Dengan demikian, metode Algoritma C4.5 dapat digunakan untuk klasifikasi seleksi mobil berdasarkan fitur.

REFERENSI

- Chapman, Peter. *CRISP-DM 1.0 Step by step Data Mining Guide*. SPSS Inc, 2000.
- Chauhan, Sucheta dan K. V. Prema. *Car Classification Using Artificial Neural Network*. 2012.
- Dawson, C. W. *Projects In Computing And Information System A Student's Guide*. England: Addison-Wesley, 2009.
- Gorunescu, F. *Data Mining Concepts, Models and Techniques*. Berlin Heidelberg: Springer Verlag, 2011.
- Kothari, C. R.. *Research Methology Methods and Techniques*. India: New Age International Limited, 2004.

Maimon, Oded&Rokach, Lior. *Data Mining and Knowledge Discovey Handbook*. New York:Springer, 2010.

Mishra, dkk. *A Probabilistic Neural Network Approach For Classification Of Vehicle*. 2013.

N.A, Lazuardi dkk. *Evaluasi Kelayakan Mobil Menggunakan K-Nearest Neighbour*. 2013.

Wah, dkk. *Predicting Car Purchase Intent Using Data Mining Approach*. 2011.

BIODATA PENULIS



Esty Purwaningsih, M.Kom.

Lahir di Jakarta, pada tanggal 28 Oktober 1987. Riwayat pendidikan: Diploma III AMIK BSI Jakarta tahun lulus 2009, Strata I STMIK Nusa Mandiri Jakarta lulus tahun 2011 dan Strata II Pascasarjana STMIK Nusa Mandiri Jakarta lulus tahun 2014. Bekerja sebagai Dosen Bina Sarana Informatika.