

**PREDIKSI PRODUK LARIS MOBIL HONDA DENGAN METODE KLASIFIKASI
MENGUNAKAN ALGORITMA C4.5 “(STUDI KASUS : DATA PENJUALAN SALES PT
PROSPECT MOTOR, CIKARANG)”**

Aswan S Sunge¹⁾, Heri Fidiawan²⁾

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa
aswan.sunge@pelitabangsa.ac.id

Disetujui, 27 Juni 2019

Abstraksi

Banyaknya persaingan dalam dunia bisnis khususnya dalam industri penjualan mobil menuntut para pengembang untuk menemukan suatu strategi yang dapat meningkatkan penjualan dan pemasaran produk yang dijual. Perusahaan harus memperhatikan jenis transaksi penjualan produk baik produk baru maupun produk lama yang di pasarkan dengan berbagai cara sehingga dapat meningkatkan efektifitas kinerja perusahaan dalam mengolah data transaksi penjualan. Mengetahui hasil prediksi dengan melihat akurasi algoritma C4.5 agar penjualan mobil honda dapat memperoleh target sesuai planning yang telah di tentukan perusahaan. Data sekunder yang digunakan pada penelitian ini berupa data penjualan sales executive PT Prospect Motor Cikarang. Pembentukan model prediksi menggunakan metode C4.5. pada algoritma C4.5 dilakukan perhitungan entropy dan information gain dimana atribut laris sebagai atribut tujuan, sedangkan kelas, model, transmisi, pendapatan, leasing, tenor dan diskon sebagai atribut sumber untuk memperoleh node akar dan node lainnya. Berdasarkan hasil klasifikasi menggunakan algoritma C4.5 menunjukkan bahwa diperoleh akurasi mencapai 67.5%, yang menunjukkan bahwa algoritma C4.5 cocok digunakan untuk mengukur prediksi penjualan produk laris mobil honda.

Kata Kunci : *Data Mining, Penjualan Produk, Decision Tree, Algoritma C4.5.*

Abstract

The amount of competition in the business world, especially in the car sales industry requires developers to find a strategy that can increase sales and marketing of products sold. Companies must pay attention to the type of product sales transactions both new products and old products which are marketed in various ways so as to improve the effectiveness of the company's performance in processing sales transaction data. Knowing the prediction results by looking at the accuracy of the C4.5 algorithm so that Honda car sales can obtain targets according to the planning that has been determined by the company. Secondary data used in this study are sales data of PT Prospect Motor Cikarang's sales executive. Forming a prediction model using the C4.5 method. In C4.5 algorithm, entropy and information gain calculations are performed where the best-selling attribute is the destination attribute, while the class, model, transmission, income, leasing, tenor and discount as source attributes to obtain root nodes and other nodes. Based on the results of the classification using the C4.5 algorithm shows that the accuracy reached 67.5%, which shows that the C4.5 algorithm is suitable for measuring sales predictions for sales of Honda cars.

Keywords : *Data Mining, Product Sale, Decision Tree, Algoritma C4.5.*

1. Pendahuluan

Teknologi pengumpulan dan penyimpanan data telah memudahkan organisasi untuk mengumpulkan sejumlah data berukuran besar. Ketersediaan data dan informasi yang lengkap, benar dan juga tepat sudah menjadi kebutuhan pokok bagi kelangsungan dan kelancaran suatu organisasi.

PT Prospect Motor adalah sebuah perusahaan yang bergerak di bidang penjualan mobil honda. Dalam meningkatkan kinerja operasional, perusahaan berusaha untuk melayani kebutuhan customernya. Perusahaan harus memperhatikan jenis transaksi penjualan produk yang di pasarkan dengan berbagai cara sehingga dapat meningkatkan efektifitas kinerja perusahaan dalam mengolah data transaksi penjualan.

Data penjualan yang terus bertambah membuat perusahaan kesulitan mendapatkan informasi strategis seperti mendapatkan data penjualan produk- produk laris. Berdasarkan permasalahan dibuatlah analisa data mining pada data penjualan “Sales Person di PT Prospect Motor Cikarang”. Ketersedian data yang banyak dan kebutuhan akan informasi atau pengetahuan sebagai pendukung pengambilan keputusan untuk

membentuk solusi bisnis dibidang teknik informatika merupakan cikal bakal dari lahirnya teknologi data mining.

Penggunaan teknik data mining diharapkan dapat membantu mempercepat proses pengambilan keputusan, memungkinkan perusahaan untuk mengelola informasi yang terkandung didalam data transaksi menjadi pengetahuan yang baru. dan tak secara langsung mengetahui data penjualan menjadi data mining.

2. Tinjauan Studi

2.1. Pengertian Produk

Produk adalah semua hal yang dapat ditawarkan kepada pasar yang dapat memuaskan keinginan atau kebutuhan konsumen (Kotler, Armstrong, Brown, & Adam, Principles of Marketing, 2006). Sedangkan Kotler dan Keller mendefinisikan produk adalah sesuatu yang dapat ditawarkan kepada pasar untuk memuaskan suatu keinginan atau kebutuhan, termasuk barang fisik, jasa, pengalaman, acara, orang, tempat, properti, organisasi, informasi dan ide (Kotler & Keller, 2009).

2.2. Pengertian kecerdasan buatan

Pengertian kecerdasan buatan adalah salah satu bagian ilmu komputer yang membuat agar mesin (komputer) dapat melakukan pekerjaan seperti dan sebaik yang dilakukan manusia menurut (Kusumadewi, 2003).

2.3. Machine Learning (ML)

Machine learning (ML), bagian dari kecerdasan buatan (*artificial intelligence*), merupakan metode untuk mengoptimalkan performa dari sistem dengan mempelajari data sampel atau data histori (Alpaydin, 2009). Machine learning adalah suatu area dalam artificial intelligence atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik pemrograman berdasarkan pembelajaran masa lalu dan bersinggungan dengan ilmu statistik kadang juga optimasi (Mujiasih, 2011)

2.4. Data Mining

Data Mining adalah proses ekstraksi informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang ilmu statistik, mesin pembelajaran, dan sistem manajemen database (Ranjan, 2007). Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose D. T., 2005).

2.5. Algoritma C4.5

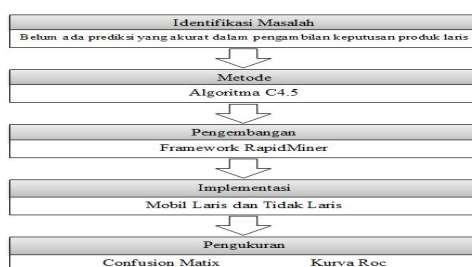
Algoritma C4.5 adalah pohon keputusan. Pohon keputusan (*Decision Tree*) adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*). Simpul pada sebuah pohon keputusan dibedakan menjadi tiga yaitu akar simpul, simpul percabangan dan simpul akhir (Abdillah, 2015).

3. Desain Penelitian/Metodologi

Penelitian ini merupakan penelitian dengan menggunakan metode penelitian eksperimen yaitu suatu cara untuk mencari hubungan sebab akibat (hubungan kausal) antara dua faktor yang sengaja ditimbulkan oleh peneliti dengan mengeliminasi atau mengurangi atau menyisihkan faktor-faktor lain yang mengganggu (Arikunto, 2006).

Dalam pengumpulan data terdapat sumber data yang dihimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder (Riduwan, 2008). Data sekunder yang digunakan pada penelitian ini adalah berupa data penjualan sales executive PT Prospect Motor Cikarang berjumlah 100 data kemudian dibagi 2 yaitu 80% data training dan 20% data testing. Hasil data training digunakan untuk memperoleh hasil klasifikasi produk laris dalam bentuk pohon keputusan dan data testing digunakan untuk tingkat akurasi dari hasil klasifikasi tersebut.

Dalam penelitian ini akan dilakukan analisa menggunakan metode algoritma C4.5. Data dihitung dengan algoritma sesuai dengan metodenya kemudian dicari hasil akurasinya. Dalam tahapan ini akan dilakukan beberapa langkah pengujian data sebagai berikut :



Gambar 1. Langkah Pengujian Data

4. Hasil Penelitian dan Pembahasan

4.1. Pengujian

4.1.1. Pengujian data

Data yang digunakan pada penelitian ini yaitu 100 data penjualan, 20 data testing dan 80 data training untuk menentukan akurasi prediksi produk laris. Data yang digunakan dalam perhitungan decision tree secara manual menggunakan 20% dari data penjualan yaitu 20 data. Data kemudian dikategorikan dengan variabel, atribut dengan nilai laris dan tidak laris.

4.1.2. Pemodelan Data Mining Menggunakan Algoritma C4.5

Berikut ini merupakan langkah-langkah perhitungan nilai Entrophy dan Gain data testing pada setiap atribut yang memiliki label laris dan tidak laris :

1. Perhitungan Nilai Entrophy

Langkah awal algoritma C4.5 adalah mencari nilai Entrophy. Pertama tentukan terlebih dahulu nilai entrophy total dalam kasus dengan rumus sebagai berikut :

Maka :

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

$$Entropy \text{ Total Data Testing } (S) = -(total \text{ laris} / total \text{ kasus}) * \log_2 (total \text{ laris} / total \text{ kasus}) + -(total \text{ tidak laris} / total \text{ kasus}) * \log_2 (total \text{ tidak laris} / total \text{ kasus})$$

2. Perhitungan Nilai Gain

Setelah nilai entropy sudah diketahui, maka langkah selanjutnya adalah mencari nilai gain, dengan rumus :

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Keterangan :

S = Himpunan Kasus A = fitur

N = jumlah partisi atribut A

|Si| = proporsi Si terhadap S

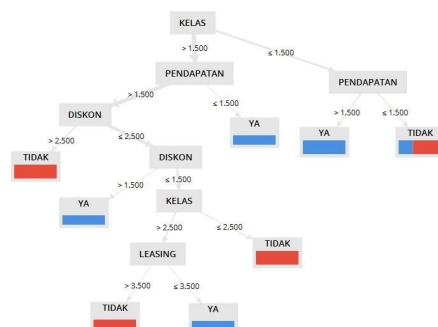
|S| = jumlah kasus dalam S Maka :

Gain = Entrophy Total – (kasus / total kasus * Entrophy kasus) Tabel 1. Nilai Entrophy dan Gain Data Testing

Tabel 1. Rule untuk Prediksi Produk Laris

ATRIBUT		JUMLAH KASUS	YA	TIDAK	ENTROPY	GAIN
	TOTAL	20	10	10	1	
KELAS	LCGC	6	4	2	0,918296	0,042738
	MPV	8	3	5	0,954434	
	SUV	6	3	3	1	

Nilai gain tertinggi adalah nilai gain dari kelas mobil yang dijadikan node akar pertama, untuk node akar selanjutnya, dilakukan perhitungan lagi dan seterusnya.



Gambar 2. Decision Tree dari Data Testing

Maka diperoleh rule untuk prediksi produk laris yaitu :

Tabel 2. Rule untuk Prediksi Produk Laris

Rule 1	Jika kelas mpv dan suv, pendapatan sedang dan tinggi, diskon tinggi maka tidak
--------	--

	laris
Rule 2	Jika kelas mpv dan suv, pendapatan sedang dan tinggi, diskon sedang maka laris
Rule 3	Jika kelas mpv dan suv, pendapatan sedang dan tinggi, diskon rendah, kelas suv, leasing bca dan mbf maka tidak Laris
Rule 4	Jika kelas mpv dan suv, pendapatan sedang dan tinggi, diskon rendah, kelas suv, leasing u-finance, mtf dan tidak maka laris
Rule 5	Jika kelas mpv dan suv, pendapatan sedang dan tinggi, diskon rendah, kelas lcgc dan mpv maka tidak laris
Rule 6	Jika kelas mpv dan suv, pendapatan rendah maka laris
Rule 7	Jika kelas lcgc, pendapatan sedang dan tinggi maka laris
Rule 8	Jika kelas lcgc, pendapatan rendah maka tidak laris

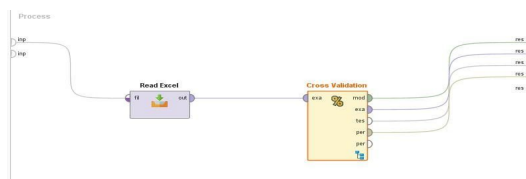
4.1.1 Pengujian Data Menggunakan Tool RapidMiner

1. Proses Import Data
Data penjualan diimport kedalam tool RapidMiner 8.0. Data yang digunakan dalam aplikasi ini type Excel dengan jumlah 100 data (80% data training dan 20% data testing).

KELAS	MODEL	TRANSMISI	PENDAPAT	LEASING	TENOR	DISKON	LARIS
integer	integer	integer	integer	integer	integer	integer	polyno...
attribute	attribute	attribute	attribute	attribute	attribute	attribute	label
3	3	2	1	2	3	1	TIDAK
2	2	2	1	3	3	2	YA
3	3	2	2	4	3	1	YA
3	4	2	2	5	2	1	YA
1	1	2	2	3	3	2	YA

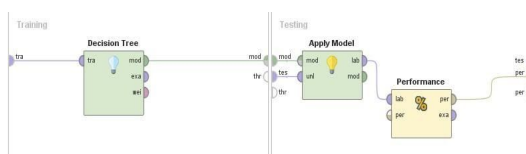
Gambar 3.. Import Data Penjualan

2. Proses Validasi



Gambar 4. Proses Validasi

3. Proses Training dan Testing
Proses training digunakan untuk proses blok modecision Tree, dihubungkan dengan garis penghubung pada blok Apply Model dan blok Performance dibagian proses testing sebagai penampil hasil dari pengujian data, hasil dari pengujian ini akan menghasilkan algoritma C4.5.



Gambar 5. Proses Training dan Testing

4.2. Pembahasan

Uji coba dalam penelitian ini dilakukan sebanyak 2 kali yaitu menggunakan 20% data testing dan 80% data training.

4.2.1. Confusion Matrix

Confusion matrix adalah alat yang berguna untuk menganalisis seberapa baik classifier mengenali tuple dari kelas yang berbeda. TP dan TN memberikan informasi ketika classifier benar, sedangkan FP dan FN memberitahu ketika classifier salah (Han & Kamber, 2011).

		True Class	
		Positive	Negative
Predicted Class	Positive	true positives count (TP)	false negatives count (FP)
	Negative	false positives count (FN)	true negatives count (TN)

Tabel 3. Confusion Matrix

1. Akurasi Data Testing
Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total data sample yang diuji.
TP=5 TN=2 FP=5 FN=8

$$\begin{aligned}
 accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{5+2}{5+2+8+5} \\
 &= \frac{7}{20} \\
 &= 35\%
 \end{aligned}$$

accuracy: 35.00% +/- 23.91% (mikro: 35.00%)

	true YA	true TIDAK	class precision
pred YA	5	8	38.46%
pred TIDAK	5	2	28.57%
class recall	50.00%	20.00%	

Gambar 6. Akurasi Data Testing

2. Akurasi Data Training

TP=31
TN=23 FP=17 FN=9

$$\begin{aligned}
 accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{31+23}{31+23+17+19} \\
 &= \frac{54}{80} \\
 &= 67,5\%
 \end{aligned}$$

accuracy: 67.50% +/- 12.73% (mikro: 67.50%)

	true TIDAK	true YA	class precision
pred TIDAK	23	17	57.50%
pred YA	9	31	77.50%
class recall	71.88%	64.58%	

Gambar 7. Akurasi Data Training

4.2.2. Kurva ROC/AUC

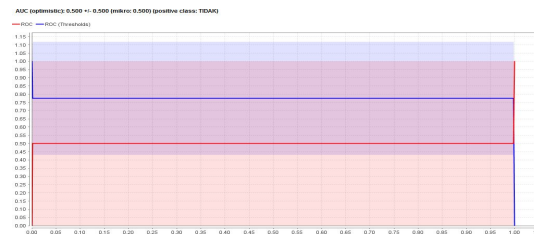
Confusion matrix hanya menyajikan informasi dalam bentuk angka, maka jika ingin menampilkan informasi kinerja algoritma klasifikasi dalam bentuk grafik dapat digunakan Kurva *Receiver Operating Characteristic* (ROC). Kurva ROC digunakan untuk membandingkan kinerja diagnostik dari dua atau lebih tes laboratorium atau diagnostik (Griner, 1981). Kurva ROC digunakan untuk melakukan sebuah analisa terhadap model classifier yang telah dibuat. Kurva ROC dibuat berdasarkan nilai yang telah didapatkan pada perhitungan *confusion matrix*, yaitu antara False Positive Rate dengan True Positive Rate.

AUC (*Area Under Curve*) adalah luas daerah di bawah kurva ROC, bila nilainya mendekati satu, maka model yang didapat lebih akurat. AUC memiliki tingkat diagnosa (Hidayati, Listyawati, & Setyawan, 2008), yaitu :

1. Akurasi bernilai 0,90 – 1,00 = Excellent Classification
2. Akurasi bernilai 0,80 – 0,90 = Good Clasification
3. Akurasi bernilai 0,70 – 0,80 = Fair Clasification
4. Akurasi bernilai 0,60 – 0,70 = Poor Clasification
5. Akurasi bernilai 0,50 0,60 = Failur

4.2.3. Kurva ROC/AUC Optimis Data Testing

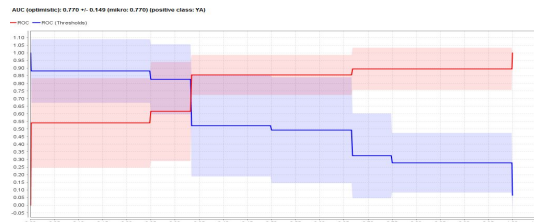
Hasil yang di dapat dari pengolahan AUC optimis untuk metode algoritma C4.5 dengan menggunakan tool RapidMiner sebesar 0,458 dapat dilihat gambar di bawah ini dengan tingkat diagnosa *Failur Clasification*.



Gambar 8. AUC Optimistic Data Testing

4.2.4. Kurva ROC/AUC Optimis Data Training

Hasil yang di dapat dari pengolahan AUC optimis data training sebesar 0,770 dengan tingkat diagnosa *Fair Clasification*.



Gambar 9. AUC Optimistic Data Training

5. Kesimpulan

Dari hasil penelitian yang telah dilakukan oleh peneliti, maka dapat disimpulkan bahwa produk laris dapat diprediksi dan dievaluasi dengan memanfaatkan teknik data mining menggunakan algoritma decision tree C4.5 untuk memprediksi (menentukan kelas) dari produk laris dengan data training yang telah diperoleh. Percobaan dan pengujian prediksi produk laris dengan rapidMiner menggunakan metode decision tree C4.5, diperoleh akurasi sebesar 76,25% dengan kriteria akurasi *Fair Classification* menggunakan confusion matrix. Berdasarkan pembahasan yang telah di lakukan dengan algoritma decision tree C4.5 dan di lakukan pengujian dengan aplikasi RapidMiner maka penulis dapat menarik kesimpulan sebagai berikut :

1. Data Mining dapat di implementasikan dengan menggunakan Database penjualan mobil Honda untuk memprediksi (menentukan kelas) dari produk laris dengan data training yang telah diperoleh.
2. Percobaan dan pengujian prediksi produk laris dengan rapidMiner menggunakan metode decision tree C4.5, diperoleh akurasi sebesar 72,5% dengan kriteria akurasi *Good Classification* menggunakan *confusion matrix*.

Daftar Pustaka

- Abdillah, S. (2015). Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi Data Mining Pada Rumah Sakit Santa Maria Pematang.
- Abdulla, Nair, Ghazali, & Khade. (2015). Measure Customer Behaviour using C4.5 Decision Tree Map Reduce Implementation in Big Data Analytics and Data Visualization.
- Alpaydin, E. (2009). Introduction to Machine Learning.
- Arikunto. (2006). Prosedur Penelitian Suatu Pendekatan Praktek. Jakarta : PT. Rineka Cipta.
- Assauri, S. (2004). Manajemen Pemasaran. Edisi II. Rajawali Press. Jakarta.
- Connell, M., & Servaes, H. (1995). Equity Ownership and the Two Faces of Debt.
- Cynthia, & Ismanto. (2018). Metode Decision Tree Alogoritma C4.5 Dalam Mengklarifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji.
- Danukusumo, K. P. (2017). Implementasi Deep Learning Menggunakan Convolutional Neural Network Untuk Klasifikasi Citra Candi Berbasis Gpu.
- Daryanto. (2013). Sari Kuliah Manajemen Pemasaran.
- Gorunescu, F. (2011). Data Mining Concept Model and Techniques. Griner. (1981). ROC Curve Analysis in MedCalc.
- Han, J., & Kamber, M. (2001). Data Mining: Concepts and Technique. sanfransisco.
- Han, J., & Kamber, M. (2006). Concepts and Techniques, Second Edition. Morgan Kaufmann Publishers.
- Han, J., & Kamber, M. (2011). Data Mining : Concepts and Techniques.
- Han, J., & Kamber, M. (2012). Data Mining Concepts and Techniques, Third Edition.
- Hermawati, F. (2013). Data Mining.
- Hidayati, N. A., Listyawati, S., & Setyawan, A. (2008). Kandungan kimia dan uji antiinflamasi ekstrak etanol *Lantana camara L.* pada tikus putih (*Rattus norvegicus L.*) jantan.
- Iskandar, D., & Suprpto, Y. K. (2013). Perbandingan akurasi klasifikasi tingkat kemiskinan antara algoritma C4.5 dan Naïve Bayes Clasifier.

- Julianto, W., Yunitarini, R., & Sophan, M. (2014). Algoritma C4.5 Untuk Penilaian Kinerja Karyawan. Universitas Trunojoyo Madura.
- Kotler, & Armstrong. (2008). *Principles of Marketing, 12th Edition*. Kotler, & Keller. (2009). Manajemen Pemasaran, Jilid 1 Edisi ke 13. Kotler, Armstrong, Brown, & Adam. (2006). *Principles of Marketing*.
- Kusnawi. (2007). Sistem Pendukung Keputusan Dengan Menggunakan Decision Tree Dalam Pemberian Beasiswa Di Sekolah Menengah Pertama.
- Kusumadewi, S. (2003). Artificial Intelligence (Teknik dan Aplikasinya). Laksana, F. (2008). Manajemen Pemasaran. 67.
- Larose. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. Larose, D. T. (2005). Discovering knowledge in data.
- Mujiasih, S. (2011). Pemanfaatan Data Mining Untuk Prakiraan Cuaca.
- Ranjan, J. (2007). Application of Data Mining Technique in Pharmaceutical Industry. *Journal of Theoretical and Applied Information Technology*.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier.
- Santoso. (2013). Analisa dan Penerapan Metode C4.5 untuk Prediksi Loyalitas Pelanggan.
- Swastha. (2004). Azas-azas Marketing. Swastha, B. (2005). Manajemen Pemasaran.
- Turnip, & Wijaya. (2016). Penerapan Algoritma C4.5 untuk Penentuan Tingkat Konsumsi Konsumen pada Medan Solusindo.
- Widiyono, & Pakkanna, M. (2013). Pengantar Bisnis : Respon terhadap Dinamika Global.
- Zulianto, A. (2003). Kontribusi Industri Kecil dan Menengah terhadap Peningkatan.