Penerapan Data Mining dalam Memprediksi Pembelian cat

Fitriana Harahap
STMIK POTENSI UTAMA
Jl. KL. Yos Sudarso KM 6,5 No 3 A Tj. Mulia Medan
Email : fitrianaharahap1@gmail.com

Abstrak

Untuk memudahkan dalam pengambilan keputusan dalam proses pembelian cat di Departement penjualan Home Smart Medan yang selama ini pengambilan keputusan seorang manager penjualan dalam mengambil keputusan dengan melihat seberapa dekat hubungan supplier dan seberapa banyak dana sponsor yang diberikan kepada perusahaan. Selain harga, type merek juga masih kalah saing dengan perusahaan lain. Pembelian cat yang kurang efektif, menyebabkan produk khususnya cat pada perusahaan ini kurang diminati oleh customer . Dengan menerapkan teknik klasifikasi data mining pada pembelian cat yang efektif Pada Departement Penjualan Home Smart, diharapkan nantinya dapat menghasilkan suatu pengetahuan yang dapat digunakan dalam pengambilan keputusan dalam melakukan pembelian cat yang efektif. Algoritma C4.5 adalah algoritma klasifikasi data bertipe pohon keputusan. Pohon keputusan Algoritma C4.5 dibangun dengan beberapa tahap yang meliputi pemilihan atribut sebagai akar, membuat cabang untuk tiap-tiap nilai dan membagi kasus dalam cabang. Tahapan-tahapan ini akan diulangi untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Dari penyelesaian pohon keputusan maka akan didapatkan beberapa rule. Dalam hal ini penulis mengklasifikasikan pembelian cat berdasarkan penjualan pada Departement Penjualan Home Smart. Penerapan Algoritma C4.5 ini dapat membantu Departement Penjuan Home Smart dalam menentukan pembelian cat dari Suplier.

Kata kunci: *Data Mining, Pembelian cat, Decision Tree, Algoritma C4.5*

1. Pendahuluan

Dalam ekonomi dan bisnis masih banyak perusahaan-perusahaan yang menggunakan selembar kertas ataupun hanya menggunakan aplikasi *Microsoft Excel* untuk mengolah data perusahaan. Seperti halnya pada *Departement* Penjualan *Home Smart* Medan. Meskipun pemanfaatan teknologi komputerisasi sudah terealisasikan, namun tidak begitu dalam mengambil keputusan untuk pembelian cat. Pengambilan keputusan seorang *manager purchasing* dalam mengambil keputusan apakah produk cat yang ditawarkan oleh perusahaan cat yaitu dengan melihat seberapa dekat hubungan *supplier* dan seberapa banyak dana sponsor yang diberikan kepada perusahaan. Sehingga terkadang tidak dapat bersaing dengan perusahaan lain. Selain harga, *type* merek juga masih kalah saing dengan perusahaan lain. Pembelian cat yang kurang efektif, menyebabkan produk khususnya cat pada perusahaan ini kurang diminati oleh *customer.*

Data mining adalah proses yang mengunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Salah satu teknik yang ada pada *data mining* adalah klasifikasi [1][2].

Dalam penelitian sebelumnya dengan judul " Perbandingan kinerja pohon keputusan *ID3* dan *C4.5* dalam identifikasi kelayakan kredit sepeda motor". Penelitian tersebut dilakukan untuk mengidentifikasi kelayakan kredit menggunakan algoritma pohon keputusan *ID3* dan *C4.5* serta untuk mengukur kinerja algoritma *ID3* dan *C4.5* dari sisi keakuratan hasil prediksi. Pengukuran kinerja yang dilakukan menggunakan sekelompok data uji untuk mengetahui persentase *precision*, *recall* dan *accurancy*. Hasil akhir dari penelitian ini menunjukkan bahwa algoritma *C4.5* memiliki tingkat akurasi yang lebih tinggi daripada algoritma *ID3[3]*.

Yi Jiang et al melakukan penilaian terhadap kredit debitur. Penelitiannya menyatakan bahwa*C4.5* adalah algoritma pembelajaran yang mengadopsi strategi pencarian lokal, dan dapatmemperoleh aturan keputusan terbaik[5].

2. Metode Penelitian

Metode penelitian dilakukan dengan studi literatur terhadap sumber-sumber yang relevan, analisis pengetahuan terhadap faktor pembelian cat menggunakan algoritma C4.5. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5 . Algoritma C4.5 merupakan pengembangan dari algoritma ID3 [4].

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang memprediksikan aturan. Aturan dapat dengan mudah dipahami dengan alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu [2].

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule dan menyederhanakan rule. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [4] :

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus di bawah ini :

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

Di mana :

S = Himpunan kasus
A = Atribut
n = Jumlah partisi atribut A
$|S_i|$= Jumlah kasus pada partisi ke-i
$|S|$= Jumlah kasus dalam S

Sementara itu, perhitungan nilai *entropy* adalah seperti persamaan 2 di bawah ini :

$$Entropy(S) = \sum_{i=1}^{n} - pi * \log_2 pi \qquad (2)$$

Di mana :

S = Himpunan kasus *n* = Jumlah partisi S *A* = Fitur
Pi = Proporsi dari $|S_i|$ terhadap S

Analisa Data

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon, salah satu atribut merupakan atribut yang menyatakan data solusi per item data yang disebut target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*.

Pemilihan Variabel.

Dari data-data yang telah diperoleh, maka akan ditentukan suatu variabel yang menjadi variabel keputusan dalam pembelian cat. Diketahui beberapa faktor yang menjadi penentu dalam pembelian cat oleh *home smart* adalah tingkat animo *customer* terhadap produk cat yang dapat dilihat berdasarkan hasil penjualan merek cat dengan warna cat tertentu pada *home smart*, jika hasil penjualan cat tersebut tinggi maka animo *customer* tinggi terhadap cat tersebut. Kompetisi *supplier* juga menjadi indikator dalam pembelian cat dimana persaingan supplier dalam menawarkan produk mereka kepada *home smart* medan

dengan memberikan cat dengan kualitas Super, Medium dan standar dengan penawaran harga yang berbeda. Tingkat kompetisi dikatakan tinggi jika hasil penjualan cat dari supplier tersebut tinggi dan banyak diminati oleh *customer*.Berdasarkan analisa tersebut dapat ditentukan variabel-variabel yang digunakan dalam penentuan pembelian cat dengan mempertimbangkan faktor di atas yaitu : kualitas, harga, animo dan kompetisi.

Tabel 1. Data Hasil Pra-Proses

No	Kualitas	Animo	Harga	Kompetisi	Beli
1	Medium	Rendah	Mahal	Sedang	Ya
2	Medium	Sedang	Mahal	Sedang	Ya
3	Super	Tinggi	Mahal	Rendah	Ya
4	Standar	Sedang	Mahal	Sedang	Ya
5	Standar	Rendah	Terjangkau	Sedang	Ya
6	Standar	Rendah	Terjangkau	Tinggi	Ya
7	Standar	Rendah	Mahal	Sedang	Tidak
8	Super	Rendah	Mahal	Sedang	Ya
9	Super	Tinggi	Normal	Sedang	Ya
10	Super	Rendah	Terjangkau	Sedang	Ya
11	Medium	Rendah	Terjangkau	Sedang	Tidak
12	Medium	Sedang	Terjangkau	Sedang	Tidak
13	Standar	Rendah	Mahal	Rendah	Tidak
14	Standar	Sedang	Mahal	Rendah	Tidak
15	Super	Rendah	Terjangkau	Rendah	Tidak
16	Super	Sedang	Terjangkau	Rendah	Tidak
17	Super	Rendah	Mahal	Sedang	Ya
18	Super	Sedang	Mahal	Sedang	Ya
19	Medium	Tinggi	Mahal	Tinggi	Tidak
20	Medium	Sedang	Normal	Tinggi	Tidak
21	Medium	Rendah	Terjangkau	Tinggi	Tidak

Adapun Pra-Proses yang dilakukan dalam mempertimbangkan faktor diatas diambil berdasarkan sampel data penjualan

1. **Mengelompokkan Kualitas Cat.**

Tabel 2. Tabel Klasifikasi Kualitas

Klasifikasi	Kualitas
>10 liter	Super
>20 kg	Medium
5 kg-20kg	Standar

2. **Mengelompokkan Harga Cat.**

Pengelompokkan Harga cat diklasifikasikan menjadi tiga kelas yaitu, harga dikatakan Mahal jika harga lebih besar dari Rp. 150.000, harga dikatakan normal jika harga mencapai Rp76.000 sampai dengan 150.000 dan harga dikatakan terjangkau jika harga dibawah dari Rp. 76.000.Berikut harga cat dalam *range* yang tampak seperti tabel 4.2 dibawah ini :

Tabel 3. Tabel Klasifikasi Harga

Harga	Klasifikasi
>150.000	Mahal

76.000-150.000	Normal
10.000-75.000	Terjangkau

3. Mengelompokkan Animo

Pengelompokkan Animo diambil berdasarkan hasil penjualan per produk cat yang dipasarkan dengan berbagai cara yang dilakukan pihak produsen cat. Animo dikatakan rendah jika hasil penjualan per produk mencapai Rp.500.000 sampai Rp 5.500.000, Animo dikatakan sedang jika hasil penjualan mencapai 5.600.000 sampai 16.000.000 dan animo dikatakan tinggi jika hasil penjualan lebih besar dari Rp. 16.000.000

Tabel 4. Tabel Klasifikasi Animo

Animo	Klasifikasi
>16.000.000	Tinggi
5.600.000-16.000.000	Sedang
500.000-5.500.000	Rendah

4. Mengelompokkan Kompetisi

Pengelompokkan kompetisi diambil berdasarkan hasil penjualan produk cat per supplier yang memasarkan produk cat tersebut.Kompetisi diklasifikasikan menjadi tinggi, sedang dan rendah. Kompetisi dikatakan tinggi jika hasil penjualan per sulier mencapai lebih besar dari Rp. 50.000.000, kompetisi dikatakan sedang jika hasil penjualan per supplier mencapai Rp. 41.000.000 sampai Rp. 50.000.000 dan kompetisi dikatakan rendah jika hasil penjualan mencapai Rp. 10.000.000 sampai 40.000.000

Tabel 5. Tabel Klasifikasi Kompetisi

Kompetisi	Klasifikasi
>50.000.000	Tinggi
41.000.000-50.000.000	Sedang
10.000.000-40.000.000	Rendah

Perhitungan *Entropy* dan *Gain.*

Entropy (Total) $= \left(-\frac{10}{21} * \log_2 \left(\frac{10}{21}\right)\right) + \left(-\frac{11}{21} * \log_2 \left(\frac{11}{21}\right)\right)$

$= 0.99836$

Entropy (Kualitas Super) $= \left(-\frac{2}{8} * \log_2 \left(\frac{2}{8}\right)\right) + \left(-\frac{6}{8} * \log_2 \left(\frac{6}{8}\right)\right)$

$= 0.81128$

Entropy (Kualitas Medium) $= \left(-\frac{5}{7} * \log_2 \left(\frac{5}{7}\right)\right) + \left(-\frac{2}{7} * \log_2 \left(\frac{2}{7}\right)\right)$

$= 0.86312$

Entropy (Kualitas Standar) $= \left(-\frac{3}{6} * \log_2 \left(\frac{3}{6}\right)\right) + \left(-\frac{3}{6} * \log_2 \left(\frac{3}{6}\right)\right)$

$= 1.$

Entropy (Animo Tinggi) $= \left(-\frac{1}{3} * \log_2 \left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2 \left(\frac{2}{3}\right)\right)$

$= 0.9183$

Entropy (Animo Sedang) $= \left(-\frac{4}{7} * \log_2 \left(\frac{4}{7}\right)\right) + \left(-\frac{3}{7} * \log_2 \left(\frac{3}{7}\right)\right)$

$= 0.98523$

Entropy (Animo Rendah) $= \left(-\frac{5}{11} * \log_2 \left(\frac{5}{11}\right)\right) + \left(-\frac{6}{11} * \log_2 \left(\frac{6}{11}\right)\right)$

$= 0.99403$

Entropy (Harga Mahal) $= \left(-\frac{4}{11} * \log_2 \left(\frac{4}{11}\right)\right) + \left(-\frac{7}{11} * \log_2 \left(\frac{7}{11}\right)\right)$

= 0.94566

Entropy (Harga Normal) $= (-\frac{1}{2} * \log_2 (\frac{1}{2})) + (-\frac{1}{2} * \log_2 (\frac{1}{2})) = 1.$

Entropy (Harga Terjangkau) $= (-\frac{4}{5} * \log_2 (\frac{4}{5})) + (-\frac{1}{5} * \log_2 (\frac{1}{5}))$

= 0.72193

Entropy (Kompetisi Tinggi) $= (-\frac{3}{4} * \log_2 (\frac{3}{4})) + (-\frac{1}{4} * \log_2 (\frac{1}{4}))$

= 0.81128.

Entropy (Kompetisi Sedang) $= (-\frac{3}{12} * \log_2 (\frac{3}{12})) + (-\frac{9}{12} * \log_2 (\frac{9}{12}))$

= 0.81128

Entropy (Kompetisi Rendah) $= (-\frac{3}{12} * \log_2 (\frac{3}{12})) + (-\frac{9}{12} * \log_2 (\frac{9}{12}))$

= 0.81128.

Sementara itu, nilai *Gain* pada baris kualitas dihitung dengan menggunakan persamaan 1 :

Gain(Total, kualitas) = $0.99836 - ((\frac{8}{21} * 0.81128) + (\frac{7}{21} * 0.86312) + (\frac{6}{21} * 1))$

= 0.11588

Gain(Total, Animo) $= 0.99836 - ((\frac{3}{21} * 0.9183) + (\frac{7}{21} * 0.98523) + (\frac{11}{21} * 0.99403))$ = 0.01809

Gain(Total, Harga) $= 0.99836 - ((\frac{11}{21} * 0.94566) + (\frac{2}{21} * 1) + (\frac{8}{21} * 0.95443))$

= 0.04419

Gain (Total,Kompetisi) =

$0.99836 - ((\frac{4}{21} * 0.81128) + (\frac{12}{21} * 0.81128) + (\frac{5}{21} * 0.72193))$ = 0.21687.

Tabel 6. Perhitungan Node 1

Node			Jlm Kasus	Tidak	Ya	Entropy	Gain
1	Total		21	10	11	0.99836	
	Kualitas						0.11588
		Super	8	2	6	0.81128	
		Medium	7	5	2	0.86312	
		Standar	6	3	3	1	
	Animo						0.01809
		Tinggi	3	1	2	0.9183	
		Sedang	7	4	3	0.98523	
		Rendah	11	5	6	0.99403	
	Harga						0.04419
		Mahal	11	4	7	0.94566	
		Normal	2	1	1	1	
		Terjangkau	8	5	3	0.95443	
	Kompetisi						0.21687
		Tinggi	4	3	1	0.81128	
		Sedang	12	3	9	0.81128	
		Rendah	5	4	1	0.72193	

Dari hasil pada tabel 6 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah kompetisi, yaitu sebesar 0.21687. Dengan demikian, kompetisi dapat menjadi node akar. Ada tiga nilai atribut dari kompetisi, yaitu tinggi, sedang dan rendah, sehingga perluh dilakukan perhitungan lagi.

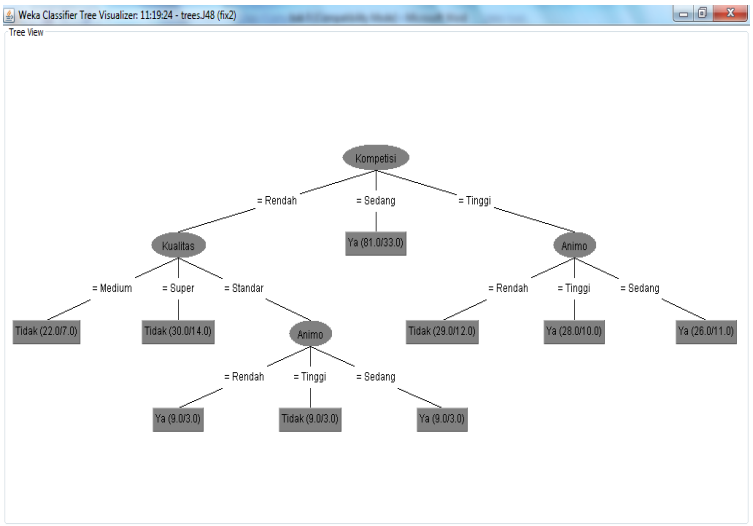
3. Hasil dan Pembahasan

Pengujian terhadap analisa, sangat penting dilakukan untuk menentukan dan memastikan apakah hasil analisa tersebut telah sesuai dengan keputusan yang diharapkan. Untuk menguji kebenaran dari hasil pengolahan data yang dilakukan secara manual, maka dapat menggunakan salah satu software aplikasi WEKA 3.5.5 *knowledge Explorer.*

Langkah - langkah Implementasi

Seluruh variabel yang terdiri dari atribut kondisi dan atribut keputusan yang digunakan untuk menentukan pembelian cat disimpan pada Microsoft excel dengan nama file datacat.xls (yang berisi kasus atau kriteria dalam menghasilkan rule). Selanjutnya proses *transformation* data dimana File datacat.xls kemudian disimpan dengan extension csv, selanjutnya file dibuka dengan notepad atau editor teks lainnya dan data sudah berubah dalam format command separated seperti gambar 4.2. Kemudian data disesuaikan dengan menambahkan informasi awal dan data tersebut sudah dapat digunakan sebagai inputan dalam WEKA 3.5.5.

Selanjutnya Klik *buttom Choose*, pilih *J48* dalam pembentukan pohon keputusan dan klik menu *Start* .
Pada tahap ini proses data mining dilakukan dengan memilih algoritma yang akan dipakai dalam menghasilkan sebuah pohon keputusan, algoritma C4.5.



Gambar 1. Tree View

4. Simpulan

Berdasarkan hasil penelitian yang penulis lakukan pada *Home Smart* Medan, maka penulis dapat menarik kesimpulan bahwa pembelian cat dengan menggunakan metode *Data Mining* khususnya *Algoritma C4.5* akan bermanfaat sekali dalam proses pengambilan keputusan dalam pembelian cat pada *Home Smart* Medan.

1. Yang menjadi faktor tertinggi yang mempengaruhi pembelian cat pada *Home Smart* adalah faktor kompetisi supplier dalam memasarkan produknya.

2. Faktor kedua yang mempengaruhi pembelian cat *Home Smart* adalah Kualitas cat dan Animo Masyarakat untuk mengetahui dan membeli produk cat yang dipasarkan dengan berbagai cara yang dilakukan pihak produsen cat tersebut.

3. Faktor Harga tidak mempengaruhi pembelian pada *Departement* Penjualan *Home Smart* Medan, karena cat dengan harga mahal ternyata masih diminati oleh pelanggan *Home Smart* Medan.

Selanjutnya penulis menyarankan agar dapat membandingkan metode pangambilan keputusan dengan metode *Data Mining* dengan metode lainnya.

Daftar Pustaka

[1] Efraim Turban, Jay E. Aronson, Ting Peng Liang, 2005. *Decision Support System and Intelligent Systems Edisi 7 Jilid 1,* Andi Yogyakarta.

[2] Kusrini, (2009). *Algoritma Data Mining* ,Andi Yogyakarta

[3] Budanis Dwi Meilani Achmad dan Fauzi Slamat, 2012.”*Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree “,Vol 16, No.1,Mei.*

[4] Muhammad Syahril, 2011 .”*Konversi Data Training Tentang Penyakit Hipertensi Menjadi Bentuk Pohon Keputusan dengan Teknik Klasifikasi Menggunakan Tools Rapid Miner 4.1 ,Vol 10, No.2,Mei.*

[5] Jiang,Yi. et al, (2007*).”A Bank Customer CreditEvaluation Based on the Decision Tree and theSimulated Annealing Algorithm. Journal of Department of Computer Science XiamenUniversity (IEEE International Co 8-11 July 2008)”.*