PENERAPAN ALGORITMA C4.5 DALAM *DATA MINING* UNTUK MENENTUKAN KONTRAK KERJASAMA KONSINYASI DI CV. PANJI URBAN BANDUNG

ARTIKEL

LUQMAN PANJI BUDIMAN

10511214



PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS KOMPUTER INDONESIA
BANDUNG
2015

ABSTRACT

Papersmooth is one of the Creative Industry engaged in a fashion that sells products such as T-shirt, Shirts, Jackets, Accessories, etc. The company is based in Bandung and has a lot of marketing points around Indonesia. The company's distribution point over 40 marketing point such Jakarta city, Bandung, Yogyakarta, etc.

With the number of consignment points there are also sales data which can be used as a decision support process in the business world. The data currently owned by Papersmooth not used optimally. This is causes many of stored data but little information that can be used to support decision making strategically

Keywords : Algoritma C4.5, decision tree, sales strategy, papersmooth

BAB I

PENDAHULUAN

1.1 Latar Belakang

Didalam dunia bisnis setiap perusahaan sudah tentu harus bisa bersaing dan berpikir bagaimana caranya agar perusahaan bisa terus berkembang dan dapat memperluas skala bisnisnya tersebut. Namun, masalah – masalah yang dihadapi perusahaan memiliki tingkat kesulitan dan kompleksitas yang bervariasi, mulai dari masalah yang sangat sederhana sampai dengan masalah yang sangat rumit. Salah satu contoh masalahnya adalah tentang menentukan keputusan, memilih keputusan bukanlah perkara yang mudah dengan banyaknya pertimbangan yang ada terkadang untuk mendapatkan suatu keputusan itu sulit mempertimbangkannya.

Untuk mendukung agar dapat terwujudnya suatu keputusan yang dapat digunakan untuk mendapatkan solusi dengan tingkat resiko yang kecil diperlukanlah suatu keputusan yang tepat dalam menentukan segala macam strategi untuk memajukan perusahaan. Dalam mengembangkan skala bisnis dan meningkatkan pendapatan perusahaan harus ditentukannya strategi pemasaran yang tepat. Untuk dapat melakukan hal tersebut, informasi yang dibutuhkan harus mencukupi hal ini agar dapat di analisis lebih lanjut lagi.

1.2 Identifikasi dan Rumusan Masalah

1.2.1 Identifikasi Masalah

Berdasarkan uraian dari latar belakang permasalahan yang dihadapi oleh Papersmooth adalah kesulitan dalam hal melakukan pengolahan informasi penjualan sebagai informasi daya serap penjualan pada setiap titik konsinyasi dalam mempertimbangkan kelanjutan kerjasama.

1.2.2 Rumusan Masalah

Pokok masalah yang akan dibahas melalui penelitian ini adalah dengan mengacu pada uraian latar belakang masalah diatas yaitu:

1. Bagaimana cara membuat pohon keputusan dalam Algoritma C4.5 untuk menentukan kerjasama kontrak konsinyasi

2. Bagaimana menentukan aturan keputusan kerjasama kontrak konsinyasi

1.3 Maksud dan Tujuan Penelitian

1.3.1 Maksud Penelitian

Maksud dari penelitian ini adalah untuk menerapkan teknik *data mining* dalam mengolah informasi mengenai penjualan produk berdasarkan data – data penjualan dari setiap titik konsinyasi yang ada sehingga dapat menentukan kelanjutan kontrak konsinyasi.

1.3.2 Tujuan Penelitian

Memberikan informasi untuk pengambilan keputusan dengan menggunakan teknik *data mining* dengan metode algoritma C4.5 ini guna membantu memberikan suatu keputusan terhadap kerjasama konsinyasi.

1.4 Kegunaan Penelitian

Kegunaan dari penelitian yang dari hasil penelitian yang penulis lakukan adalah:

1.4.1 Kegunaan Praktis

1. Memberikan informasi guna mendukung dalam pengambilan keputusan dalam mendistribusikan produk.

2. Memberikan keputusan tentang kerjasama dengan titik konsinyasi yang paling berpotensi untuk memperpanjang masa kerjasamanya.

1.4.2 Kegunaan Akademis

1. Sebagai penerapkan dan mengembangkan kemampuan yang diperoleh selama kuliah di UNIKOM Bandung dan ide-ide baru selama penelitian, dan belajar menerapkan teknik *data mining* untuk menangani masalah-masalah yang terdapat di luar lingkungan perkuliahan, serta sebagai salah satu syarat untuk menyalesaikan studi di UNIKOM.

2. Berguna dalam mengetahui kempampuan dan pengetahuan terhadap teori – teori yang telah diperoleh semala kuliah. Mengetahui kesiapan mahasiswa dalam menghadapi dunia kerja yang sebenarnya.

3. Hasil penelitian ini diharapkan dapat menambah atau memberikan pemikiran ide kepada peneliti lain yang akan mengambil matakuliah Skripsi atau tugas akhir dalam kajian yang sama sebagai bahan referensi.

1.5 Batasan Masalah

Untuk menghindari pembahasan yang meluas, maka penulis hanya membatasi pembahasan permasalahan hanya pada Penerapan *data mining untuk* menentukan kerjasama konsinyasi dengan menggunakan Algoritma C4.5 mengikuti tahapan *Knowledge Discovery in Database (KDD)*, data yang dianalisis adalah data penjualan selama 4 bulan terakhir.

BAB II

LANDASAN TEORI

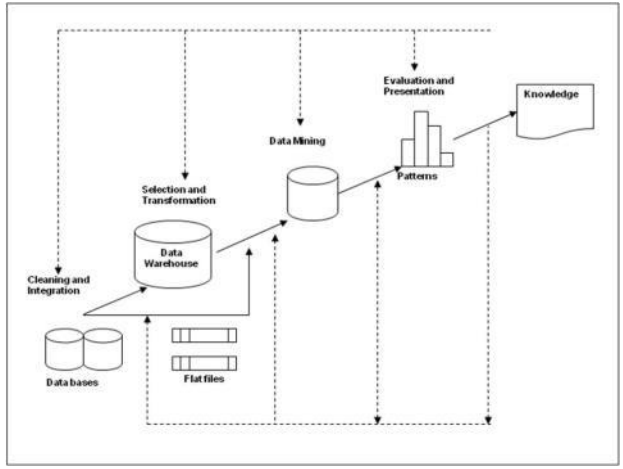
2.1 *Data Mining*

2.1.1 Pengertian *Data Mining*

Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, *data mining* dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field* dari sebuah relasional *database* yang besar. [1]

2.2 Metode *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Database adalah proses untuk menggali dan menganalisis sejumlah data dan mengestrak informasi dan pengetahuian yang berguna.Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan untuk keperluan pengambilan keputusan. Proses dalam KDD adalah proses yang digambarkan pada dan terdiri dari rangkaian proses sebagai berikut :



Gambar 2.1 Tahapan *Knowledge Discovery in Database*
Sumber : (Iko, 2003)

Tahap-tahap *data mining* ada 6 yaitu :

1. Pembersihan data (*Data Cleaning)*
2. Integrasi Data (*Data Integration)*
3. Transformasi Data (*Data Transformation)*
4. Proses *Mining*
5. Evaluasi Pola (*Pattern Evaluation)*
6. Presentasi Pengetahuan (*Knowledge Presentation)*

2.2.1 Model Pohon Keputusan

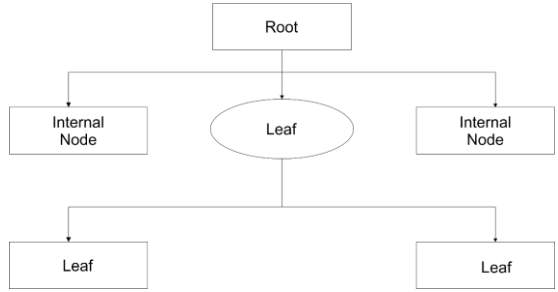
Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. *Decision tree* merupakan metode klasifikasi yang paling popular digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami. Pada *decision tree* terdapat 3 jenis *node*, yaitu:

a. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.

b. *Internal Node* , merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output minimal* dua.

c. *Leaf node* atau terminal *node* , merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

Contoh dari pohon keputusan dapat dilihat di Gambar 2.3 berikut ini :



Gambar 2.3 Model Pohon Keputusan
Sumber : (Hamidah, 2012)

2.3 Algoritma C4.5

2.3.1 Pengertian Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma *decision tree*. Algoritma ini mempunyai input berupa *training samples* dan *samples*. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field-field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data [2].

Algoritma C4.5 adalah pengembangan dari algoritma ID3. Oleh karena pengembangan tersebut algoritma C4.5 mempunyai prinsip dasar kerja yang sama dengan algoritma ID3. Hanya saja dalam algoritma C4.5 pemilihan atribut dilakukan dengan menggunakan *Gain Ratio* dengan rumus :

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

Keterangan :

S = Himpunan Kasus
A = Atribut
i = Jumlah Partisi Atribut
|Si| = Jumlah Kasus pada partisi ke i
|S| = Jumlah Kasus dalam S

Sedangkan penghitungan nilai *entropy* dapat dilihat pada rumus berikut

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

Dengan diketahui:

S : Himpunan Kasus

n : Jumlah partisi S

pi : Proporsi dari S_i terhadap S

Secara umum algoritma C4.5 dapat dibangun menggunakan pohon keputusan dengan cara :

a. Pilih atribut sebagai akar.
b. Buat cabang untuk masing-masing nilai.
c. Menghitung entropi kelas dan gain dari setiap atribut

Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang telah habis.

BAB III

OBJEK DAN METODE PENELITIAN

3.1 Objek Penelitian

Dalam penyusunan laporan penelitian ini penulis melakukan penelitian dengan mengambil objek penelitian pada kantor Papersmooth Clothing yang bertempat di Jl. Jend. Ahmad Yani No. 909 Bandung.

3.2 Metode Penelitian

Menurut Sugiyono (Sugiyono, 2007) mendefinisikan Metode Penelitian sebagai berikut :

> "Metode penelitian pada dasarnya merupakan cara ilmiah untuk mendapatkan data yang valid dengan tujuan dapat ditemukan, dibuktikan dan dikembangkan suatu pengetahuan sehingga pada gilirannya dapat digunakan untuk memahami, memecahkan, dan mengantisipasi masalah."

3.2.1 Desain Penelitian

Dalam menentukan desain penelitian, penulis melakukannya pada CV. Panji Urban / Papersmooth yang berada di Jl. Jendral Ahmad Yani no. 909 Bandung. Maka penulis akan melakukan penelitian pada Papersmooth ini untuk menganalisis data - data penjualan, penulis menggunakan metode kualitatif yang bersifat deskriptif pada pendekatan kasus di Papersmooth, yaitu metode yang pada tahap pertama memulai penulis melakukan dengan cara mengumpulkan data – data dan bahan yang diperlukan terlebih dahulu dan pada tahap berikutnya penulis mengolah dan membahas sampai pada suatu kesimpulan yang pada ada.

3.2.2 Jenis dan Metode Pengumpulan Data

Dalam pembuatan penelitian ini metode pengumpulan data yang digunakan penulis adalah metode pengumpulan dengan cara menggabungkan data dari sumber data primer dan sumber data sekunder yang menunjang pada tujuan penelitian.

3.2.2.1 Sumber Data Primer

Dalam penulisan penelitian ini, penulis mengumpulkan data sebagai bahan dari pembuatan laporan dengan menggunakan metode deskriptif, yaitu teknik pengumpulan data yang digunakan adalah :

a. Wawancara (*Interview)*
b. Studi Pustaka
c. Observasi

3.2.2.2 Sumber Data Sekunder

Dalam pengumpulan data sekunder penulis menggunakan cara dokumentasi yaitu mengumpulkan bahan – bahan yang tertulis berupa data/laporan yang diperoleh dari Papermooth selama melakukan kegiatan penelitian. Selain itu penulis juga melihat referensi dari penulisan penelitian sebelumnya yang berkaitan dengan judul.

BAB IV
HASIL DAN PEMBAHASAN

4.1 Penilaian dan Perhitungan

Dari hasil analisis dokumen penjualan, aset dan biaya operasional maka dapat dilihat pada tabel dibawah ini :

2015																		
BULAN	MAGNUM YOGYA		R	MAGNUM MAKASAR		R	7SOUL		R	INSPIRED		R	ISSUE MAKASAR		R	BLOOP		R
	ASSET	OMSET		ASSET	OMSET		ASSET	OMSET		ASSET	OMSET		ASSET	OMSET		ASSET	OMSET	
JANUARI	206,885,000	19,385,000	9.37%	208,500,000	28,920,000	13.87%	137,585,000	19,990,000	14.53%	148,070,000	22,309,000	15.07%	221,530,000	28,160,000	12.71%	174,750,000	31,235,000	17.87%
FEBRUARI	251,285,000	17,535,000	6.98%	251,258,000	20,510,000	8.16%	189,640,000	23,150,000	12.21%	203,095,000	18,039,000	8.88%	286,365,000	23,405,000	8.17%	217,340,000	29,690,000	13.66%
MARET	256,635,000	17,645,000	6.88%	260,390,000	21,195,000	8.14%	191,390,000	19,045,000	9.95%	205,585,000	19,675,000	9.57%	282,155,000	22,005,000	7.80%	255,910,000	30,309,000	11.84%
APRIL	238,990,000	17,775,000	7.44%	234,285,000	26,105,000	11.14%	174,620,000	19,990,000	11.45%	185,910,000	20,150,000	10.84%	260,150,000	23,685,500	9.10%	263,050,000	32,035,000	12.18%
MEI																		
JUNI																		
JULI																		
AGUSTUS																		
SEPTEMBER																		
OKTOBER																		
NOVEMBER																		
DESEMBER																		
AVERAGE	238,448,750	18,085,000	7.58%	238,608,250	24,182,500	10.13%	173,308,750	20,543,750	11.85%	185,665,000	20,043,250	10.80%	262,550,000	24,313,875	9.26%	227,762,500	30,817,250	13.53%

Gambar 4.1 Hasil Analisis Dokumen Penjualan dan Aset

FAMILLY AFFAIR			POPSHOP			WIDELY			HIPNOTIK			URBIE		
ASSET	OMSET	R	ASSET	OMSET	R	ASSET	OMSET	R	ASSET	OMSET	R	ASSET	OMSET	R
190,760,000	37,640,000	19.73%	154,201,000	15,955,000	10.35%	102,165,000	7,050,000	6.90%	103,845,000	9,525,000	9.17%	199,060,000	26,095,000	13.11%
262,795,000	25,210,000	9.59%	184,085,000	10,115,000	5.49%	86,345,000	5,730,000	6.64%	135,900,000	9,200,000	6.77%	257,840,000	16,490,000	6.40%
218,310,000	26,955,000	12.35%	109,720,000	7,350,000	6.70%	88,315,000	4,575,000	5.18%	126,700,000	11,065,000	8.73%	241,350,000	18,200,000	7.54%
253,750,000	30,790,000	12.13%	102,370,000	4,135,000	4.04%	83,740,000	3,765,000	4.50%	115,635,000	10,855,000	9.39%	223,150,000	22,750,000	10.19%
231,403,750	30,148,750	13.03%	137,594,000	9,388,750	6.82%	90,141,250	5,280,000	5.80%	120,520,000	10,161,250	8.52%	230,350,000	20,883,750	9.07%

Gambar 4.2 Hasil Analisis Dokumen Penjualan dan Aset (lanjutan)

Dari hasil analisis diatas dapat dibuat kedalam bentuk *dataset* yang dimana bentuk nominal dari penjualan, rasio penjualan, aset, biaya operasional diubah kedalam bentuk kategori dengan mengambil data rata-rata dari setiap atribut yang ada sehingga didapat sebuah tabel sebagai berikut :

No	brand	Rata - rata penjualan	Rata-rata rasio penjualan	Rata - rata aset	Biaya operasional	Perpanjang kontrak?
1	sevensoul	penjualan baik	baik	aset cukup	biaya mahal	ya
2	bloop	penjualan baik	baik	aset banyak	biaya murah	ya
3	family affair	penjualan baik	baik	aset banyak	biaya murah	ya
4	hipnotik	penjualan cukup	cukup	aset cukup	biaya murah	ya
5	inspired	penjualan baik	baik	aset cukup	biaya mahal	ya
6	issue makassar	penjualan baik	cukup	aset banyak	biaya mahal	tidak
7	magnum makassar	penjualan baik	cukup	aset banyak	biaya mahal	tidak
8	magnum yogya	penjualan cukup	cukup	aset banyak	biaya mahal	tidak
9	popshop	penjualan kurang	kurang	aset cukup	biaya murah	tidak
10	urbie	penjualan cukup	cukup	aset banyak	biaya murah	ya
11	widely	penjualan kurang	kurang	aset kurang	biaya murah	tidak

Tabel 4.1 Tabel *Dataset*

Berdasarkan data pada tabel pertama diatas maka dibuat calon percabangan, adapun calon percabangan yang mungkin terjadi dari tabel pertama adalah :

Tabel 4.2 Tabel Calon Percabangan

Penentuan Calon Cabang			
1	Penjualan Baik	Penjualan Cukup	Penjualan Kurang
2	Aset Banyak	Aset Cukup	Aset Kurang
3	Rasio Baik	Rasio Cukup	Rasio Kurang
4	Biaya Mahal	Biaya Murah	

Kemudian menghitung entropi dengan menggunakan dengan rumus sebagai berikut :

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

$$Entropy(ya) = \left(-\left(\frac{6}{11}\right) * \log_2 \left(\frac{6}{11}\right)\right)$$
$$= \left(-(0.55) * \log_2 (0.55)\right)$$
$$= (-(0.55) * -0{,}87))$$
$$= 0.48$$

$$Entropy(tidak) = \left(-\left(\frac{5}{11}\right) * \log_2\left(\frac{5}{11}\right)\right)$$
$$= \left(-(0.45) * \log_2\ (0.45)\right)$$
$$= (-(0.36) * -1.14))$$
$$= 0.52$$

Jadi entropi(S) didapat sebagai berikut :

$$Entropy(S) = 0.48 + 0.52 = 1.00$$

Tabel 4.3 Hasil Penghitungan Entropi

Total Kasus	Jumlah (Ya)	Jumlah (Tidak)	Entropi (Ya)	Entropi (Tidak)	Entropi Total
11	6	5	0.48	0.52	1.00

Jadi terlebih dahulu mencari masing-masing entropi pada setiap calon cabang dengan cara sebagai berikut

$$Entropy(Penjualan\ Baik) = \left(-\left(\frac{5}{6}\right) * \log_2\left(\frac{5}{6}\right)\right) + \left(-\left(\frac{1}{6}\right) * \log_2\left(\frac{1}{6}\right)\right)$$
$$= \left(-(0.,83) * \log_2\ (0.83)\right) + \left(-(0.17) * \log_2\ (0.17)\right)$$
$$= (-0.83 * -0{,}26)) + (-0.17 * -2.58)$$
$$= 0.22 + 0.43$$
$$= 0.65$$

$$Entropy(Penjualan\ Cukup) = \left(-\left(\frac{2}{3}\right) * \log_2\left(\frac{2}{3}\right)\right) + \left(-\left(\frac{1}{3}\right) * \log_2\left(\frac{1}{3}\right)\right)$$
$$= \left(-(0.,67) * \log_2\ (0.67)\right) + \left(-(0.33) * \log_2\ (0.33)\right)$$
$$= (-0{,}67 * -0{,}58)) + (-0.33 * -1.58)$$
$$= 0.39 + 0.53$$
$$= 0.92$$

$$Entropy(Penjualan\ Kurang) = \left(-\left(\frac{0}{2}\right) * \log_2\left(\frac{0}{2}\right)\right) + \left(-\left(\frac{2}{2}\right) * \log_2\left(\frac{2}{2}\right)\right)$$
$$= \left(-(0) * \log_2\ (0)\right) + \left(-(1) * \log_2\ (1)\right)$$
$$= (-(0) * 0)) + (-(1) * 0)$$
$$= 0 + 0$$
$$= 0$$

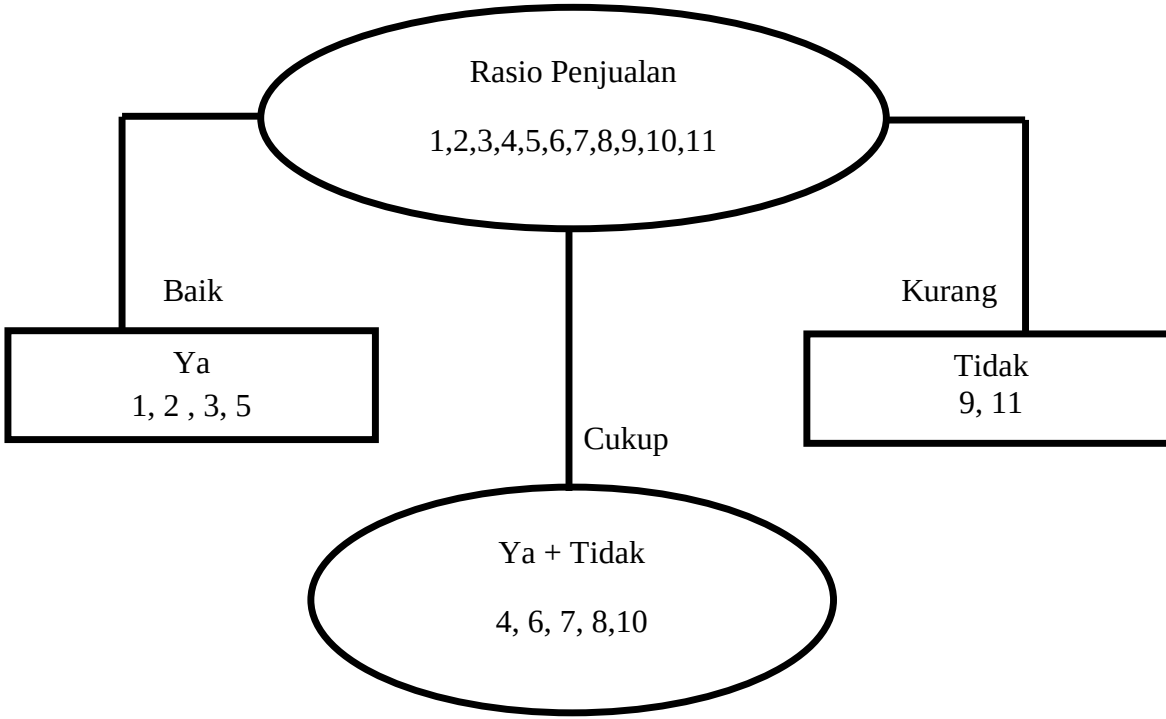
Maka *Gain(*Penjualan) dapat diperoleh dengan cara sebagai berikut :

$$Gain(S, A) = 0.95 - \left(\left(\frac{6}{11}\right) * 0.65 + \left(\frac{3}{11}\right) * 0.92 + \left(\frac{2}{11}\right) * 0\right)$$
$$= 1.00 - (0.35 + 0.25 + 0)$$
$$= 1.00 - 0.60$$
$$= 0.40$$

Tabel 4.5 Hasil Penghitungan Entropi dan Gain

Node	Kriteria	Kategori	Jumlah Kasus (S)	Ya (s1)	Tidak (s2)	Entropy	Gain
1	Total		11	7	4	0.95	
	Penjualan						
		Baik	6	5	1	0.65	0.40
		Cukup	3	2	1	0.92	
		Kurang	2	0	2	0	
	Rasio Penjualan						
		Baik	4	4	0	0	0.56
		Cukup	5	3	2	0.97	
		Kurang	2	0	2	0	
	Aset						0.21
		Banyak	6	4	2	0.92	
		Cukup	4	3	1	0.81	
		Kurang	1	0	1	0	
	Biaya Operasional						0.06
		Mahal	5	2	3	0.97	
		Murah	6	4	2	0.92	

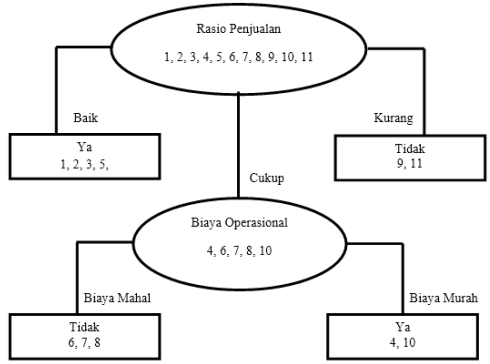
dari kolom tersebut telah diperoleh dengan nilai Gain tertinggi berjumlah 0,56 yang akhirnya rasio penjualan menjadi caban



Gambar 4.3 Pohon Keputusan Level 1

terlihat simbol berbeda yaitu simbol kotak dan simbol elips, simbol kotak ini menandakan bahwa keputusan final artinya, jika rasio penjualan baik maka penjualan akan kontrak akan diperpanjang dan jika rasio penjualannya kurang maka dapat dipastikan kontrak kerjasama tidak akan diperpanjang namun jika rasio penjualannya cukup kita tidak bisa menentukan apakah kontrak kerjasama akan diperpanjang atau tidak akan diperpanjang, karena keputusannya belum final maka dilakukan pengulangan langkah – langkah diatas dengan menghapus bagian rasio penjualan dengan nilai baik dan juga menghapus titik konsinyasi yang sudah masuk pada pohon keputusan.

Sehingga terbentuk suatu pohon keputusan akhir seperti gambar berikut



Gambar 4.4 Pohon Keputusan Akhir

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dengan menggunakan Algoritma C4.5, dimana menghasilkan suatu pohon keputusan dengan aturan yang telah dibuat berdasarkan pohon keputusan yang sama maka kesimpulan yang didapat adalah :

1. Metode *data mining* algoritma C4.5 ini dapat membantu dalam menentukan kontrak kerjasama setiap titik konsinyasi.
2. Dengan menggunakan Algoritma C4.5 ini dapat diterapkan kedalam bentuk software yang menghasilkan rekomendasi dalam menentukan kontrak kerjasama konsinyasi.

5.2 Saran

Semua hasil yang telah dicapai saat ini, tentunya masih banyak kekurangan yang harus diperbaiki atau ditambah. Saran-saran terhadap penggunaan sistem yang telah dibuat adalah sebagai berikut :

1. Perlu adanya data yang lebih komplex agar keakuratan prediksi semakin tinggi.
2. Perlu adanya pengembangan lagi agar pohon keputusan ini bisa memprediksi penjualan yang lebih tepat, terukur dan akurat.

DAFTAR PUSTAKA

[1] Mabrur, Angga Ginanjar dan Lubis, Riani (8 Mei 2015), "Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit" [online], 2012, *Available*:http://komputa.if.unikom.ac.id/_s/data/jurnal/volume-1/komputa-1-1-penerapan-data-mining-riani-8.pdf/pdf/komputa-1-1-penerapan-data-mining-riani-8.pdf

[2] Hamidah, Ida (31 Mei 2015), "Aplikasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5" [online], 2012, Available:http://elib.unikom.ac.id/gdl.php?mod=browse&op=read&id=jbptunikomp-gdl-idahamidah-29367

[3] Ikhsan, m. R. (2011). Pegaruh Kualitas Pelayanan Dan Promosi Penjualan Terhadap Loyalitas Pelanggan Indra Celllular Kudus. Semarang.