

Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5

Eki Ruziqa Maris

Program Studi Teknik Informatika, Fakultas Ilmu Komputer
 Universitas Dian Nuswantoro Semarang
 Jl. Nakula I No 5-11 Semarang 50131
 Telp : (024) 3517361, Fax : (024) 3520165
 Email : ruziqa922@gmail.com

Abstrak

Kepuasan pelanggan Telkomsel merupakan suatu keadaan dimana keinginan, harapan dan kebutuhan pelanggan terpenuhi. Penelitian ini bertujuan untuk mengetahui dan menganalisis kepuasan pelanggan telkomsel di Kota Semarang. Penelitian ini menggunakan kuesioner sebanyak 150 responden.

Pada tugas akhir ini dilakukan sebuah penelitian agar dapat mengetahui seberapa besar kepuasan pelanggan menggunakan kartu Telkomsel di Kota Semarang menggunakan penerapan ilmu data mining. Metode yang digunakan adalah Algoritma C4.5.

Algoritma C4.5 merupakan salah satu teknik terkenal dalam data mining karena metode ini tidak memerlukan proses pengelolaan pengetahuan terlebih dahulu dan dapat menyelesaikan dengan sederhana kasus-kasus yang memiliki dimensi yang besar. Akurasi sangat baik asalkan data yang akan dijadikan patokan merupakan data yang akurat.

Proses evaluasi dilakukan dengan menghitung akurasi, precision, recall dan f-measure. Hasil menunjukkan bahwa data 70% : 30% akurasi 87% dan data 90% : 10% akurasi meningkat menjadi 93%.

Kata Kunci : pohon keputusan, data mining, kepuasan pelanggan, C4.5

Abstract

Telkomsel Customer satisfaction is a situation where desires, expectations and needs of customers are met. This study aims to determine and analyze customer satisfaction Telkomsel in Semarang. This study used a questionnaire of 150 respondents.

In this thesis conducted a study in order to find out how much customer satisfaction using Telkomsel card in Semarang using data mining application of science. The method used is a C4.5 algorithm.

C4.5 algorithm is a well-known technique in data mining because this method does not require prior knowledge management process and may resolve with simple cases which have large dimensions. Accuracy is very good as long as the data that will be used as a benchmark is accurate data.

The evaluation process is done by calculating the accuracy, precision, recall and F-measure. The results show that the data is 70% : 30% accuracy 87% and 90% of data: 10% accuracy increased to 93%.

Keywords : decision tree , data mining , customer satisfaction , C4.5

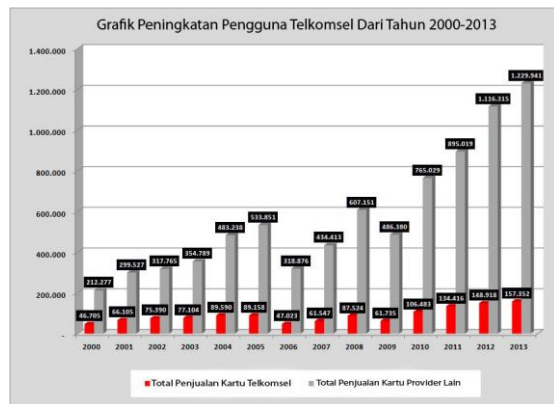
PENDAHULUAN

1. Latar Belakang

Perangkat elektronik khususnya di bidang telekomunikasi terlihat sangat jelas semakin berkembang pesat. Berbagai inovasi dimunculkan demi mencapainya kemudahan dalam penggunaannya. Sampai saat ini manusia di berikan berbagai media dan salah satunya adalah dengan perkembangan ponsel yang dinilai paling mengalami masa transisi paling cepat dalam satu dekade terakhir. Salah satu contoh adalah dalam dunia komunikasi. Surat menyurat tidaklah lagi menjadi satu-satunya jalan untuk melakukan komunikasi jarak jauh, handphone

menjadi pilihan utama, selain bisa digunakan untuk komunikasi jarak jauh, dapat dibawa kemana-mana, handphone juga memiliki banyak kegunaan lain. Pemakaiannya semakin bertambah luas, boleh dikatakan setiap orang memakainya, seperti kebutuhan pokok, dengan kondisi semacam ini makin banyak pilihan kartu SIM Card yang ditawarkan oleh penyedia jaringan telekomunikasi. Kondisi semacam ini memaksa perusahaan produsen telekomunikasi untuk melakukan strategi yang tepat untuk mempertahankan produk mereka di pasaran. Salah satu perusahaan produsen telekomunikasi yang terbesar di Indonesia saat ini adalah Telkomsel. Karena jumlah

pelanggan yang semakin banyak dan muncul pesaing seperti terlihat di gambar 1.1, maka perlukiranya perusahaan mengenal seperti apa kelebihan ataupun kekurangan produknya di pasaran oleh perusahaan Telkomsel. Fungsi, kualitas dan benefits dari suatu produk merupakan fokus perhatian konsumen [1].



Gambar : Grafik Peningkatan Pengguna Telkomsel

Seorang konsumen memungkinkan mengalami berbagai tingkat kepuasan yaitu bila produk tidak sesuai dengan harapan setelah dikonsumsi, maka konsumen tersebut akan merasa tidak puas. Namun bila terjadi sebaliknya yaitu produk sesuai dengan harapannya, maka konsumen merasakan puas sehingga suatu saat mengkonsumsi kembali produk tersebut. Konsumen yang puas adalah konsumen yang akan berbagai kepuasan dengan produsen [2].

Dalam penelitian ini hasil survey kepuasan konsumen tersebut juga akan dianalisis menggunakan pendekatan data mining dengan algoritma C4.5 dikarenakan algoritma ini secara umum lebih cepat, mudah diinterpretasikan dan dipahami pengguna. Di samping itu, berdasarkan beberapa penelitian menunjukkan bahwa performansi lebih konsisten dibandingkan beberapa algoritma klasifikasi jenis pohon keputusan lainnya [2].

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numeric dan diskret, dapat

menangan ini lain atribut yang hilang, yang menghasilkan aturan-aturan mudah diinterpretasikan dan tercepat diantara algoritma yang lainnya. Keakuratan prediksi yaitu kemampuan model untuk dapat memprediksi label kelas terhadap data-data yang baru ataupun data yang belum diketahui sebelumnya dengan baik [2].

Liliana Swastina [3] melaporkan bahwa penelitian menggunakan algoritma C4.5 untuk penerapan penentuan jurusan mahasiswa. Anik Andriani [4] juga melaporkan bahwa penelitiannya menggunakan algoritma C4.5 untuk penerapan algoritma C4.5 pada program klasifikasi mahasiswa dropout dengan hasil baik. R.A. Nurlinda [5] juga melaporkan bahwa penelitiannya tentang kepuasan/ketidakpuasan pelanggan sebagai respon pelanggan terhadap evaluasi ketidaksesuaian yang dipersepsikan antara harapan awal sebelum pembelian dan kinerja aktual produk yang dirasakan setelah pemakaiannya. Selanjutnya Yoan Santosa Putra [6] juga melaporkan bahwa penelitiannya tentang menguji pengaruh variabel independen yaitu bukti fisik, kehandalan, daya tanggap, jaminan, dan empati terhadap kepuasan pelanggan dengan hasil baik.

Mengingat atribut harga, kekuatan sinyal, dan pelayanan sangat penting dalam memberikan kepuasan pada pelanggan, maka tertarik untuk menulis Tugas Akhir dengan judul **“Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5”**.

Rumusan Masalah

Berdasarkan uraian pada latar belakang yang telah dibahas maka dapat diambil suatu rumusan masalah sebagai berikut :

1. Bagaimana menganalisis kepuasan pelanggan menggunakan algoritma C4.5?
2. Berapa akurasi algoritma menggunakan C4.5 untuk prediksi kepuasan pelanggan?

Batasan Masalah

Agar tidak terlepas dari maksud dan tujuan dari disusunnya laporan ini, maka membatasi pokok permasalahan tentang harga, kekuatan sinyal, dan pelayanan terhadap kepuasan pelanggan melalui pendekatan data mining menggunakan algoritma C4.5. Data diambil dari Telkomsel.

Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan penelitian adalah :

1. Menganalisis kepuasan pelanggan menggunakan algoritma C4.5.
2. Akurasi algoritma menggunakan C4.5 untuk prediksi kepuasan pelanggan.

Maanfaat Penelitian

Bagi Penulis

Bisa menambah pengetahuan khususnya berhubungan dengan kepuasan pelanggan dilihat dari faktor harga, kekuatan sinyal, dan pelayanan sebagai sarana untuk penerapan ilmu pengetahuan yang diperoleh pada saat perkuliahan.

Bagi Manajemen Perusahaan

Memberikan masukan untuk pengembangan berbagai kebijakan pemasaran untuk meningkatkan kepuasan pelanggan.

Bagi Universitas

Memberikan tambahan sumbangsih perpustakaan khususnya yang berhubungan dengan perilaku konsumen dan kepuasan pelanggan. Selain itu juga sebagai salah satu sarana untuk mengukur sejauh mana pemahaman mahasiswa terhadap materi yang telah diberikan.

METODE PENELITIAN

Tahapan-Tahapan Algoritma C4.5

Berikut ini adalah uraian langkah-langkah dalam Algoritma C4.5 untuk menyelesaikan kasus penentuan kepuasan pelanggan puas atau tidak puas. Sebagian data yang ada pada Tabel 3.1, akan digunakan untuk membentuk pohon keputusan.

Pada Tabel 3.2, atribut-atribut adalah Harga, Kekuatan Sinyal, dan Pelayanan. Sedangkan pada kolom Hasil yaitu “Ya” dan “Tidak”. Kemudian langkah tersebut data dianalisis, dataset tersebut memiliki 10 kasus yang terdiri dari 5 “Ya” dan 5 “Tidak”.

Tabel 3.1 Data Training Kepuasan Pelanggan

Nama Pelanggan	Harga	Kekuatan Sinyal	Pelayana n	Hasil
Rocky Ervin	3	5	4	Ya
Edi Bajang	3.5	3	3.2	Ya
Syaiful Luqman	3.2	2.5	3.2	Tidak
Wiwit Kurniawan	2.7	4	3.2	Tidak
Ruly Yahya	2.8	3.5	2.6	Tidak
Joko Romadhon	3.2	4	3.6	Tidak
Sava Randu	3.2	5	4.8	Ya
Doni Cahya	3.5	2	3.8	Tidak
Sarmin	4.7	4	5	Ya
Slamet Mitulegi	4.3	5	5	Ya

$$\text{Jadi Entropy } (S) = \left(-\left(\frac{5}{10}\right) \times \log_2\left(\frac{5}{10}\right)\right) + \left(-\left(\frac{5}{10}\right) \times \log_2\left(\frac{5}{10}\right)\right) = 1$$

Tabel 3.2 Perhitungan Entropy

Total Kasus	Sum (Ya)	Sum (Tidak)	Entropi Total
10	5	5	1

Setelah mendapatkan entropy dari keseluruhan kasus, lakukan analisis pada setiap atribut, nilai-nilainya, dan hitung entropy seperti yang ditampilkan pada Tabel 3.3.

Analisis Entropy dan Gain

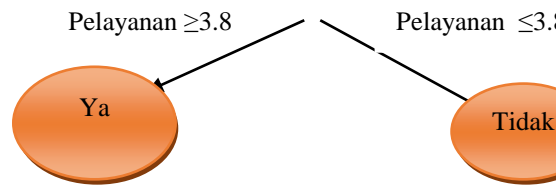
Jadi :

Node	Atribut		Tota 1	Sum	Sum	Entropy	Gain
				(Ya)	(Tidak)		
1	Harga	2.7	1	0	1	0	0.69304183
		2.8	1	0	1	0	
		3	1	1	0	0	
		3.2	3	1	2	0.35652722	
		3.5	2	1	1	1	
		4.3	1	1	0	0	
		4.7	1	1	0	0	
	Kekuatan Sinyal	2	1	0	1	0	0.89304183
		2.5	1	0	1	0	
		3	1	1	0	0	
		3.5	1	0	1	0	
		4	3	1	2	0.35652722	
		5	3	3	0	0	
	Pelayanan	2.6	1	0	1	0	0.89304183
		3.2	3	0	3	0	
		3.6	1	0	1	0	
		3.8	1	0	1	0	
		4	1	1	0	0	
4.8		1	1	0	0		
5		2	2	0	0		

$$Gain (Asal) = 1 - \left(\left(\frac{1}{10} \right) \times 0 + \left(\frac{1}{10} \right) \times 0 + \left(\frac{1}{10} \right) \times 0 + \left(\frac{3}{10} \right) \times 0.35652722318 + \left(\frac{2}{10} \right) \times 1 + \left(\frac{1}{10} \right) \times 0 + \left(\frac{1}{10} \right) \times 0 \right)$$

$$= 0.69304183304$$

Hitung pula Gain (Kekuatan Sinyal), dan Gain (Pelayanan). Hasilnya dapat dilihat pada Tabel 3.4. Karena nilai gain terbesar adalah Gain (Pelayanan). Maka Pelayanan menjadi node akar (root node).



Gambar 3.1 Pohon Keputusan Node 1

Instrumental Penelitian

Berdasarkan penelitian yang dibuat untuk menyelesaikan tugas akhir ini, instrument peralatannya meliputi [17]:

Kebutuhan Perangkat Lunak

- a. Sistem Operasi Microsoft Windows Seven
 - Sistem operasi yang digunakan pada penelitian ini minimal yaitu Microsoft Windows Seven, karena pada sistem operasi ini bahasa pemrograman matlab sudah dapat dijalankan.
- b. Matlab

Software ini dapat digunakan untuk menerapkan metode yang akan digunakan dalam penelitian tugas akhir ini.

Kebutuhan Perangkat Keras

Selain itu kebutuhan perangkat lunak, juga diperlukan perangkat keras yang dibutuhkan dalam penyelesaian

Untuk menghitung gain setiap atribut rumusnya adalah :

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

penelitian ini. Personal Computer dengan spesifikasi minimal :

- a. Prosesor Intel(R) Core(TM) i3 CPU M 350 @ 2,27GHz 2,27GHz
- b. Harddisk 320 GB
- c. RAM 2 GB
- d. 32-bit Operating System

Metode Pengumpulan Data

Metode yang digunakan untuk mendapatkan data-data yang dibutuhkannya itu metode studi kuesioner. Kuesioner yaitu teknik pengumpulan data yang dilakukan dengan cara memberikan seperangkat pertanyaan atau pernyataan kepada orang lain yang dijadikan respon dan untuk dijawabnya. Beberapa hal yang perlu diperhatikan dalam penyusunan kuesioner, terkait dengan prinsip an kuesioner, prinsip pengukuran dan penampilan fisik [18] :

1. Isi dan tujuan pertanyaan artinya jika isi pertanyaan ditujukan untuk mengukur maka harus ada skala yang jelas dalam pilihan jawaban.
2. Bahasa yang digunakan harus disesuaikan dengan kemampuan responden. Tidak mungkin menggunakan bahasa yang penuh istilah-istilah bahasa inggris pada responden yang tidak mengerti bahasa inggris.
3. Tipe dan bentuk pertanyaan apakah terbuka atau tertutup. Jika terbuka artinya jawaban yang diberikan adalah bebas, sedangkan jika pernyataan tertutup maka respon dan hanya diminta untuk memilih jawaban yang disediakan.

Kuesioner ini dibuat pada tahun 2014 di Kota Semarang. menyebarkan kuesioner sebanyak kurang lebihnya 150 kuesioner untuk dibagikan kepada para pengguna yang ada di Kota Semarang. Atribut yang digunakan dalam pembuatan kuesioner ini yaitu harga, kekuatan sinyal, dan pelayanan.

Teknik Analisis Data

Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan, yaitu uji kepuasan pelanggan pengguna yang digunakan untuk menguji seberapa puaskah pengguna menggunakan kartu . Metode yang diusulkan untuk proses seperti yang telah dijelaskan diatas yaitu metode algoritma C4.5. Prosedur perhitungan kepuasan pelanggan yang diusulkan sebagai berikut :

Keterangan :

- a. Langkah pertama yaitu membuat kuesioner kemudian menyebarkan kuesioner kepada para pelanggan di Kota Semarang sebanyak 150 kuesioner. Dari kuesioner tersebut, bisa mengambil data-data yang diperlukan, kemudian data-data tersebut akan dijadikan sebagai acuan untuk pengambilan keputusan.
- b. Data-data tersebut diolah menggunakan Microsoft Excel dengan cara mencari rata-rata setiap atributnya.
- c. Menerapkan perhitungan Decision Tree C4.5 pada kepuasan pelanggan di Kota Semarang.
- d. Setelah itu dilakukan pembuatan pohon keputusan dengan Matlab menggunakan data-data pelanggan yang sudah diolah tersebut.
- e. Kemudian dilakukan pengujian sebanyak 2 kali, untuk mengetahui tingkat keakuratan dari setiap percobaan-percobaan yang sudah diuji.
- f. Untuk bisa mengetahui keakuratan maka dilakukan perhitungan akurasi.
- g. Evaluasi.

HASIL DAN PEMBAHASAN

Proses Pengelompokan Data

Dalam penelitian ini, yang pertama yaitu mengumpulkan data dengan cara membuat kuesioner, kurang lebihnya 150

yang akan dibagikan kepada para pelanggan yang berada di Kota Semarang.

Setelah kuesioner itu sudah jadi, maka kuesioner tersebut akan dibagikan kepada para pelanggan yang berada di Kota Semarang. Setelah kuesioner tersebut sudah terkumpul, kemudian mengolah data-data tersebut dengan menggunakan Microsoft Office Excel seperti yang terlihat pada gambar dibawah ini :

No.	Nama	x1 = Harga					Rata-rata	x2 = Kekuatan Sinyal					Rata-rata	x3 = Pelayanan					Rata-rata	Y
		1	2	3	4	5	6	x1	7	8	9	10	11	12	13	x3				
1	Kukuh Supriyatno	4	4	3	3	4	4	3,66666667	4	3	3,5	4	4	4	4	3	3,8	Ya		
2	Edi Bajang	4	3	5	2	4	3	3,5	3	3	3	4	3	3	4	3	3,4	Ya		
3	Harjo Minangun	5	5	5	5	5	3	4,66666667	3	5	4	5	5	5	5	5	5	Ya		
4	Suparnasih	5	5	5	4	1	1	3,5	4	5	4,5	5	5	5	4	4	4,6	Ya		
5	Guntur Marsidi	5	3	3	2	4	4	3,5	4	5	4,5	4	4	5	4	4	4,2	Ya		
6	Siti Muttupiah	5	5	5	5	4	4	4,66666667	5	5	5	5	5	5	5	5	5	Ya		
7	Muh. Canggih Orianto	5	3	5	2	3	3	3,5	5	5	5	3	3	5	5	5	4,2	Ya		
8	Siti Sholikhah	4	4	3	4	4	2	3,5	2	2	2	3	3	3	3	2,4	Tidak			
9	Abdul Haris	4	4	4	5	4	4	4,16666667	4	3	3,5	3	3	3	3	4	3,2	Ya		
10	Imawan	3	2	3	2	3	2	2,5	3	2	2,5	2	2	2	2	4	2,4	Tidak		
11	Nur Sida	4	4	4	3	4	3	3,66666667	3	3	3	3	5	5	4	4	4,2	Ya		
12	Alfa Rizal Akbar Murwanda	5	4	5	3	1	1	3,16666667	5	5	5	5	5	4	4	5	4,6	Ya		
13	Muh. Samsul Bakri	4	5	5	5	1	1	3,5	5	5	5	5	5	5	3	3	4,2	Ya		
14	Darus	4	4	4	4	4	4	4	3	3	3	4	3	3	3	3,2	Ya			
15	Sarmin	5	5	5	5	5	3	4,66666667	3	5	4	5	5	5	5	5	5	Ya		
16	Slamet Mihallegi	5	5	5	5	3	3	4,33333333	5	5	5	5	5	5	5	5	5	Ya		
17	Solikin Huda	5	5	5	5	3	3	4,33333333	5	5	5	5	5	5	5	5	5	Ya		
18	Darmadi Jamal	5	5	5	5	3	3	4,33333333	5	5	5	5	5	5	5	5	5	Ya		
19	Kresndi	4	4	4	4	4	4	4	3	3	3	4	4	4	3	3,8	Ya			
20	Novita Sari	4	3	3	4	3	3	3,33333333	2	2	2	3	3	3	2	3	2,8	Ya		
21	Linda Silvia	4	3	3	3	4	3	3,33333333	2	2	2	3	3	2	2	3	2,6	Ya		
22	Rully Yalrya	4	2	4	3	2	2	2,83333333	4	4	4	4	4	4	3	3,8	Ya			
23	Wisnu Wijanarko	3	3	3	4	4	3	3,33333333	3	3	3	3	4	4	4	4	3,8	Ya		
24	Indra Ridunika	4	4	3	3	3	4	3,5	4	4	4	3	3	4	4	3	3,4	Ya		
25	Dewi Cahya M.	4	3	4	4	4	4	3,83333333	4	4	4	3	3	4	3	3	3,2	Ya		
26	Ervanda L I I	4	4	4	3	2	2	3,16666667	5	5	5	4	3	5	3	3,6	Ya			
27	Miliah NurChaidah	5	3	5	3	3	3	3,66666667	5	5	5	5	4	4	4	4,2	Ya			
28	Rina Agustiana	5	3	3	2	2	2	2,83333333	5	5	5	4	4	5	4	4,4	Ya			
29	Yenny Laila	5	3	3	3	3	3	3,5	5	5	5	4	4	5	4	4,4	Ya			

Data kuesioner

Setelah data sudah diolah menggunakan Microsoft Office Excel, maka dilakukan lagi pengolahan data kuesioner tersebut dengan cara mengambil data rata-rata x1 (Harga), rata-rata x2 (Kekuatan Sinyal), rata-rata x3 (Pelayanan) dan Y (Hasil).

No.	Nama	Rata-rata	Rata-rata	Rata-rata	Y
		x1	x2	x3	
1	Kukuh Supriyatno	3,7	3,5	3,8	Ya
2	Edi Bajang	3,5	3	3,4	Ya
3	Harjo Minangun	4,7	4	5	Ya
4	Sujarningsih	3,5	4,5	4,6	Ya
5	Guntur Marsidi	3,5	4,5	4,2	Ya
6	Siti Muttupiah	4,7	5	5	Ya
7	Muh. Canggih Orianto	3,5	5	4,2	Ya
8	Siti Sholikhah	3,5	2	2,4	Tidak
9	Abdul Haris	4,2	3,5	3,2	Ya
10	Imawan	2,5	2,5	2,4	Tidak
11	Nur Sida	3,7	3	4,2	Ya
12	Alfa Rizal Akbar Murwanda	3,2	5	4,6	Ya
13	Muh. Samsul Bakri	3,5	5	4,2	Ya
14	Darus	4	3	3,2	Ya
15	Sarmin	4,7	4	5	Ya
16	Slamet Mihallegi	4,3	5	5	Ya
17	Solikin Huda	4,3	5	5	Ya
18	Darmadi Jamal	4,3	5	5	Ya
19	Kresndi	4	3	3,8	Ya
20	Novita Sari	3,3	2	2,8	Ya
21	Linda Silvia	3,3	2	2,6	Ya
22	Rully Yalrya	2,8	4	3,8	Ya
23	Wisnu Wijanarko	3,3	3	3,8	Ya
24	Indra Ridunika	3,5	4	3,4	Ya

Data Kuesioner Yang Sudah Diolah

Proses Data Mining

Setelah dijelaskan proses penerapan data mining dengan teknik *decision tree* secara teoritis pada bab-bab sebelumnya, maka disini pula akan dijelaskan proses tentang *data mining* secara aplikatif dimana proses data mining yang akan dilakukan menggunakan *Software Matlab*. Sebelum data diproses di Matlab akan dijelaskan secara singkat bagaimana cara perhitungan secara manual kepuasan pelanggan . Data pelanggan tercantum pada lampiran.

Proses Data Mining Pada Matlab

Data pelanggan disimpan di notepad. Hanya data dari parameter Harga, Kekuatan Sinyal dan Pelayanan yang dicantumkan. File disimpan dengan nama data pelanggan.

THEN Hasil = Tdk
7. IF kekuatan sinyal ≥ 2.75 and pelayanan ≥ 2.7 and harga ≥ 2.75 and pelayanan < 3.3 and harga ≥ 4.25 THEN Hasil = Tdk
Rules untuk Hasil = Iya
1. IF kekuatan sinyal < 2.75 and harga ≥ 2.9 and harga < 3.4 and harga < 3.25 and kekuatan sinyal ≥ 1.75 and pelayanan < 3.9 THEN Hasil = Iya
2. IF kekuatan sinyal < 2.75 and harga ≥ 2.9 and harga < 3.4 and harga ≥ 3.25 and kekuatan sinyal < 2.25 THEN Hasil = Iya
3. IF kekuatan sinyal ≥ 2.75 and pelayanan > 2.7 and harga < 2.75 THEN Hasil = Iya
4. IF kekuatan sinyal ≥ 2.75 and pelayanan ≥ 2.7 and harga ≥ 2.75 and pelayanan < 3.3 and harga < 4.25 THEN Hasil = Iya
5. IF kekuatan sinyal ≥ 2.75 and pelayanan ≥ 2.7 and harga ≥ 2.75 and pelayanan ≥ 3.3 THEN Hasil = Iya

Pengujian Rules Terhadap Data Pelanggan

Pengujian terhadap validitas sistem bertujuan untuk mengetahui apakah solusi yang dihasilkan oleh pohon keputusan tersebut valid atau tidak. Rules tersebut dikatakan valid jika jumlah dan nama pelanggan yang puas sesuai dengan data set.

Data set dibagi menjadi dua bagian, yaitu data training dan data testing. Pengujian dilakukan dua kali dengan jumlah data training dan data testing yang berbeda. Uraianya sebagai berikut:

- Data training 70% dan data testing 30%
- Data training 90% dan data testing 10%

Pengujian I

Data set dibagi menjadi dua 70% data training dan 30% data testing. Keseluruhan data set berjumlah 150, data training sebanyak 105 data dan data testing sebanyak 45 data.

Tabel 4.1 Data training 70%

Nama Pelanggan	Harga	Kekuatan Sinyal	Pelayanan	Harga
1	3.7	3.5	3.8	Ya
2	3.5	3	3.4	Ya
3	4.7	4	5	Ya
...
73	3.7	3	3.4	Ya
74	3.3	5	5	Ya
...
...
81	3.7	2.5	3.4	Tidak
...
99	3	2.5	4.6	Tidak
100	2.5	5	2.6	Tidak
...
...
105	3.3	4	3.8	Ya

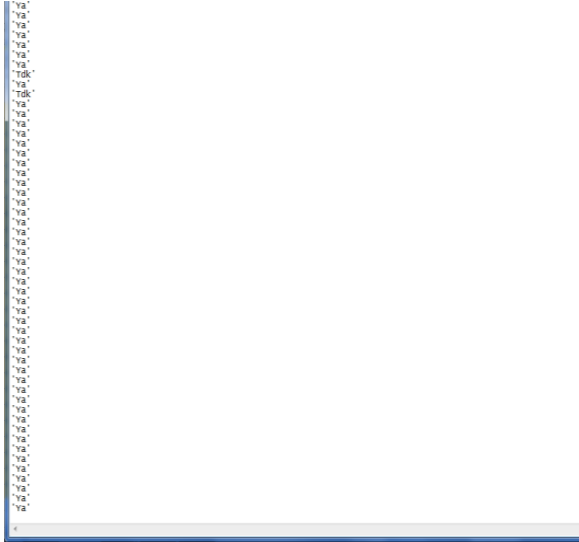
Tabel diatas merupakan tabel data pelanggan yang akan di training menggunakan Software Matlab. Data training berjumlah 98 data pelanggan. Kemudian data tersebut di buat di notepad dengan parameter harga, kekuatan sinyal, dan pelayanan. Data tersebut disimpan dengan nama data1.

The screenshot shows a Notepad window with a menu bar (File, Edit, Format, View, Help) and a list of data points. Each line contains three numerical values representing price, signal strength, and service, respectively. The data points are as follows:

Price	Signal Strength	Service
3.3	2.5	2.8
3.2	2.5	3.2
3.7	4.0	3.2
2.5	2.5	2.6
3.5	4.0	2.6
3.5	4.0	3.4
3.3	3.0	4.0
3.3	4.0	3.4
3.2	3.0	3.6
3.5	5.0	4.8
3.2	5.0	4.0
3.2	5.0	4.6
3.0	2.0	2.6
2.8	3.5	2.8
3.5	4.5	3.4
3.8	3.5	3.8
3.7	3.0	3.4
3.3	5.0	5.0
3.8	2.5	4.2
3.2	1.5	3.2
3.7	3.5	3.4
3.5	4.0	3.0
3.0	4.5	2.0
4.3	4.0	3.2
3.7	2.5	3.4
3.5	4.0	3.4
3.3	3.5	3.8
3.5	5.0	4.4
3.5	4.0	4.4
3.2	3.5	4.0
3.5	4.0	2.8
3.5	3.0	3.0
3.8	4.5	4.0
3.5	5.0	3.8
3.5	4.0	3.4
3.8	5.0	4.2
3.3	1.5	3.2
3.0	4.0	3.0
3.2	2.0	4.0
3.7	3.0	2.6
3.7	3.0	3.6
4.3	5.0	5.0
3.0	2.5	4.6
3.5	3.0	2.6
3.0	5.0	3.4
4.0	4.0	3.4
3.8	3.0	3.0

Data1

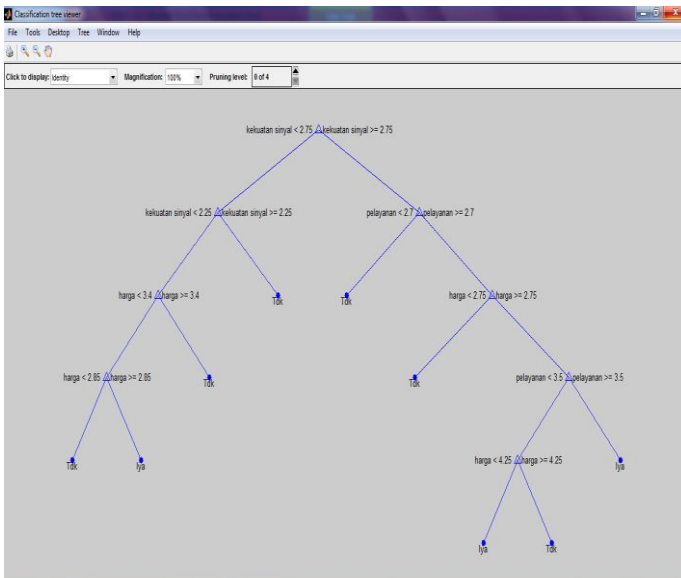
Kemudian data dari parameter Hasil tersebut disimpan di notepad. File tersebut disimpan dengan nama hasil1. Pada penulisan data di notepad diberi tanda petik satu (') karena merupakan String.



Hasil1

Berikut adalah pohon keputusan yang dihasilkan dari

Data training :



Pohon Keputusan Data Training 70%

Pohon keputusan diatas menghasilkan *rules* sebagai berikut :

Rules untuk Hasil = Tdk

1. IF kekuatan sinyal <2.75 and kekuatan sinyal <2.25 and harga <3.4 and harga <2.85 THEN Hasil = Tdk
2. IF kekuatan sinyal <2.75 and kekuatan sinyal <2.25 and harga >=3.4 THEN Hasil = Tdk
3. IF kekuatan sinyal <2.75 and kekuatan sinyal >=2.25 THEN Hasil = Tdk
4. IF kekuatan sinyal >=2.75 and pelayanan <2.7 THEN Hasil = Tdk
5. IF kekuatan sinyal >=2.75 and pelayanan >=2.7 and harga <2.75 THEN Hasil = Tdk
6. IF kekuatan sinyal <3.5 and harga >=4.25 THEN Hasil = Tdk

Rules untuk Hasil = Iya

1. IF kekuatan sinyal <2.75 and kekuatan sinyal <2.25 and harga <3.4 and harga >=2.85 THEN Hasil = Iya
2. IF kekuatan sinyal >=2.75 and pelayanan >=2.7 and harga >=2.75 and pelayanan <3.5 and harga <4.25 THEN Hasil = Iya
3. IF kekuatan sinyal >=2.75 and pelayanan >=2.7 and harga >=2.75 and pelayanan >=3.5 THEN Hasil = Iya

Setelah dilakukan training terhadap 105 data tersebut, selanjutnya dilakukan *testing data* terhadap *data set* yang tersisa yaitu 45 data pelanggan.

Data testing 30%

Nama Pelanggan	Harga	Kekuatan Sinyal	Pelayanan	Hasil	Predicted
106	3.2	5	4.4	Ya	Ya
...
115	4.3	5	3.4	Ya	Tidak
...
120	2.3	5	3.4	Ya	Tidak
121	2.8	5	3.8	Ya	Ya
...
129	3.2	1.5	3	Tidak	Ya
...
141	3	1.5	2	Tidak	Ya
...
146	3.2	2.5	3.2	Ya	Tidak
..
148	3.5	2.5	3.4	Ya	Tidak
...
150	3	2.5	3	Tidak	Tidak

Pada Tabel 4.5 menunjukkan bahwa ada 6 data yang hasil prediksinya tidak tepat, ditunjukkan dengan kolom yang berwarna biru.

Hasil Testing 1

Jumlah Data testing	Error	Positif Benar (TP)	Positif Salah (FP)	Negatif Salah (FN)
45	6	27	2	4

Pada tabel di atas data dengan perbandingan 70% : 30% terdapat *error* sebanyak 6 data, positif benar 27, positif salah 2 dan negatif salah 4.

Evaluasi

Presentasi Data	Data Training	Data Testing	Akurasi	Precision	Recall	data. Measure
70% : 30%	105	45	87%	93%	87%	90%

Pada tabel di atas data dengan perbandingan 70% : 30% memiliki nilai akurasi sebesar 87%, *recall* sebesar 87%, *precision* 93% dan *f-measure* 90%.

Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total sample *data testing* yang diuji.

$$\text{Akurasi} = \frac{\text{Data Testing} - \text{Error}}{\text{Jml Data testing}} = \frac{45 - 6}{45} = \frac{0.866666666666}{1} = 87\%$$

Nilai *precision* dihitung dengan cara membagi jumlah data benar yang bernilai positif (*True Positive*) dibagi dengan jumlah data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai positif (*False Negative*).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{27}{27 + 2} = 0.931034482759 = 93\%$$

Sedangkan *recall* dihitung dengan cara membagi data benar yang bernilai positive (*True Positive*) dengan hasil penjumlahan dari data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai negatif (*False Negative*).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{27}{27 + 4} = 0.870967741935 = 87\%$$

Nilai *F-Measure* didapat dari perhitungan pembagian hasil dari perkalian *precision* dan *recall* dengan hasil penjumlahan *precision* dan *recall*, kemudian dikalikan dua.

$$\text{F-Measure} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \cdot \frac{0.93 \cdot 0.87}{0.93 + 0.87} = 0.9 = 90\%$$

Pengujian II

Data set dibagi menjadi dua 90% *data training* dan 10% *data testing*. Keseluruhan *data set* berjumlah 150, *data training* sebanyak 135 data dan *data testing* sebanyak 15

Data training 90%

Nama Pelanggan	Harga	Kekuatan Sinyal	Pelayanan	Harga
1	3.7	3.5	3.8	Ya
2	3.5	3	3.4	Ya
3	4.7	4	5	Ya
...
...
73	3.7	3	3.4	Ya
74	3.3	5	5	Ya
...
...
...
81	3.7	2.5	3.4	Tidak
...
99	3	2.5	4.6	Tidak
100	2.5	5	2.6	Tidak
...
...
135	3.7	3.5	3.6	Ya

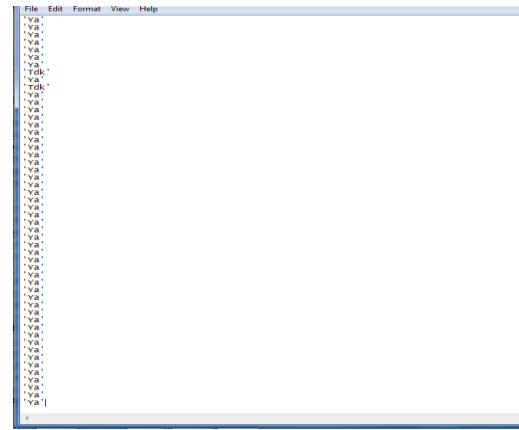
Tabel diatas merupakan tabel data pelanggan yang akan di training menggunakan. *Data training* berjumlah 135 data pelanggan. Setelah itu maka data training yang sebanyak 135 itu kemudian di simpan di notepad dengan parameter harga, kekuatan sinyal, dan pelayanan. Data tersebut disimpan dengan nama data2.

```

3.7 3.5 3.8
3.5 2.0 3.4
3.7 4.0 5.0
3.5 4.5 4.6
3.5 4.5 4.2
4.7 5.0 5.0
3.5 5.0 4.2
3.5 2.0 2.4
4.2 3.5 3.2
3.5 2.5 2.4
3.7 3.0 4.2
3.2 5.0 4.6
3.5 5.0 4.2
4.0 3.0 3.2
4.7 4.0 5.0
4.3 4.0 5.0
4.3 5.0 5.0
4.3 5.0 5.0
4.3 5.0 5.0
4.0 3.0 3.8
3.3 2.0 2.8
3.3 2.0 2.6
3.8 4.0 3.8
3.3 3.0 3.8
3.3 4.0 3.4
3.8 4.0 3.2
3.2 5.0 3.4
3.7 5.0 4.2
3.8 5.0 4.4
3.8 5.0 4.4
4.0 3.0 3.4
3.2 3.5 3.4
3.3 3.5 4.0
3.2 3.5 3.0
3.3 3.0 3.0
3.2 3.5 3.2
3.8 3.5 3.4
4.0 3.5 4.0
3.7 3.3 3.4
3.8 3.0 3.8
3.8 3.0 3.8
3.3 5.0 5.0
3.8 3.0 3.8
3.8 3.0 3.8
4.3 5.0 5.0
3.5 5.0 4.6
4.3 5.0 5.0
3.5 5.0 4.4
3.3 5.0 4.8
3.3 4.5 4.0
3.8 4.5 4.0
3.3 5.0 3.0
    
```

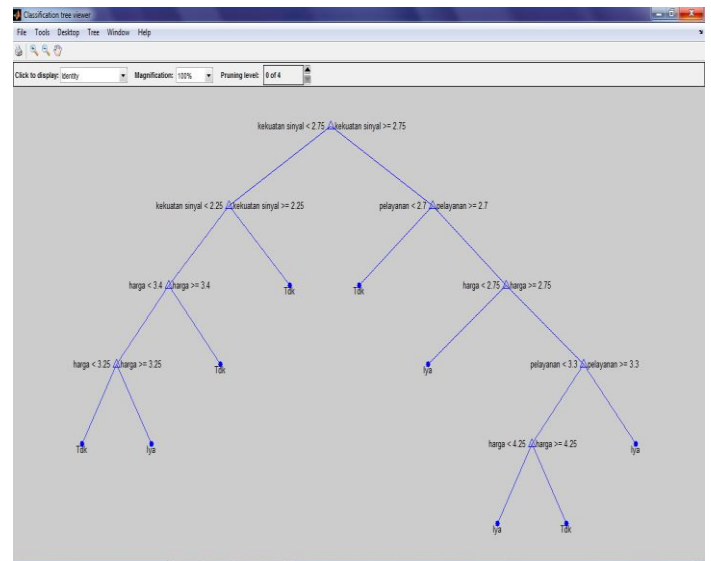
Data2

Kemudian langkah selanjutnya yaitu data dari parameter Hasil tersebut disimpan di notepad. File tersebut disimpan dengan nama hasil2. Pada penulisan data di notepad diberi tanda petik satu (') karena merupakan String



Hasil2

Berikut ini adalah pohon keputusan yang dihasilkan dari *Data training* :



Pohon Keputusan Data Training 90%

Dari pohon keputusan diatas didapat *rules* sebagai berikut :

- Rules untuk Hasil = Tdk**
1. IF kekuatan sinyal < 2.75 and kekuatan sinyal < 2.25 and harga < 3.4 and harga < 3.25 THEN Hasil = Tdk
 2. IF kekuatan sinyal < 2.75 and kekuatan sinyal < 2.25 and harga >= 3.4 THEN Hasil = Tdk
 3. IF kekuatan sinyal < 2.75 and kekuatan sinyal >= 2.25 THEN Hasil = Tdk

<p>4. IF kekuatan sinyal ≥ 2.75 and pelayanan < 2.7 THEN Hasil = Tdk</p> <p>5. IF kekuatan sinyal > 2.75 and pelayanan ≥ 2.7 and harga ≥ 2.75 and pelayanan < 3.3 and harga ≥ 4.25 THEN Hasil = Tdk</p>
Rules untuk Hasil = Iya
<p>1. IF kekuatan sinyal < 2.75 and kekuatan sinyal < 2.25 and harga < 3.4 and harga ≥ 3.25 THEN Hasil = Iya</p> <p>2. IF kekuatan sinyal ≥ 2.75 and pelayanan ≥ 2.7 and harga < 2.75 THEN Hasil = Iya</p> <p>3. IF kekuatan sinyal ≥ 2.75 and pelayanan ≥ 2.7 and harga ≥ 2.75 and pelayanan < 3.3 and harga < 4.25 THEN Hasil = Iya</p> <p>4. IF kekuatan sinyal ≥ 2.75 and pelayanan ≥ 2.7 and harga ≥ 2.75 and pelayanan ≥ 3.3 THEN Hasil = Iya</p>

Setelah dilakukan *training* terhadap 135 data tersebut, selanjutnya dilakukan *testing data* terhadap *data set* yang tersisa yaitu 15 data pelanggan.

Data testing 10%

Nama Pelanggan	Harga	Kekuatan Sinyal	Pelayanan	Hasil	Prediksi
136	3.5	4	4.6	Ya	Ya
137	4.3	5	5	Ya	Ya
...
146	3.2	2.5	3.2	Ya	Tidak
...
...
150	3	2.5	3	Tidak	Tidak

Hasil Testing 2

Jumlah Data testing	Error	Positif Benar (TP)	Positif Salah (FP)	Negatif Salah (FN)
15	1	8	0	1

Pada tabel di atas diketahui bahwa terdapat *error* atau kesalahan prediksi sebanyak 1 data, positif benar 8, positif salah 0 dan negatif salah 1.

Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total sample *data testing* yang diuji.

$$\text{Akurasi} = \frac{\text{Data Testing} - \text{Error}}{\text{Jml Data testing}} = \frac{15 - 1}{15} = \frac{14}{15} = 0.933333333333 = 93\%$$

Nilai *precision* dihitung dengan cara membagi jumlah data benar yang bernilai positif (*True Positive*) dibagi dengan jumlah data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai positif (*False Negative*).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{8}{8 + 0} = 0 = 0\%$$

Sedangkan *recall* dihitung dengan cara membagi data benar yang bernilai positive (*True Positive*) dengan hasil penjumlahan dari data benar yang bernilai positif

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{8}{8 + 1} = 0.888888888889 = 88\%$$

Nilai *F-Measure* didapat dari perhitungan pembagian hasil dari perkalian *precision* dan *recall* dengan hasil penjumlahan *precision* dan *recall*, kemudian dikalikan dua.

$$\text{F-Measure} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = 0 = 0\%$$

Evaluasi

Presentasi Data	Data Training	Data Testing	Akurasi	Precision	Recall	M
90% : 10%	135	15	93%	0%	88%	

Pada tabel di atas data dengan perbandingan 90% : 10% memiliki nilai akurasi sebesar 93%, *recall* sebesar 88%, *precision* 0% dan *f-measure* 0%.

Analisa Hasil Pengujian

Pengujian I dan II. menghasilkan akurasi, *precision*, *recall* dan *f-measure* sebagai berikut :

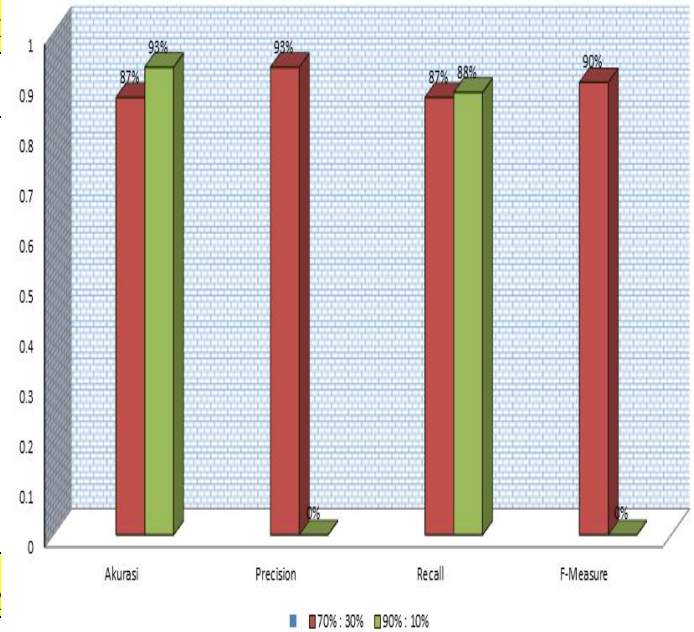
Hasil Pengujian

Persentase Data	Data Training	Data Testing	Akurasi	Precision	Recall	M
70% : 30%	105	45	87%	93%	87%	90%
90% : 10%	135	15	93%	0%	88%	0%

Tabel di atas menunjukkan bahwa nilai *precision* untuk data dengan perbandingan *data training* dan *data testing* 70% : 30% adalah 93% dan *recall* 87%. Sedangkan data perbandingan 90% : 10% memiliki nilai *precision* 0% dan *recall* 88%.

Data dengan perbandingan 70% : 30% memiliki nilai *F-Measure* 90%, sedangkan data dengan perbandingan 90% : 10% yaitu sebesar 0%.

Berikut ini adalah grafik tentang pengujian yang pertama dan kedua :



Gambar 4.12 Diagram Grafik Performance

PENUTUP

Kesimpulan

Berdasarkan hasil pengujian metode *decision tree* terhadap data pelanggan dapat disimpulkan :

1. Pengimplementasian metode *C4.5* menggunakan data pelanggan dapat digunakan untuk menentukan kepuasan pelanggan.
2. *Ratio data training* yang digunakan mempengaruhi nilai akurasi pada setiap percobaan. Pada percobaan ke- 1 nilai akurasi adalah 86% dengan *data training* 70% dan *data testing* 30%. Pada percobaan ke-2 nilai akurasi adalah 93% dengan *data training* 90% dan *data testing* 10%. Dengan hasil ini pengujian yang terbaik yaitu dengan akurasi 93%.

Saran

Adapun saran-saran yang disampaikan berdasarkan hasil pengamatan dan analisa selama melakukan penelitian data kepuasan pelanggan di adalah sebagai berikut:

- a. Penelitian selanjutnya hendaknya menggunakan data yang lebih banyak agar menghasilkan *rules* yang lebih akurat.
- b. Pengujian metode ini belum sampai pada implementasi, maka perlu dibuat sistem aplikasi.

DAFTAR PUSTAKA

- [1] Syukron, "Sejarah Telekomunikasi," Jakarta, 2012.
- [2] Bayu Adhi Tama, "Analisis Kepuasan Konsumen Terhadap Restoran Cepat Saji Melalui Pendekatan Data Mining Studi Kasus XYZ," Vol.6 No.1, pp. 25-28, 2011.
- [3] Liliana Swastina, "Penerapan Algoritma Untuk Penentuan Jurusan Mahasiswa," vol. 2 No. 1, pp. 93-98, Banjarmasin, 2013.
- [4] Anik Andriani, "Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout," pp. 139-147, Jakarta, 2012.
- [5] R.A. Nurlinda, "Customer Satisfaction Strategy Terhadap Peningkatan Kepuasan Konsumen," pp. 172-181, Jakarta, 2013.
- [6] Yoan Santosa Putra, Eris Dianawati, Endi Sarwoko "Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pelanggan Pengguna Jasa Parkir," pp.1-14, Malang.
- [7] Yudianto Tri Atmojo, "Mengukur Kepuasan Pelanggan" 2006.
- [8] Zeithaml, "Kepuasan Pelanggan," 2012l.
- [9] Andhi Wijaya, "Pengertian Dan Konsep Data Mining," 2013.
- [10] Nuqson Masykur Huda, "Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa," Semarang, 2010.
- [11] Moh Nugroho Wibowo, "Implementasi Dan Demo Pohon Keputusan ID3 Dan C4.5," 2012.
- [12] Kusriani, "Algoritma Data Mining," Yogyakarta, Andi, 2009.
- [13] Astuti Fajar, "Data Mining," Andi, Surabaya, 2009.
- [14] Santosa Budi, "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis", Yogyakarta : Graha Ilmu, 2007.
- [15] Gorunescu, Florin, "Data Mining-Concepts, Model and Techniques," 2011.
- [16] Hamid Abdul, Al-Ghamdi Abdullah, Naoman Amin, Madbouly Ayman, "A Comparative Analysis of Classification Algorithm for student college enrollment approval using data mining" pp. 1-8, 2012.
- [17] Kumar Rakesh and Badal Dharmendra, "Admission Management through Data Mining using WEKA," vol. 3, issue 10, pp. 674-678, Oct. 2013.
- [18] Hendryadi, "Metode Pengumpulan Data," Jakarta, Uma Sekaran, 2009.