

Penerapan Data Mining Untuk Menganalisis Kepuasan Konsumen Menggunakan Metode Algoritma C4.5

Sutriawaty Takalapeta,

Jurusan Teknik Informatika, FT Universitas Widyagama, Malang
e-mail : sutriawatytakalapeta@gmail.com

Abstract

Consumer satisfaction includes the difference between the level of importance and performance or perceived results, as well as a full-time evaluation in which the chosen alternatives can at least yield the same or exceed consumer expectations, while dissatisfaction can occur if the results do not meet the desired expectations of the consumer. There are 5 dimensions designed to measure the quality of service based on the difference between the expected value and the perceived performance value of the consumers, namely tangible, reliability, responsiveness, assurance, and empathy. Five dimensions are very influential in the progress of a company therefore using data mining methods can be found important trends for the company. Assessment of consumer satisfaction based on a questionnaire that has been filled by consumers. The questionnaire results will be processed using C4.5 algorithm.

The C4.5 algorithm is one of the algorithms of the classification method and produces a decision tree. The decision tree method transforms a very large fact into a decision tree that represents the rule. Rules can be easily understood with natural language. Based on the results of research that has been done, the use of C4.5 algorithm can help the resto in improving the service in accordance with the results of questionnaires that have been filled by the customers. Based on the calculation, customers are satisfied with Tangibles (Facilities), Tangibles (Tools), Tangibles (Employee Appearance). While customers are not satisfied with one variable that is Tangibles (Communication). The highest gain value is variable tangibles with value 0,378118. While the indicator of tangibles variable which has the highest entropy value is facility indicator with value 0,764205.

Keywords : *Algoritma C4.5, Data Mining, Customer Satisfaction*

Abstrak

Kepuasan konsumen mencakup perbedaan antara tingkat kepentingan dan kinerja atau hasil yang dirasakan, serta merupakan evaluasi purna beli dimana alternatif yang dipilih sekurang-kurangnya dapat memberikan hasil sama atau melampaui harapan konsumen, sedangkan ketidakpuasan dapat terjadi apabila hasil yang diperoleh tidak memenuhi harapan yang diinginkan konsumen. Terdapat 5 dimensi yang dirancang untuk mengukur kualitas pelayanan yang didasarkan pada perbedaan antara nilai harapan dengan nilai kinerja yang dirasakan oleh konsumen yaitu tangible, reliability, responsiveness, assurance, dan empathy. Lima dimensi tersebut sangat berpengaruh dalam kemajuan suatu perusahaan maka dari itu dengan menggunakan metode data mining dapat ditemukan tren-tren penting bagi perusahaan. Penilaian kepuasan konsumen berdasarkan angket yang telah diisi oleh konsumen. Hasil angket tersebut akan diolah menggunakan algoritma C4.5.

Algoritma C4.5 adalah salah satu algoritma dari metode klasifikasi dan menghasilkan sebuah pohon keputusan. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Berdasarkan hasil penelitian yang telah dilakukan, penggunaan algoritma C4.5 dapat membantu pihak resto dalam meningkatkan pelayanan sesuai dengan hasil kuisioner yang telah diisi oleh para pelanggan. Berdasarkan hasil perhitungan, pelanggan puas terhadap variabel Tangibles (Fasilitas), Tangibles (Peralatan), Tangibles (Penampilan Karyawan). Sedangkan pelanggan tidak puas terhadap satu variabel yaitu Tangibles (Komunikasi). Nilai gain tertinggi adalah variabel tangibles dengan nilai 0,378118. Sedangkan indikator dari variabel tangibles yang mempunyai nilai entropy tertinggi adalah indikator fasilitas dengan nilai 0,764205. Total kuisioner yang diisi sejumlah 15 kuisioner, 11 orang menyatakan puas dan hanya 4 orang yang menyatakan tidak puas.

Kata kunci : Algoritma C4.5, Data Mining, Kepuasan Konsumen

I. PENDAHULUAN

Pada dasarnya kepuasan konsumen mencakup perbedaan antara tingkat kepentingan dan kinerja atau hasil yang dirasakan, serta merupakan evaluasi purna beli dimana alternatif yang dipilih sekurang-kurangnya dapat memberikan hasil sama atau melampaui harapan konsumen, sedangkan ketidakpuasan dapat terjadi apabila hasil yang diperoleh tidak memenuhi harapan yang diinginkan konsumen. Masalah pelayanan sebenarnya bukanlah hal yang sulit atau rumit, tetapi apabila hal ini tidak diperhatikan maka dapat menimbulkan permasalahan bagi perusahaan.

Sementara itu, data mining adalah proses pengekstrakan informasi dari kumpulan data yang besar. Pengekstrakan informasi dilakukan berdasarkan metode data mining yang akan digunakan. Ada beberapa metode serta algoritma data mining yang digunakan untuk mengekstrak informasi antara lain: metode klasifikasi, metode asosiasi, metode clustering, metode prediksi, dan metode estimasi.

Algoritma C4.5 adalah salah satu algoritma dari metode klasifikasi algoritma C4.5 adalah algoritma yang digunakan untuk membuat pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami.

Dalam kasus ini penyusun membandingkan dengan jurnal milik Eki[1]. Pengimplementasian metode C4.5 menggunakan data pelanggan dapat digunakan untuk menentukan kepuasan pelanggan. 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%. Variabel yang paling berpengaruh dari penelitian ini adalah variabel pelayanan dengan nilai gain sebesar 1.

II. METODE PENELITIAN

2.1 Penelitian Terdahulu

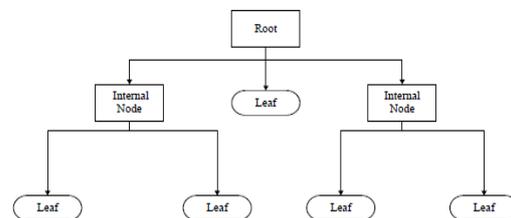
Bahan perbandingan untuk penelitian ini salah satunya oleh Tama [2] berjudul Analisis Kepuasan Konsumen Terhadap Restoran Cepat Saji Melalui Pendekatan Data Mining. Penjelasan mengenai kepuasan konsumen dalam penelitian ini, mencakup produk, pelayanan, fasilitas, harga, dan lokasi yang berasal dari data hasil survei. Dalam penelitian ini, didapatkan bahwa rules yang dibangkitkan dari beberapa atribut masukan menghasilkan hubungan sebab-akibat dalam mengklasifikasikan konsumen puas dan tidak puas. Penelitian ini diharapkan dapat membantu pihak manajemen dalam meningkatkan kepuasan konsumen untuk mempertahankan konsumen dan meningkatkan laba mereka.

2.2 Data Mining

Data mining diartikan sebagai suatu proses ekstraksi informasi berguna dan berpotensi dari sekumpulan data yang terdapat secara implisit dalam suatu basis data. Data mining meliputi tugas-tugas sebagai ekstraksi pengetahuan, arkeologi data, eksplorasi dalam pemrosesan pola data dan memanen informasi. Semua aktifitas ini dilakukan secara otomatis dan mengijinkan adanya penemuan cepat bahkan oleh programmer.[3]

2.3 Pohon Keputusan (*Decision Tree*)

Decision tree adalah flowchart seperti struktur tree, dimana tiap internal node menunjukkan sebuah test pada sebuah atribut, tiap cabang menunjukkan hasil dari test, dan leaf node menunjukkan class-class atau class distribution.[4]



Gambar 1. Model Pohon Keputusan

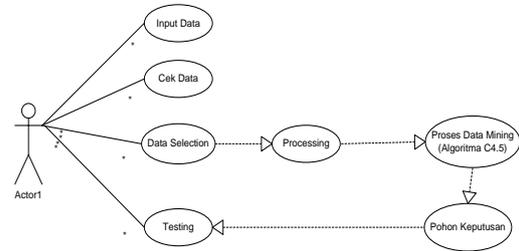
2.4 Ranking pada Kamus

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. Algoritma C4.5 adalah algoritma hasil pengembangan dari algoritma ID3. Perbaikan algoritma C4.5 dari algoritma ID3 dilakukan dalam hal [5]:

1. Bisa mengatasi missing value
2. Bisa mengatasi data kontinyu
3. Pruning
4. Adanya aturan

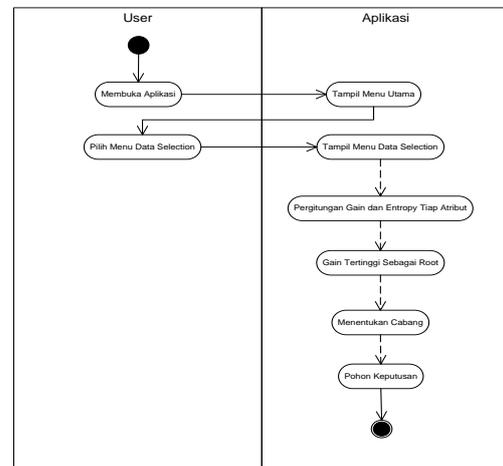
4. Melakukan pemilihan gain tertinggi yang akan dijadikan root awal
5. Penentuan cabang
6. Hasil pohon keputusan

3.2 Use Case



Gambar 3. Use Case

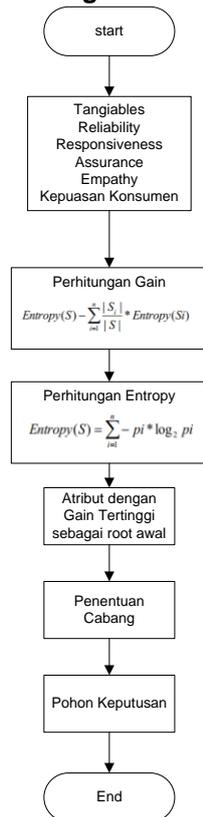
3.3 Activity Diagram Data Selection



Gambar 4. Activity Diagram

III. HASIL DAN PEMBAHASAN

3.1 Flowchart Algoritma C4.5



Gambar 2. Flowchart C4.5

1. Menentukan variabel yang akan digunakan
2. Menghitung nilai gain
3. Menghitung nilai entropy

User terlebih dahulu membuka aplikasi dan aplikasi akan merespon dengan menampilkan menu utama, user memilih menu data selection dan memilih tombol perhitungan, maka aplikasi akan merespon dengan melakukan perhitungan gain dan entropy yang hasilnya akan dijadikan rule pohon keputusan.

3.4 Halaman Login



Gambar 5. Form Menu Login

Form login digunakan untuk user masuk ke dalam halaman utama. User

memasukkan data username dan password, jika username salah atau password salah maka akan kembali ke halaman login jika username benar dan password benar, maka user akan masuk ke halaman utama.

3.5 Halaman Variabel



Gambar 6. Halaman Proses Variabel

Pada halaman ini user melakukan pengecekan terhadap data variabel yang digunakan.

3.6 Halaman Data Training



Gambar 7. Halaman Data Training

Pada halaman data training user dapat melakukan tambah data training, total data, kuisisioner, reset data dan proses data. Jika user memilih tambah data maka user akan masuk ke halaman tambah data, jika user memilih total data maka user akan melihat jumlah data yang masuk, jika user memilih reset data maka data training pada database akan dihapus. Sedangkan jika user memilih proses data maka sistem akan melakukan proses data training dengan menggunakan algoritma C4.5 untuk menganalisa kepuasan konsumen.

3.7 Halaman Algoritma C4.5

Pada halaman ini user dapat melihat hasil dari proses algoritma C4.5.

Data Perhitungan Entropy Dan Gain						
Nama Variabel	Indikator	Jumlah	Tidak	Ya	Entropy	Gain
Kepuasan Pelanggan		15	4	11	0.836641	0.378118
Tangibles						
	Facilitas	9	2	7	0.764205	
	Peralatan	3	0	3	0	
	Penampilan Karyawan	1	0	1	0	
	Komunikasi	2	2	0	0	
Reliability						
	Pelayanan Sesuai Dengan Yang Dijanjikan Kepada Pelanggan	7	1	6	0.591673	
	Pelayanan Sesuai Dengan Kebutuhan Pelanggan	8	3	5	0.954434	
Responsiveness						
	Kecapatan Pelayanan	10	3	7	0.881291	
	Cepat Tanggap Dalam Pelayanan	5	1	4	0.721928	
Assurance						
	Pengetahuan Karyawan Tentang Produk	3	1	2	0.918296	
	Kesopanan Karyawan	7	2	5	0.863121	
	Kemampuan Komunikasi Karyawan	5	1	4	0.721928	
Empathy						
	Kepedulian Terhadap Pelanggan	3	3	0	0	
	Perhatian Terhadap Pelanggan	12	2	10	0.650022	

Rule 1	
Indikator	Kepuasan Pelanggan
Tangibles(Facilitas)	PUAS
Tangibles(Peralatan)	PUAS
Tangibles(Penampilan Karyawan)	PUAS
Tangibles(Komunikasi)	TIDAK

Gambar 8. Halaman Data Rule

Pada halaman ini user dapat melihat hasil perhitungan entropy, gain, dan rule yang dihasilkan berdasarkan data training yang sudah ada.

3.8 Pengujian

Pengujian dilakukan dengan membandingkan perhitungan program, excel, dan manual. Berikut ini merupakan hasil perbandingannya.

Tabel 1. Perbandingan Perhitungan

Variabel	Indikator	Kepuasan Pelanggan			Manual		Excel		Program	
		Kasus	Tidak	Ya	Entropy	Gain	Entropy	Gain	Entropy	Gain
Total		15	4	11	0.836641		0.836641		0.836641	
Tangibles (Wujud)	V1	9	2	7	0.764205		0.764205		0.764205	
	V2	3	0	3	0		0		0	
	V3	1	0	1	0		0		0	
	V4	2	2	0	0		0		0	
Reliability (Keandalan)	V5	7	1	6	0.591673		0.591673		0.591673	
	V6	8	3	5	0.954434		0.954434		0.954434	
Responsiveness (Daya Tanggap)	V7	10	3	7	0.881291		0.881291		0.881291	
	V8	5	1	4	0.721928		0.721928		0.721928	
Assurance (Jaminan)	V9	3	1	2	0.918296		0.918296		0.918296	
	V10	7	2	5	0.863121		0.863121		0.863121	
Empathy (Empati)	V11	5	1	4	0.721928		0.721928		0.721928	
	V12	3	3	0	0		0		0	
	V13	12	2	10	0.650022		0.650022		0.650022	

Berdasarkan hasil perhitungan pada tabel 4.5 diketahui bahwa nilai gain terbesar adalah variabel tangibles, sehingga variabel ini menjadi tolak ukur awal terhadap kepuasan pelanggan. Pada variabel tangibles memiliki 4 indikator yang mempengaruhi kepuasan pelanggan. Sehingga dihasilkan rule sebagai berikut.

Tabel 2. *Perbandingan Hasil Perhitungan*

Indikator	Kepuasan Pelanggan		
	Manual	Excel	Program
Tangibles(Fasilitas)	PUAS	PUAS	PUAS
Tangibles(Peralatan)	PUAS	PUAS	PUAS
Tangibles(Penampilan Karyawan)	PUAS	PUAS	PUAS
Tangibles(Komunikasi)	TIDAK	TIDAK	TIDAK

Rule pada tabel 2 menghasilkan ada 3 indikator pada variabel tangibles yang menghasilkan pelanggan puas yaitu fasilitas, peralatan, dan penampilan karyawan, sedangkan indikator komunikasi menghasilkan pelanggan tidak puas. Berdasarkan hasil kuisioner, pelanggan yang memilih indikator fasilitas (7 orang), peralatan (3 orang), dan penampilan karyawan (1 orang) menyatakan puas, sedangkan pelanggan yang berjumlah 2 orang memilih indikator komunikasi dan hasilnya adalah pelanggan tidak puas. Maka indikator yang harus diperbaiki adalah komunikasi.

IV. KESIMPULAN

1. Penggunaan algoritma C4.5 dapat membantu pihak resto dalam meningkatkan pelayanan sesuai dengan hasil kuisioner yang telah diisi oleh para pelanggan. Berdasarkan hasil perhitungan, didapatkan nilai gain tertinggi adalah variabel tangibles dengan nilai 0,378118 dan untuk indikatornya pelanggan menyatakan puas terhadap Fasilitas, Peralatan, dan Penampilan Karyawan. Sedangkan pelanggan tidak puas terhadap satu indikator yaitu Komunikasi.
2. Memudahkan pihak resto dalam mengumpulkan data dari pelanggan untuk meningkatkan pelayanan.
3. Bisa untuk memberikan masukan kepada Sunrise Resto & Cafe dalam meningkatkan pelayanan kepada pelanggan.

V. SARAN

Berikut ini adalah beberapa saran yang perlu dipertimbangkan untuk menjadi landasan penelitian selanjutnya kuisioner yang digunakan sebagai data training sebaiknya lebih banyak lagi, karena semakin banyak data yang digunakan, data yang dihasilkan juga semakin baik..

UCAPAN TERIMA KASIH

Terimakasih kepada Orangtua, keluarga dan para sahabat, kepada dosen pembimbing (Ibu Fitri dan Pak Indra) serta dosen penguji (Pak Istiadi) dan kepada Universitas Widyagama Malang. Terima kasih juga kepada para peneliti terdahulu .Semoga yang ditulis peneliti dapat bermanfaat untuk siapa saja yang ingin belajar. Aamiin...

DAFTAR PUSTAKA

- [1] EKI, Maris Ruziqa. 2015. Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5. Skripsi, Fakultas Ilmu Komputer.
- [2] Tama, Bayu Adhi. 2011. Analisis Kepuasan Konsumen Terhadap Restoran Cepat Saji Melalui Pendekatan Data Mining: Studi Kasus XYZ. Jurnal Generic Vol.6 No.1.
- [3] Agrawel, R., et al., 1993, Mining Association Rules. Between Sets Of Items In Large Databases, Proceeding of the ACM. SIGMOD conference on Management of Data, May 1993, Washington.
- [4] Sunjana, 2010. Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree. Seminar Nasional Aplikasi Teknologi Informasi. Vol 7 pp. 24-29.
- [5] Santoso, T. B. 2014. Analisa Dan Penerapan Metode C4.5 Untuk Prediksi Loyalitas Pelanggan. Jurnal Ilmiah Fakultas Teknik Limit's Vol. 10 No.1.