
DATA MINING UNTUK MENGUKUR TINGKAT KEPUASAN PESERTA BPJS KETENAGAKERJAAN MENGGUNAKAN ALGORITMA C4.5

Afifah Junia¹, Fristi Riandari²

^{1,2}Program Studi Teknik Informatika, STMIK Pelita
Nusantara Jl. St. Iskandar Muda No.1 Medan
Email : afifah.junia1995@gmail.com

***Abstract-** BPJS Employment (Managing Social Security Agency) is a form of public service. Efforts in improving services need to have benchmarks on the level of participant satisfaction both in terms of tangibles, reliability. Responsiveness, assurance and empathy. Based on the research and the problem, a research about participant satisfaction was conducted. Participants are given a questionnaire which will be processed using the C45 algorithm. For this reason, it is necessary to evaluate the participant's side to find out whether the participant is satisfied or not satisfied with the services provided. Seeing the problems above, a research will be conducted with the title "Data mining to measure the level of satisfaction of BPJS Employee participants using the C45 algorithm". The purpose of this study is to find out whether BPJS Employment has provided quality services that are in line with participants' expectations.*

Keyword : data mining, C4.5 algorithm, participant satisfaction

Abstrak- BPJS Ketenagakerjaan (Badan Penyelenggara Jaminan Sosial Ketenagakerjaan) salah satu bentuk pelayanan publik. Upaya dalam meningkatkan pelayanan perlu adanya tolak ukur terhadap tingkat kepuasan peserta baik dalam segi *tangibles, reliability. Responsiveness, assurance* dan *empathy*. Berdasarkan penelitian dan permasalahan tersebut dilakukan penelitian mengenai kepuasan peserta. Peserta diberikan kuesioner yang akan diolah menggunakan algoritma C45. Untuk itu, maka perlu dilakukan evaluasi dari sisi peserta untuk mengetahui apakah peserta puas atau tidak puas dengan pelayanan yang di berikan. Melihat permasalahan diatas maka akan dilakukan penelitian dengan judul "Data mining untuk mengukur tingkat kepuasan peserta BPJS Ketenagakerjaan menggunakan algoritma C45". Tujuan dari penelitian ini untuk dapat mengetahui apakah BPJS Ketenagakerjaan telah memberikan kualitas jasa yang sesuai dengan harapan peserta.

Kata kunci : data mining, algoritma C4.5, kepuasan peserta

1. PENDAHULUAN

1.1 Latar Belakang

BPJS Ketenagakerjaan (Badan Penyelenggara Jaminan Sosial Ketenagakerjaan) salah satu bentuk pelayanan publik. Dalam memberikan pelayanan terhadap peserta BPJS Ketenagakerjaan belum memiliki tolak ukur untuk menentukan apakah peserta sudah puas terhadap pelayanan yang diberikan. Perlunya survei kepuasan pelanggan untuk mengetahui sejauh mana sudah memberikan pelayanan terhadap peserta apakah peserta sudah puas dengan pelayanan yang sudah

diberikan. Sehingga dengan adanya tolak ukur tersebut dapat meningkatkan kepuasan peserta.

Peserta yang datang ke BPJS Ketenagakerjaan akan diberikan kuesioner yang akan diolah menggunakan algoritma C45. Sehingga dapat diketahui apakah BPJS Ketenagakerjaan telah memberikan kualitas jasa yang sesuai dengan harapan peserta. Untuk itu, maka perlu dilakukan evaluasi dari sisi peserta untuk mengetahui apakah peserta puas atau tidak puas dengan pelayanan yang di berikan.

1.2 Rumusan Masalah

Berdasarkan masalah yang telah diuraikan pada latar belakang, maka dapat dirumuskan masalah pada skripsi ini sebagai berikut:

- Bagaimana akurasi algoritma C4.5 untuk memprediksi kepuasan peserta di BPJS Ketenagakerjaan?
- Bagaimana cara menerapkan algoritma C4.5 untuk mengukur tingkat kepuasan pelanggan di BPJS Ketenagakerjaan?
- Bagaimana merancang sebuah aplikasi untuk mengukur tingkat kepuasan peserta dengan algoritma C4.5?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas maka tujuan pada penelitian ini adalah sebagai berikut :

- Mengetahui akurasi algoritma C4.5 untuk memprediksi kepuasan pelanggan di BPJS Ketenagakerjaan.
- Menerapkan algoritma C4.5 pada kepuasan pelanggan di BPJS Ketenagakerjaan. Merancang sebuah aplikasi untuk mengukur tingkat kepuasan dengan algoritma C4.5

2. LANDASAN TEORI

2.1 Data Mining

Ginting, Zarman dan Hamidah (2014) berpendapat, “Data *mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis yang menemukan keteraturan, pola dan hubungan dalam set data berukuran besar. Maksud dari pengertian ini yaitu proses pencarian informasi yang tidak diketahui sebelumnya dari sekumpulan data besar.”

Sulastri dan Gufroni (2017) berpendapat, “*Preprocessing* data *mining* dalam melakukan persiapan data *mining* dapat meningkatkan kualitas data, sehingga data yang diperoleh lebih berkualitas artinya data-data tersebut bersifat objektif, representatif, memiliki *sampling error* yang kecil, terbaharui dan relevan. Persiapan tersebut antara lain:

- Data Cleaning**
Data Cleaning merupakan proses untuk dapat mengatasi nilai yang hilang, *noise* dan data yang tidak konsisten .
- Data Integration**
Data *Integration* merupakan proses menggabungkan data dari banyak database.

- Data Selection**

Data *Selection* merupakan proses meminimalkan jumlah data yang digunakan untuk proses *mining* dengan tetap merepresentasikan data aslinya. Data *selection* dapat berupa *sampling*, *denoising*, dan *feature extraction*.

- Data Transformation**

Data *transformation* dilakukan untuk mengubah bentuk dan format data. Hal ini tentunya sangat membantu memudahkan pengguna dalam proses *mining* ataupun memahami hasil yang didapat. Dalam proses data transformation bisa dilakukan dengan *centering*, *normalization*, dan *scaling*.”

2.2 Pohon Keputusan (Decision Tree)

Vulandari (2018:15) Pohon keputusan adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan aturan keputusan. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk *mem-break down* proses pengambilan keputusan akan lebih menginterpretasikan solusi dari permasalahan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel *input* dengan sebuah variabel *input* dengan sebuah variabel target. Pohon keputusan memadukan antara eksplorasi data dan permodelan, sehingga sangat bagus sebagai langkah awal dalam proses permodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

2.3 Algoritma C4.5

Rismayanti (2016) berpendapat, “Algoritma C4.5 salah satu algoritma yang digunakan untuk membentuk *decision tree* berdasarkan *training* data. Algoritma ini merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. “

Cyhnthia dan Ismanto (2018) berpendapat, “Algoritma C4.5 merupakan kelompok algoritma pohon keputusan (*Decision Tree*). Algoritma ini mempunyai *input* berupa *training samples* dan *samples*. *Training samples* berupa data yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples*

merupakan *field-field* data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data.“

Arifin dan fitriana (2018) berpendapat, “Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai *gain* tertinggi. Berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Pada tahapannya algoritma C4.5 memiliki 2 prinsip kerja, yaitu: Membuat pohon keputusan, dan membuat aturan-aturan (*rule model*). Aturan aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if then*.

Arifin dan fitriana (2018) berpendapat, “*Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impurity* dan *homogeneity* dari kumpulan data. Dari nilai *Entropy* tersebut kemudian dihitung nilai *information gain* masing-masing atribut. Penghitungan nilai *Entropy* digunakan rumus seperti dalam Persamaan (1).

$$Entropy (S) = \sum_{i=0}^n - p_i * \log_2 (p_i) \dots\dots\dots (1)$$

Rumus (1) merupakan rumus yang digunakan dalam perhitungan *entropy* yang digunakan untuk menentukan seberapa informatif atribut tersebut. Berikut keterangannya :

- S : Himpunan kasus
- N : Jumlah partisi
- P_i : Jumlah kasus pada partisi ke- i ”

Arifin dan fitriana (2018) berpendapat, “*Information Gain* adalah informasi yang didapatkan dari perubahan *entropy* pada suatu kumpulan data, baik melalui observasi atau bisa juga disimpulkan dengan cara melakukan partisipasi terhadap suatu set data.

$$GAIN (S, A) = Entropy (S) - \sum_{i=1}^n - [S_i] * Entropy(S_i) \dots\dots\dots(2)$$

| S |

Rumus (2) merupakan rumus yang digunakan dalam perhitungan *information gain* setelah melakukan perhitungan *entropy*. Berikut keterangannya :

- S : Himpunan kasus
- N : Jumlah partisi atribut A
- $|S_i|$: Jumlah kasus pada partisi ke- i
- |S| : Jumlah kasus dalam s

Dengan mengetahui rumus–rumus diatas, data yang telah diperoleh dapat dimasukkan dan diproses dengan algoritma C4.5 untuk proses pembuatan *decision tree*.”

3. PEMBAHASAN

3.1 Analisis Data

BPJS Ketenagakerjaan merupakan badan hukum publik yang memberikan perlindungan terhadap pekerja di Indonesia baik pekerja di bidang formal maupun informal. Dalam melayani pekerja Indonesia, BPJS Ketenagakerjaan memiliki standar pelayanan dalam melayani peserta, atau dikenal dengan pelayanan prima dimana peserta merasa dilayani melebihi ekspektasi. Dalam memberikan pelayanan prima terhadap peserta, BPJS Ketenagakerjaan sendiri belum mengetahui sejauh mana peserta puas terhadap pelayanan yang diberikan. Untuk itu perlu adanya tolak ukur untuk mengukur tingkat kepuasan peserta BPJS Ketenagakerjaan melalui algoritma C4.5.

3.2 Pembahasan

Ada beberapa proses yang akan dilakukan penulis dalam melakukan data mining, yaitu :

a. Data Selection

Memilih himpunan data yang akan digunakan pada penelitian ini yaitu data kuisisioner peserta BPJS Ketenagakerjaan. Untuk data kuisisioner yang berhasil terkumpul selama bulan Mei 2019 sebanyak 102 data.

b. Cleaning

yaitu membersihkan data kuisisioner peserta klaim di BPJS Ketenagakerjaan dengan menghapus data yang tidak digunakan atau yang tidak memiliki nilai *attribute* lengkap (*missing value*). Data kuisisioner yang pada awalnya terdapat 102 data menjadi 78 data yang siap di proses.

c. Data Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

Tabel 1. Range Penilaian Kepuasan Peserta

Keterangan	Range Nilai
Sangat Puas	3
Cukup Puas	2
Tidak Puas	1

Pada tahap ini data kuesioner yang berisi keterangan sangat puas diganti menjadi nilai 3, cukup puas diganti menjadi 2 dan tidak puas diganti menjadi 1. Untuk data kuesioner yang di buat oleh penulis terdiri dari 10 pertanyaan, 2 buah pertanyaan mewakili *Tangibles*, *reliability*, *Responsiveness*, *Assurance* dan *Empathy*. Data tersebut diubah menjadi angka dan dicari rata-rata di setiap pernyataan.

Terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu:

1. Memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada.
2. Membuat cabang untuk masing-masing nilai, artinya membuat cabang sesuai dengan jumlah nilai variabel *gain* tertinggi.
3. Membagi setiap kasus dalam cabang, berdasarkan perhitungan nilai *gain* tertinggi dan perhitungan dilakukan setelah perhitungan nilai *gain* tertinggi awal dan kemudian dilakukan proses perhitungan *gain* tertinggi kembali tanpa meyertakan nilai variabel *gain* awal.
4. Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama, mengulangi semua proses perhitungan *gain* tertinggi untuk masing-masing cabang kasus sampai tidak bisa lagi dilakukan proses perhitungan.

Menghitung nilai *Entropy* dan *GAIN* untuk setiap kriteria node 1

a. *Entropy Tangible*

$$Entropy (S) = \sum_{i=0}^n - p_i * \log_2 (p_i)$$

$$Entropy (S) = (-6/19 * \log_2 (6/19)) + - 13/19 * \log_2 (13/19) = \mathbf{0,899743759}$$

$$Entropy (Tangible Sangat Puas) = (-3/13 * \log_2 (3/13)) + - 10/13 * \log_2 (10/13) = 0,779349837$$

$$Entropy (Tangible Cukup Puas) = (-3/6 * \log_2 (3/6)) + - 3/6 * \log_2 (3/6) = 1$$

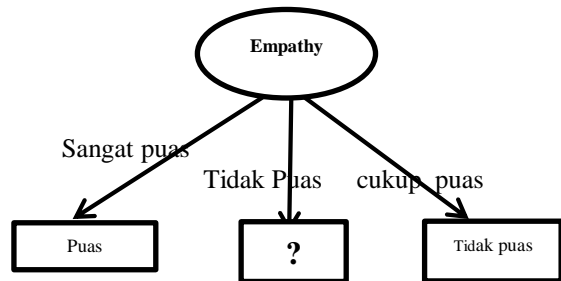
$$Entropy (Tangible Tidak Puas)$$

$$= (-0/0 * \log_2 (0/0)) + - 0/0 * \log_2 (0/0) = 0$$

Gain (Total, *Tangibles* Node 1)

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

$$= \mathbf{0,899743759} * ((13/19 * 0,779349837) + (6/19 * 1) + (0/19 * 0)) = 0,050715$$



Gambar 1. Tampilan pohon keputusan

Seperti yang terlihat pada gambar, diperoleh bahwa atribut dengan *GAIN* tertinggi oleh *Empathy* sebesar 0,430287. Maka *Empathy* menjadi node akar yang memiliki tiga nilai yaitu “ Sangat Puas”, “Cukup Puas” dan “Tidak Puas”. Untuk Sangat puas dan tidak puas hanya memiliki satu nilai yaitu puas dan tidak puas. Sedangkan untuk rule cukup puas masih terdapat dua keputusan “puas” dan “tidak puas” untuk itu dilakukan perhitungan lagi. Melakukan perhitungan lanjutan untuk mencari rule “cukup puas” setelah *empathy*.

4. KESIMPULAN

Adapun kesimpulan yang bisa diambil dalam merancang dan membangun sistem aplikasi kepuasan pelanggan yaitu :

- a. Sistem yang dibangun dapat digunakan oleh pihak BPJS Ketenagakerjaan untuk mengetahui kepuasan peserta. Sistem yang dibangun ini telah melalui proses perhitungan manual dengan menggunakan algoritma C45.
- b. Sistem aplikasi yang dibangun menggunakan Microsoft Visual Studio 2015 yang dapat diakses oleh peserta secara offline
- c. Dengan diketahuinya tingkat kepuasan peserta, diharapkan pihak BPJS Ketenagakerjaan dapat mengevaluasi

serta lebih meningkatkan pelayanan terhadap peserta.

5. DAFTAR PUSTAKA

Azwanti, Nurul. 2018. Algoritma C4.5 Untuk Memprediksi Mahasiswa yang Mengulang Mata Kuliah (Studi Kasus di AMIK Labuhan Batu). Volume 9 No 1. ISSN : 2252-4983

Cynthia Eka Pandu dan, Ismanto Edi. 2018. Metode *Decision Tree* Algoritma C4.5 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji. Volume 3. ISSN : 2527-5771

Khaerunnisa, Handini. 2013. Pengaruh Pelayanan Prima Terhadap Kepuasan Nasabah. Volume 1 No 1. ISSN : 2337-6112

Lubis Alfi Syahri dan Andayani Nur Rahmah. 2017. Pengaruh Kualitas Pelayanan (Service Quality) Terhadap Kepuasan Pelanggan PT. Succofindo Batam. Volume 1 No 2 ISSN : 2548-9909

Rismayanti. 2016. Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa di STT Harapan Medan. Volume 12 No 2. ISSN : 1858-2680

Sijabat, Alimancon. 2015. Penerapan Data Mining untuk Pengolahan Data Siswa dengan Menggunakan Metode Decision Tree. Jurnal Informasi dan Teknologi Ilmiah. Volume 5 No 3. ISSN: 2339-210X.

vrianto, Juang. 2016. Analisi Kepuasan Pelanggan Pada Restoran SOP Tunjang Pertama (M1) di Pekanbaru. Volume 2 No 2. ISSN : 2502-1419