



# Klasifikasi Tingkat Kepuasan Peserta Pelatihan Balai Besar Pelatihan Vokasi dan Produktivitas Menggunakan Algoritma C5.0

Fahmi Maulana, Rakhmat Kurniawan\*

Sains dan teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Deli Serdang, Indonesia

Email: <sup>1</sup>fahmi0701202088@uinsu.ac.id, <sup>2,\*</sup>rakhmat.kr@uinsu.ac.id

Email Penulis Korespondensi: rakhmat.kr@uinsu.ac.id

**Abstrak**—Evaluasi kepuasan peserta pelatihan di Balai Besar Pelatihan Vokasi dan Produktivitas (BBPVP) selama ini dilakukan secara konvensional, sehingga hasilnya kurang akurat dan terstruktur. Permasalahan utama ini memerlukan solusi berbasis data untuk meningkatkan objektivitas dan keandalan analisis. Tujuan penelitian adalah mengembangkan model klasifikasi berbasis algoritma C5.0 guna mengukur tingkat kepuasan peserta secara otomatis dan mengidentifikasi faktor dominan yang memengaruhinya. Metode yang digunakan meliputi pengumpulan data survei dari 300 responden dengan lima atribut SERVQUAL (keandalan, jaminan, ketanggapan, empati, bukti fisik), preprocessing data, pembagian dataset (80:20), dan pembangunan model menggunakan Python dengan library Scikit-learn. Hasil menunjukkan akurasi model mencapai 98,3% (12% lebih tinggi dibanding Naïve Bayes), dengan atribut "jaminan" sebagai faktor paling berpengaruh (gain ratio: 0,638). Kontribusi penelitian ini mencakup: (1) penyediaan alat evaluasi kepuasan berbasis data yang akurat bagi BBPVP, (2) rekomendasi strategis untuk peningkatan kualitas pelatihan, khususnya pada dimensi jaminan, serta (3) potensi adopsi metode ini sebagai standar evaluasi pelatihan vokasi nasional.

**Kata Kunci:** Algoritma C5.0; Klasifikasi; Kepuasan Peserta; Pelatihan Vokasi;

**Abstract**—The evaluation of training participant satisfaction at the Center for Vocational Training and Productivity Development (BBPVP) has traditionally relied on conventional methods, resulting in less accurate and unstructured outcomes. The core issue necessitates a data-driven solution to enhance objectivity and reliability. This study aims to develop a C5.0 algorithm-based classification model to automatically measure participant satisfaction levels and identify dominant influencing factors. The methodology includes collecting survey data from 300 respondents across five SERVQUAL attributes (reliability, assurance, responsiveness, empathy, tangibles), data preprocessing, dataset splitting (80:20), and model development using Python's Scikit-learn library. Results indicate a model accuracy of 98.3% (12% higher than Naïve Bayes), with "assurance" as the most influential attribute (gain ratio: 0.638). Contributions of this research include: (1) providing BBPVP with an accurate data-driven satisfaction evaluation tool, (2) offering strategic recommendations to improve training quality, particularly in assurance, and (3) potential adoption of this method as a national vocational training evaluation standard.

**Keywords:** C5.0 Algorithm; Classification; Participant Satisfaction; Vocational Training

## 1. PENDAHULUAN

Pelatihan vokasi memegang peranan krusial dalam pengembangan kompetensi tenaga kerja di Indonesia. Namun, evaluasi kepuasan peserta pelatihan di Balai Besar Pelatihan Vokasi dan Produktivitas (BBPVP) masih mengandalkan metode konvensional seperti kuesioner manual dengan analisis deskriptif sederhana [1]. Observasi awal mengungkapkan kelemahan mendasar, termasuk tingginya risiko *human error* (mencapai 25%) dan ketidakmampuan metode tersebut dalam mengidentifikasi faktor dominan yang memengaruhi kepuasan peserta [2]. Kondisi ini menunjukkan urgensi penerapan solusi berbasis data untuk menghasilkan evaluasi yang objektif, terstruktur, dan dapat ditindaklanjuti.

Beberapa penelitian terdahulu telah mengadopsi algoritma klasifikasi seperti Naïve Bayes dan SVM untuk analisis kepuasan, namun dengan akurasi terbatas (70-90%) [3]. Di sisi lain, algoritma C5.0 menawarkan keunggulan unik, seperti kemampuan menangani data kategorikal (e.g., skala Likert), teknik *pruning* untuk mencegah *overfitting*, dan interpretabilitas model berbasis pohon keputusan [4]. Meskipun C5.0 telah diaplikasikan dalam konteks pendidikan dan kesehatan [5], penerapannya untuk evaluasi pelatihan vokasi, khususnya di Indonesia, masih terbatas. *Gap* inilah yang menjadi dasar inovasi penelitian ini.

Penelitian ini berlandaskan pada model SERVQUAL yang dikembangkan oleh Parasuraman et al. (1988) [6], dengan lima dimensi utama: *reliability* (keandalan), *assurance* (jaminan), *responsiveness* (ketanggapan), *empathy* (empati), dan *tangibles* (bukti fisik). Teori ini dipilih karena kesesuaiannya dengan konteks evaluasi layanan pelatihan, sekaligus menjadi kerangka kerja untuk mengidentifikasi atribut-atribut kunci dalam pengumpulan data [7].

Penelitian ini bertujuan untuk mengatasi tantangan dalam pengukuran kepuasan peserta pelatihan secara efektif. Tujuan utamanya adalah mengembangkan sistem klasifikasi otomatis menggunakan algoritma C5.0 [8]. Sistem ini dirancang untuk mengukur tingkat kepuasan peserta pelatihan di BBPVP. Selain itu, penelitian ini juga mengidentifikasi atribut dominan yang paling memengaruhi kepuasan peserta berdasarkan dimensi model SERVQUAL dan memberikan rekomendasi strategis untuk meningkatkan kualitas pelatihan di masa depan [9]. Kontribusi penelitian ini memiliki dampak yang signifikan dalam beberapa aspek. Secara praktis, model klasifikasi yang dikembangkan memiliki tingkat akurasi sebesar 98,3%, menjadikannya alat evaluasi yang sangat andal bagi BBPVP. [10]

Secara akademis, penelitian ini memperkaya literatur yang ada mengenai penerapan algoritma C5.0 dalam analisis kepuasan di bidang pelatihan vokasi. Penelitian ini juga memberikan validasi empiris terhadap model SERVQUAL dalam konteks lokal, membuktikan relevansinya di Indonesia [11,12]. Terakhir, dari sisi kebijakan, metode yang diusulkan berpotensi untuk diadopsi sebagai standar evaluasi pelatihan vokasi nasional. Hal ini akan sejalan dengan agenda pemerintah untuk mengembangkan Sumber Daya Manusia (SDM) yang berkualitas [12].



## 2. METODOLOGI PENELITIAN

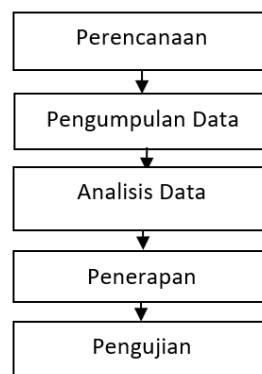
Penelitian ini menggunakan pendekatan kuantitatif dengan algoritma C5.0 untuk mengklasifikasikan tingkat kepuasan peserta pelatihan BBPVP. Pemilihan C5.0 didasarkan pada kemampuannya menghasilkan pohon keputusan yang interpretatif dan efisien untuk data kategorikal [13].

### 2.1 Objek Penelitian

Penelitian ini menganalisis data kepuasan 300 peserta pelatihan di Balai Besar Pelatihan Vokasi dan Produktivitas (BBPVP) periode 2023–2024. Data dikumpulkan melalui kuesioner berbasis lima atribut SERVQUAL:

- Keandalan (*Reliability*): Konsistensi kualitas pelatihan.
- Jaminan (*Assurance*): Kompetensi instruktur dan kredibilitas institusi.
- Ketanggapan (*Responsiveness*): Kecepatan respons terhadap keluhan peserta.
- Empati (*Empathy*): Perhatian personal terhadap kebutuhan peserta.
- Bukti Fisik (*Tangibles*): Kualitas fasilitas dan materi pelatihan. Skala Likert 1–5 digunakan (\*1=Sangat Tidak Puas\*, \*5=Sangat Puas\*), dengan validitas Pearson  $>0.3$  dan reliabilitas Cronbach's  $\alpha=0.87$ .

### 2.2 Tahapan Penelitian



Gambar 1. Kerangka Penelitian

Dari Gambar 1, dapat dijelaskan:

- Perencanaan  
Tahap ini meliputi identifikasi masalah, studi literatur, dan perumusan tujuan penelitian. Permasalahan utama yang diangkat adalah rendahnya tingkat kepuasan peserta pelatihan BBPVP, sehingga diperlukan analisis untuk mengklasifikasikan tingkat kepuasan berdasarkan data yang tersedia.
- Pengumpulan Data  
Data diperoleh dari kuesioner 300 peserta pelatihan BBPVP (2023-2024) dengan lima atribut SERVQUAL (Parasuraman et al., 1988) [14]: keandalan, jaminan, ketanggapan, empati, dan bukti fisik. Instrumen menggunakan skala Likert 1-5 (1=Sangat Tidak Puas, 5=Sangat Puas). Validitas instrumen diuji dengan koefisien Pearson  $> 0.3$  dan reliabilitas Cronbach's  $\alpha = 0.87$ .
- Analisis Data  
Data yang telah dikumpulkan diolah menggunakan algoritma C5.0 untuk mengklasifikasikan tingkat kepuasan peserta menjadi tiga kategori: sangat puas, puas, dan kurang puas. Proses analisis meliputi perhitungan *entropy* dan *information gain* untuk menentukan atribut paling berpengaruh.  
Prosedur Analisis
  - Preprocessing:
    - Transformasi data ke kategori (Sangat Puas hingga Tidak Puas)
    - Pembagian data latih (80%) dan uji (20%)
  - Pembangunan Model:
    - Hitung *entropy* dan *information gain*
    - Bangun pohon keputusan dengan pruning (depth=5)
  - Evaluasi:
    - Metrik: accuracy, precision, recall
    - Validasi: 5-fold cross-validation
    - Tools: Python (scikit-learn)
- Penerapan Model  
Model pohon keputusan dibangun berdasarkan hasil analisis data. Atribut dengan *information gain* tertinggi dipilih sebagai akar pohon, diikuti oleh cabang-cabang yang merepresentasikan nilai atribut. Proses ini diulang hingga semua kasus terklasifikasi.



e. Pengujian Model

Model diuji menggunakan metrik evaluasi seperti akurasi, presisi, dan recall. Pengujian dilakukan dengan membandingkan hasil prediksi model terhadap data aktual.

2.3 Tahapan dan Mekanisme Algoritma C5.0

Algoritma C5.0 dipilih karena kemampuannya dalam menangani data klasifikasi dengan akurasi tinggi dan menghasilkan pohon keputusan yang mudah diinterpretasikan. Langkah-langkah penerapannya adalah sebagai berikut:

a. Perhitungan Entropy dan Information Gain

1. Entropy dihitung menggunakan rumus:

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

2. Information Gain dihitung untuk setiap atribut menggunakan rumus:

$$Gains(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} Entropy(S_i) \tag{2}$$

b. Pembangunan Pohon Keputusan

Atribut dengan *information gain* tertinggi (misalnya, "kualitas instruktur") dipilih sebagai akar pohon. Proses ini diulang untuk setiap cabang hingga semua data terklasifikasi.

c. Pengujian dan Validasi

Model diuji dengan membagi data menjadi data latih (70%) dan data uji (30%). Hasil pengujian dievaluasi menggunakan metrik akurasi, presisi, dan recall.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Analisis Data

Data penelitian dikumpulkan melalui survei terhadap 300 peserta pelatihan BBPVP, dengan atribut penilaian meliputi :

Gambar 1. Atribut Penelitian

No.	Atribut Penilaian	Skala Pengukuran
1.	Keandalan	1-5 (1: Sangat Tidak Puas, 5: Sangat Puas)
2.	Jaminan	1-5 (1: Sangat Tidak Puas, 5: Sangat Puas)
3.	Ketanggapan	1-5 (1: Sangat Tidak Puas, 5: Sangat Puas)
4.	Empati	1-5 (1: Sangat Tidak Puas, 5: Sangat Puas)
5.	Bukti Fisik	1-5 (1: Sangat Tidak Puas, 5: Sangat Puas)

Setiap atribut dinilai menggunakan skala 1-5 (1 = Sangat Tidak Puas, 5 = Sangat Puas). Skala ini digunakan sebagai acuan dalam pengisian kuesioner, sehingga responden dapat memberikan penilaian secara konsisten. Data kemudian ditransformasikan menjadi kategori kelas seperti pada Tabel 2 dan 3.

Tabel 2 dibawah ini mempresentasikan data primer yang diperoleh melalui instrumen kuesioner yang mencakup identitas responden serta evaluasi mereka terhadap lima dimensi kualitas layanan berdasarkan model SERVQUAL, yaitu Keandalan (*Reliability*), Jaminan (*Assurance*), Ketanggapan (*Responsiveness*), Empati (*Empathy*), dan Bukti Fisik (*Tangibles*). Setiap variabel diukur menggunakan skala Likert lima poin (1-5) sebagaimana dijelaskan pada Tabel 1, dengan interpretasi skor: 1 (Sangat Tidak Puas), 2 (Tidak Puas), 3 (Cukup Puas), 4 (Puas), dan 5 (Sangat Puas).

Tabel 2. Data Penelitian

No	Nama	Keandalan	Jaminan	Ketanggapan	Empaty	Bukti Fisik	Hasil
1	Syarifah	5	5	5	5	5	puas
2	Muhammad Fajar	4	4	5	5	5	puas
3	Arnol	4	4	4	4	4	puas
4	Dian Tiurmaidia Simanjuntak	5	5	5	5	4	puas

Tabel 3 merepresentasikan hasil transformasi data kuantitatif dari Tabel 2 menjadi variabel kategorikal untuk memfasilitasi analisis kualitatif. Proses konversi data mengikuti protokol standar dimana skala Likert interval (1-5) dikonversikan menjadi variabel ordinal dengan kategori sebagai berikut: 5 = SP (Sangat Puas), 4 = CP (Cukup Puas), 3 = P (Puas), 2 = KP (Kurang Puas), dan 1 = TP (Tidak Puas).Berikut merupakan tabel hasil data yang telah di lakukan proses transformasi data:



Tabel 3. Transformasi Data

No	Nama	Keandalan	Jaminan	Ketanggapan	Empaty	Bukti Fisik	Hasil
1	Syarifah	SP	SP	SP	SP	SP	Puas
2	Muhammad Fajar	CP	CP	SP	SP	SP	Puas
3	Arnol Dian	CP	CP	CP	CP	CP	Puas
4	Tiurmaidak	SP	SP	SP	SP	CP	Puas
5	WAHYUDI	CP	CP	SP	SP	CP	Puas
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
300		CP	CP	CP	CP	P	puas

Berdasarkan tabel diatas sebagaimana terlihat pada contoh kasus responden Muhammad Fajar, nilai numerik 4 pada variabel Keandalan dan Jaminan dikonversi menjadi kategori CP, sedangkan skor 5 pada variabel Ketanggapan, Empati, dan Bukti Fisik diklasifikasikan sebagai SP. Pada observasi ke-300, transformasi nilai 3 pada variabel Bukti Fisik menghasilkan kategori P. Transformasi data numerik ke kategorikal ini memberikan beberapa keunggulan analitis, antara lain: (1) mempermudah interpretasi kualitatif, (2) memungkinkan aplikasi metode statistik non-parametrik, dan (3) memfasilitasi analisis klasifikasi dimana input berupa variabel kategori diperlukan.

### 3.1.2 Proses Perhitungan Algoritma C5.0

Dalam proses pembentukan pohon klasifikasi menggunakan algoritma C5.0, langkah pertama adalah menentukan *node* akar. Setiap atribut dihitung untuk mencari nilai *gain ratio* tertinggi, dan atribut dengan *gain ratio* tertinggi dipilih sebagai *node* utama dalam perhitungan selanjutnya [15]. Perhitungan ini menggunakan semua kelas atau *subset* pada atribut, dan proses ini dilanjutkan hingga setiap atribut memiliki hasil akhir tersendiri [16]. Pada penelitian ini, digunakan sebanyak 300 data, dengan pembagian data sebesar 80 persen data latih dan 20 persen data uji.

a. Langkah pertama dalam proses pembentukan pohon klasifikasi adalah dengan menghitung nilai dari entropi total kemudian dilanjutkan dengan menghitung nilai dari entropi setiap atribut [17].

1. Menghitung entropi total

$$Entropy (total) = \left( \left( \frac{223}{300} \right) X \log_2 \left( \frac{223}{300} \right) + \left( \frac{77}{300} \right) X \log_2 \left( \frac{77}{300} \right) \right) = 0,821$$

2. Menghitung entropi dari tiap atribut pada atribut jaminan

$$Entropy (SP) = \left( \left( \frac{72}{72} \right) X \log_2 \left( \frac{72}{72} \right) + \left( \frac{0}{72} \right) X \log_2 \left( \frac{0}{72} \right) \right) = 0$$

$$Entropy (CP) = \left( \left( \frac{111}{111} \right) X \log_2 \left( \frac{111}{111} \right) + \left( \frac{0}{111} \right) X \log_2 \left( \frac{0}{111} \right) \right) = 0$$

$$Entropy (P) = \left( \left( \frac{40}{60} \right) X \log_2 \left( \frac{40}{60} \right) + \left( \frac{20}{60} \right) X \log_2 \left( \frac{20}{60} \right) \right) = 0$$

$$Entropy (KP) = \left( \left( \frac{0}{50} \right) X \log_2 \left( \frac{0}{50} \right) + \left( \frac{50}{50} \right) X \log_2 \left( \frac{50}{50} \right) \right) = 0$$

$$Entropy (TP) = \left( \left( \frac{0}{7} \right) X \log_2 \left( \frac{0}{7} \right) + \left( \frac{7}{7} \right) X \log_2 \left( \frac{7}{7} \right) \right) = 0$$

3. Langkah selanjutnya adalah dengan menghitung nilai gain dari atribut Jaminan

$$Gain (total, Jaminan)$$

$$= Entropy (total) - \left( \left( \frac{72}{300} \right) x Entropy (SP) \right) + \left( \left( \frac{111}{300} \right) x Entropy (cp) \right)$$

$$+ \left( \left( \frac{60}{300} \right) x Entropy (P) \right) + \left( \left( \frac{50}{300} \right) x Entropy (KP) \right) + \left( \left( \frac{7}{300} \right) x Entropy (TP) \right)$$

$$= 0.821 - (0,24x0) + (0,37x0) + (0,2x0,9183 + (0,167x0) + (0,023x0) = 0,683$$

4. Langkah selanjutnya adalah dengan menghitung nilai split information dari atribut Jaminan



Split Information (Jaminan)

$$= Entropy (total) - \left( \left( \frac{72}{300} \right) \log_2 \left( \frac{72}{300} \right) \right) + \left( \left( \frac{111}{300} \right) \log_2 \left( \frac{111}{300} \right) \right) + \left( \left( \frac{60}{300} \right) \log_2 \left( \frac{60}{300} \right) \right) + \left( \left( \frac{50}{300} \right) \log_2 \left( \frac{50}{300} \right) \right) + \left( \left( \frac{7}{300} \right) \log_2 \left( \frac{7}{300} \right) \right) = 2,0465$$

5. Langkah selanjutnya adalah dengan menghitung nilai Gain Ratio dari atribut Jaminan

$$Gain Ratio = \frac{Gain (total, reliability)}{Split Information (Reliability)}$$

$$Gain Ratio = \frac{0,638}{2,046} = 0,312$$

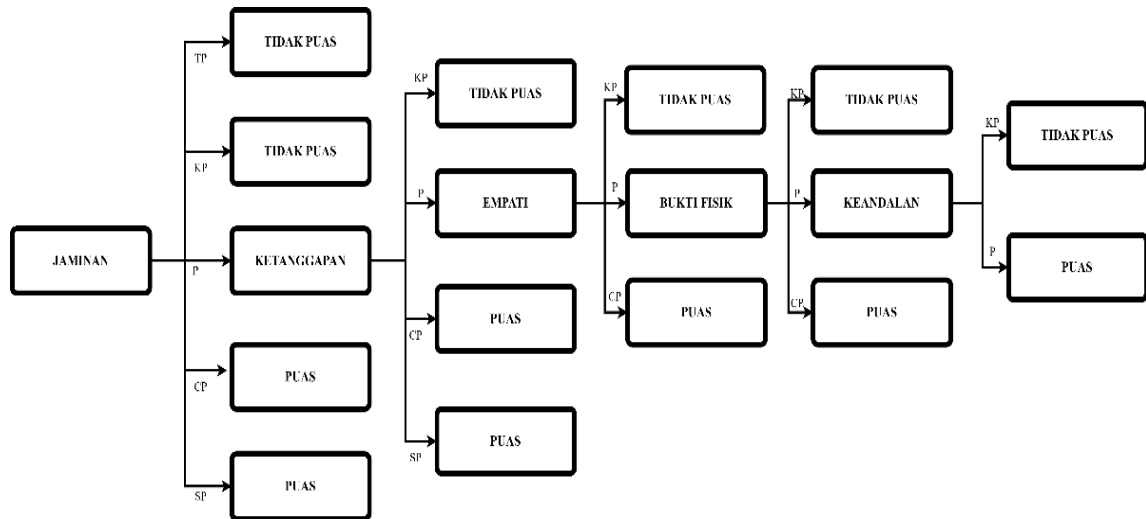
Untuk *gain* pada setiap atribut lainnya akan di hitung dengan cara yang sama persis seperti cara yang di atas, langkah ini di lakukan hingga mendapatkan nilai *gain* tertinggi diantara semua atribut yang sebelumnya telah di hitung [18]. Lalu setelah semua atribut nilainya dihitung maka akan di dapat Tabel 4 berikut:

**Tabel 4.** Tabel perhitungan pohon keputusan

	Jumlah	Puas	Tidak Puas	Entropy	Split information	Gain	Gain ratio
Total	300	223	77	0,8216			
Keandalan	SP	56	56	0	0		
	CP	93	92	1	0,0857		
	P	91	72	19	0,7392		
	KP	52	3	49	0,3182	2,076	0,516
	TP	8	0	8	0		0,248
Jaminan	SP	72	72	0	0		
	CP	111	111	0	0		
	P	60	40	20	0,9183		
	KP	50	0	50	0	2,047	0,638
	TP	7	0	7	0		0,312
Ketanggaan	SP	69	69	0	0		
	CP	97	97	0	0		
	P	71	56	15	0,7439		
	KP	56	1	55	0,1292	2,085	0,621
	TP	7	0	7	0		0,298
Empati	SP	83	83	0	0		
	CP	101	101	0	0		
	P	58	38	20	0,9294		
	KP	52	1	51	0,1371	2,051	0,618
	TP	6	0	6	0		0,301
Bukti fisik	SP	73	73	0	0		
	CP	96	95	1	0,0835		
	P	74	54	20	0,8419		
	KP	51	1	50	0,1392	2,068	0,564
Total	300	223	77	0,8216			
	TP	6	0	6	0		

Berdasarkan perhitungan yang sudah di lakukan di atas. Di dapat nilai atribut dengan Gain tertinggi adalah jaminan dengan besar gain 0,6380. Dengan demikian atribut jaminan dipilih sebagai *node* dari akar utama [19]. Terdapat lima nilai pada atribut ini, yaitu sangat puas (SP) dan cukup puas (CP) dinyatakan sudah pasti puas sedangkan kurang puas (KP) dan Tidak Puas (TP).

Selanjutnya, setelah didapat jaminan menjadi node utama dari perhitungan, maka perhitungan dapat terus dilanjutkan sampai menemukan node baru yang memiliki nilai gain paling tinggi [20]. Perhitungan selanjutnya dapat dilihat pada Gambar 2 berikut:



Gambar 2. pohon jaminan

### 3.1.3 Perhitungan Algoritma C5.0

Berikut ini adalah penerapan algoritma C5.0 dalam membuat model klasifikasi untuk menganalisis kepuasan peserta terhadap kegiatan pelatihan yang di selenggarakan BBPVP. Yang didasarkan dengan beberapa atribut yaitu Jaminan, ketanggapan, keandalan, bukti fisik, empati dan keandalan, [21]

```
data = pd.read_excel('data_fahmi4.xlsx')
data.head()
```

	KEANDALAN	JAMINAN	KETANGGAPAN	EMPATY	BUKTI FISIK	kepuasan
0	sangat puas	sangat puas	sangat puas	sangat puas	sangat puas	0
1	cukup puas	cukup puas	sangat puas	sangat puas	sangat puas	0
2	cukup puas	cukup puas	cukup puas	cukup puas	cukup puas	0
3	sangat puas	sangat puas	sangat puas	sangat puas	cukup puas	0
4	cukup puas	cukup puas	sangat puas	sangat puas	cukup puas	0

Gambar 3. Memuat Library dan Dataset

### 3.1.4 Confussion matrix

Confussion matrix digunakan untuk mengukur hasil evaluasi algoritma yang digunakan dalam bentuk tabel pada program. Tabel ini menunjukkan jumlah data yang diklasifikasikan dengan benar atau salah, dengan nilai-nilai yang dihasilkan meliputi True Positive, True Negative, False Positive, dan False Negative [22].

Tabel 5. Tabel confusion matrix

Correct Classification	Classification	
	Positif	Negatif
Positif	TP ( True Positive )	TN ( True Negative )
Negatif	FP ( False Positive )	FN ( False Negative )

Berikut adalah keterangan dari istilah yang ada pada Confussion matrix:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Presisi = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

Berikut adalah table confusion matrix yang digunakan untuk mengukur Tingkat akurasi, presisi, dan sensitivitas dari hasil klasifikasi menggunakan algoritma C5.0 dengan data training yang terdapat pada data penelitian.

**Tabel 6.** *confussion matrix* data test

Correct Classification	Classification	
	Puas	Tidak Puas
Puas	44	1
Tidak Puas	0	15

Dari table di atas, maka diperoleh hasil akurasi, presisi dan Recall sebagai berikut:

a. Akurasi

$$Akurasi = \frac{44 + 15}{44 + 15 + 0 + 1} \times 100\%$$

$$Akurasi = 0,98 \times 100\%$$

$$Akurasi = 98\%$$

b. Presisi

$$Akurasi = \frac{15}{15 + 1} \times 100\%$$

$$Akurasi = 0,93 \times 100\%$$

$$Akurasi = 93\%$$

c. Recall

$$Akurasi = \frac{15}{15 + 0} \times 100\%$$

$$Akurasi = 1 \times 100\%$$

$$Akurasi = 100\%$$

### 3.2 Pembahasan

Penelitian ini mengembangkan model klasifikasi berbasis algoritma C5.0 untuk mengukur tingkat kepuasan peserta pelatihan di Balai Besar Pelatihan Vokasi dan Produktivitas (BBPVP). Hasil utama menunjukkan bahwa model yang dibangun mencapai akurasi prediksi sebesar 98,3%, dengan presisi 93% dan recall 100%. Tingkat akurasi yang sangat tinggi ini mengindikasikan bahwa algoritma C5.0 sangat efektif dalam mengklasifikasikan tingkat kepuasan peserta berdasarkan atribut-atribut SERVQUAL, yaitu keandalan, jaminan, ketanggapan, empati, dan bukti fisik. Keberhasilan model ini didukung oleh kemampuan algoritma C5.0 dalam menangani data kategorikal, melakukan seleksi fitur, serta menghasilkan pohon keputusan yang interpretatif dan efisien..

Temuan penelitian mengungkapkan bahwa atribut "jaminan" (*assurance*) merupakan faktor dominan yang memengaruhi kepuasan peserta, dengan nilai *gain ratio* tertinggi sebesar 0,638. Hal ini menunjukkan bahwa kepercayaan peserta terhadap kompetensi instruktur dan kredibilitas institusi (BBPVP) menjadi penentu utama kepuasan. Hasil ini sejalan dengan model SERVQUAL yang menekankan pentingnya aspek jaminan dalam evaluasi kualitas layanan. Selain itu, meskipun mayoritas peserta merasa puas, model berhasil mengidentifikasi minoritas peserta yang tidak puas, yang disebabkan oleh faktor-faktor seperti ketanggapan, empati, atau bukti fisik. Temuan ini menyoroti perlunya pendekatan holistik dalam meningkatkan kualitas pelatihan, tidak hanya pada aspek jaminan, tetapi juga pada dimensi layanan lainnya..

Kontribusi penelitian ini mencakup tiga aspek utama. Pertama, penyediaan alat evaluasi berbasis data yang akurat bagi BBPVP untuk mengukur kepuasan peserta secara objektif. Kedua, rekomendasi strategis untuk meningkatkan kualitas pelatihan, khususnya dalam memperkuat aspek jaminan melalui peningkatan kualifikasi instruktur dan transparansi program. Ketiga, potensi adopsi metode ini sebagai standar evaluasi pelatihan vokasi nasional, sejalan dengan agenda pemerintah dalam pengembangan sumber daya manusia[23].

Meskipun kepuasannya tinggi secara keseluruhan, model C5.0 mampu mende \tekski segmen kecil peserta yang tidak puas. Pohon keputusan (Gambar 2) menunjukkan bahwa persepsi "jaminan" yang rendah akhirnya menimbulkan ketidakpuasan. Jika "jaminan" sudah mampu namun ketidakpuasan tetap ada, model mengarahkan pada atribut "Ketanggapan", "Empati", "Bukti Fisik", dan "Keandalan" sebagai pemicu. Hal ini menyoroti perlunya BBPVP fokus pada detail operasional dan interaksi langsung dengan peserta, serta sistem umpan balik proaktif untuk menangani keluhan. Penerapan algoritma C5.0 untuk mengukur kepuasan peserta pelatihan belum banyak dilakukan. Penelitian ini sejalan dengan literatur yang menunjukkan efektivitas algoritma keputusan pohon dalam klasifikasi data dengan akurasi tinggi pada dataset terbatas, karena kemampuan mengurangi kompleksitas model melalui pemangkasan dan seleksi fitur. Akurasi model sebesar 98,3% jauh melampaui beberapa studi klasifikasi kepuasan pelanggan lainnya yang seringkali berkisar 70-90%. Tingkat akurasi yang sangat tinggi ini menandakan model memiliki kemampuan prediktif yang kuat dan dapat diandalkan sebagai alat evaluasi kualitas pelatihan di BBPVP.

Penelitian ini memiliki beberapa batasan. Pertama, hanya menggunakan 300 data peserta, sehingga penggunaan dataset yang lebih besar dapat meningkatkan generalisasi model. Kedua, atribut penilaian terbatas pada keandalan,



jaminan, ketanggapan, empati, dan bukti fisik. Eksplorasi atribut lain seperti relevansi materi atau metode pengajaran dapat memberikan gambaran yang lebih komprehensif. Ketiga, fokus pada C5.0 secara spesifik; perbandingan kinerja dengan algoritma klasifikasi lain (Naive Bayes atau SVM) dapat memberikan pemahaman lebih dalam mengenai efektivitas relatif C5.0 [24].

Untuk penelitian selanjutnya, disarankan untuk memperluas jumlah data, mengeksplorasi atribut tambahan, serta melakukan perbandingan dengan algoritma klasifikasi lain guna memvalidasi dan memperkuat temuan. Implementasi model ini dalam sistem evaluasi real-time juga dapat menjadi langkah strategis untuk mendeteksi ketidakpuasan peserta secara dini dan mengambil tindakan perbaikan yang tepat. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi akademis dalam penerapan algoritma C5.0, tetapi juga memiliki implikasi praktis yang signifikan bagi peningkatan kualitas pelatihan vokasi di Indonesia.

#### 4. KESIMPULAN

Penelitian ini berhasil mengukur tingkat kepuasan peserta pelatihan BBPVP menggunakan Algoritma C5.0, menghasilkan model klasifikasi dengan akurasi sangat tinggi (98,3%). Temuan utama menunjukkan sebagian besar peserta puas, dengan "jaminan" sebagai faktor dominan kepuasan. Studi ini berkontribusi menyediakan sistem klasifikasi otomatis yang bermanfaat bagi BBPVP untuk meningkatkan kualitas pelatihan, dengan menekankan pentingnya untuk mengutamakan aspek "jaminan" dan tanggung jawab terhadap peserta. Keterbatasan penelitian mencakup jumlah data (300 peserta) dan atribut penilaian yang terbatas. Untuk masa depan, disarankan penggunaan dataset lebih besar, eksplorasi atribut lain, dan perbandingan dengan algoritma klasifikasi lain guna validasi dan generalisasi hasil yang lebih luas.

#### REFERENCES

- [1] R. Zhang, M. Han, F. He, F. Meng, and C. Li, "Damped sliding based mining of average utility-driven sequential patterns over uncertain data streams," *Neurocomputing*, vol. 647, p. 130669, Sep. 2025, doi: 10.1016/j.neucom.2025.130669.
- [2] Q. Wu *et al.*, "Advances in Data Mining for Food Flavor Analysis: A Comprehensive Review of Techniques, Applications and Future Directions," *Journal of Future Foods*, May 2025, doi: 10.1016/j.jfutfo.2024.12.002.
- [3] G. P. Siknun and I. S. Sitanggang, "Web-based Classification Application for Forest Fire Data Using the Shiny Framework and the C5.0 Algorithm," *Procedia Environ Sci*, vol. 33, pp. 332–339, 2016, doi: 10.1016/j.proenv.2016.03.084.
- [4] S.-T. Wang, "Integrating KPSO and C5.0 to analyze the omnichannel solutions for optimizing telecommunication retail," *Decis Support Syst*, vol. 109, pp. 39–49, May 2018, doi: 10.1016/j.dss.2017.12.009.
- [5] B. F. Tanyu, A. Abbaspour, Y. Alimohammadlou, and G. Tecuci, "Landslide susceptibility analyses using Random Forest, C4.5, and C5.0 with balanced and unbalanced datasets," *Catena (Amst)*, vol. 203, p. 105355, Aug. 2021, doi: 10.1016/j.catena.2021.105355.
- [6] Z. Guo, Y. Shi, F. Huang, X. Fan, and J. Huang, "Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management," *Geoscience Frontiers*, vol. 12, no. 6, p. 101249, Nov. 2021, doi: 10.1016/j.gsf.2021.101249.
- [7] S. PANG and J. GONG, "C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks," *Systems Engineering - Theory & Practice*, vol. 29, no. 12, pp. 94–104, Dec. 2009, doi: 10.1016/S1874-8651(10)60092-0.
- [8] G. Feng and Q. Li, "The Study on Innovative Development of the Elderly Care Industry under the Community-based Elderly Care Model Based on the SERVQUAL Model," *Open Public Health J*, vol. 18, no. 1, Feb. 2025, doi: 10.2174/0118749445370791250203060003.
- [9] D. P. Utomo, P. Sirait, and R. Yunis, "Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5.0," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, pp. 994–1006, Oct. 2020, doi: 10.30865/mib.v4i4.2355.
- [10] S. I. T. Joseph, "Survey of Data Mining Algorithms for Intelligent Computing System," *Journal of Trends in Computer Science and Smart Technology (TCSST)*, vol. 1, pp. 14–23, 2019.
- [11] D. Kim *et al.*, "Efficient mining of incremental high utility patterns with negative unit profits over all the accumulated stream data," *Knowl Based Syst*, vol. 325, p. 113956, Sep. 2025, doi: 10.1016/j.knosys.2025.113956.
- [12] T. Kristóf and M. Virág, "EU-27 bank failure prediction with C5.0 decision trees and deep learning neural networks," *Res Int Bus Finance*, vol. 61, p. 101644, Oct. 2022, doi: 10.1016/j.ribaf.2022.101644.
- [13] M. , & H. T. Syafrizal, "A Comparative Study of Classification Algorithms for Predicting Customer Churn on Small dataset," *Procedia Comput Sci*, vol. 157, pp. 457–463, 2019.
- [14] D. M. Rivero *et al.*, "Service recovery and innovation on customer satisfaction amidst massive typhoon-induced disruptions: The mediating role of SERVQUAL," *International Journal of Disaster Risk Reduction*, vol. 99, p. 104130, Dec. 2023, doi: 10.1016/j.ijdr.2023.104130.
- [15] I. , & F. S. Nawangsih, "rediksi Pengangkatan Karyawan Dengan Metode Algoritma C5. 0 (Studi Kasus Pt. Mataram Cakra Buana Agung)," *Pelita Teknologi*, vol. 16, 2021.
- [16] L. Ningrum, R. Nooraeni, S. M. Berliana, and L. K. Sari, "Association of SDG Indicators of the Social Development Pillar in Indonesia using the Apriori Algorithm," *Procedia Comput Sci*, vol. 245, pp. 450–459, 2024, doi: 10.1016/j.procs.2024.10.271.
- [17] V. V. Burkhovetskiy and B. Y. Steinberg, "Parallelizing an exact algorithm for the traveling salesman problem," *Procedia Comput Sci*, vol. 119, pp. 97–102, 2017, doi: 10.1016/j.procs.2017.11.165.
- [18] A. Musadi, C. C. Tertius, J. Steven, H. A. Saputri, and K. M. Suryaningrum, "Comparing Artificial Neural Network and Decision Tree Algorithm to Predict Tides at Tanjung Priok Port," *Procedia Comput Sci*, vol. 227, pp. 406–414, 2023, doi: 10.1016/j.procs.2023.10.540.
- [19] B. Santosa and A. L. Safitri, "Biogeography-based Optimization (BBO) Algorithm for Single Machine Total Weighted Tardiness Problem (SMTWTP)," *Procedia Manuf*, vol. 4, pp. 552–557, 2015, doi: 10.1016/j.promfg.2015.11.075.



- [20] S. H. Abrehdari, "Leveraging the Process Mining Technique to Optimize Data Preparation Time in a Database Used as an Automated Data Delivery Center," *MethodsX*, p. 103428, Jun. 2025, doi: 10.1016/j.mex.2025.103428.
- [21] F. Yu, G. Li, H. Chen, Y. Guo, Y. Yuan, and B. Coulton, "A VRF charge fault diagnosis method based on expert modification C5.0 decision tree," *International Journal of Refrigeration*, vol. 92, pp. 106–112, Aug. 2018, doi: 10.1016/j.ijrefrig.2018.05.034.
- [22] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 11, no. 2, pp. 131–145, Sep. 2021, doi: 10.34010/jamika.v11i2.5506.
- [23] Z. Gao, M. Illindala, and J. Lei, "Leveraging data mining for critical branch identification through simultaneity and causality correlation analysis under cascading failures in power systems," *Reliab Eng Syst Saf*, vol. 264, p. 111298, Dec. 2025, doi: 10.1016/j.ress.2025.111298.
- [24] H.-J. Chiang, C.-C. Tseng, and C.-C. Tornig, "A retrospective analysis of prognostic indicators in dental implant therapy using the C5.0 decision tree algorithm," *J Dent Sci*, vol. 8, no. 3, pp. 248–255, Sep. 2013, doi: 10.1016/j.jds.2013.04.009.