

Klasifikasi Tingkat Kepuasan Wali Murid Terhadap Hasil Belajar Anak Menggunakan Algoritma C5.0

Firda Siska Audina¹, Ricky Eka Putra²

^{1,2} S1 Teknik Informatika, Universitas Negeri Surabaya

¹firda.19031@mhs.unesa.ac.id

²rickyeka@unesa.ac.id

Abstrak— Pada era digital saat ini, banyak sekali jasa bimbingan belajar baik itu secara online maupun offline. Banyak pesaing di bidang jasa bimbingan belajar, sehingga diperlukan sebuah analisis yang dapat digunakan untuk memprediksi tingkat kepuasan wali murid terhadap hasil belajar anak di PT. Lesin Aja Edukasi Indonesia dan mengetahui atribut yang dominan dalam mempengaruhi kepuasan wali murid. Analisis kepuasan menggunakan metode survei untuk mengumpulkan data dari 100 responden yang merupakan wali murid dari PT. Lesin Aja Edukasi Indonesia. Responden diminta untuk mengisi kuesioner terkait tingkat kepuasan mereka terhadap hasil belajar anak mereka. Atribut yang digunakan dalam penelitian ini adalah penguasaan materi oleh guru bimbingan belajar, sikap guru bimbingan belajar saat mengajar, kualitas materi relevan dengan mata pelajaran di sekolah, durasi mengajar sesuai dengan harga yang ditawarkan, dan hasil belajar anak ketika di sekolah. Data yang diperoleh dari survei kemudian diolah menggunakan algoritma decision tree C5.0 untuk mengklasifikasikan tingkat kepuasan wali murid menjadi dua kategori, yaitu puas dan tidak puas. Evaluasi dilakukan dengan menggunakan metode confusion matrix cross validation. Hasil analisis menunjukkan bahwa algoritma decision tree C5.0 dapat mengklasifikasikan tingkat kepuasan wali murid dengan tingkat akurasi yang cukup tinggi yaitu sebesar 95%. Model klasifikasi yang dikembangkan dapat membantu PT. Lesin Aja Edukasi Indonesia dalam memprediksi tingkat kepuasan wali murid dan memperbaiki kualitas layanan pendidikan yang diberikan kepada anak-anak mereka.

Kata Kunci— Klasifikasi, Kepuasan, Hasil belajar anak, PT. Lesin Aja Edukasi Indonesia, Algoritma decision tree C5.0.

I. PENDAHULUAN

Pada zaman sekarang kualitas pendidikan adalah hal yang paling utama. Kualitas Pendidikan dapat menentukan bagaimana hasil belajar dari seorang anak. Kualitas Pendidikan yang bagus dapat memberikan ilmu yang mudah dipahami dan diterapkan dalam kehidupan sehari-hari[1]. Untuk menunjang seorang anak dalam meningkatkan kualitas Pendidikan, seringkali orang tua dari murid mempercayakan anaknya untuk mengikuti tambahan bimbingan belajar (bimbingan belajar) diluar sekolah. Semua orang tua tentu ingin anaknya mendapatkan tempat bimbingan belajar dengan kualitas pendidik yang berpengalaman, menguasai materi yang diajarkan, serta sajian materi yang mudah dipahami dan relevan terhadap pelajaran di sekolah anaknya.

Usaha di bidang jasa Pendidikan atau yang lebih akrab disebut dengan bimbingan belajar sudah banyak ditemui saat ini. Persaingan usaha di bidang tersebut juga sangat ketat. Diperlukan suatu Analisis untuk mengetahui faktor apa saja

yang dapat membuat usaha bimbingan belajar memberikan rasa puas bagi wali murid terhadap hasil belajar anaknya yang mengikuti bimbingan belajar.

Salah satu usaha yang bergerak di bidang Pendidikan adalah PT. Lesinaja Edukasi Indonesia. PT. Lesinaja Edukasi Indonesia adalah sebuah bimbingan belajar yang terletak di daerah Sidoarjo. Bimbingan belajar ini menerima murid dari playgroup sampai dengan SMA. PT. Lesinaja Edukasi Indonesia memerlukan suatu analisis untuk mengetahui hal apa saja yang membuat wali murid merasa puas atau tidak puas terhadap hasil belajar anaknya di bimbingan belajar ini.

Untuk menghadapi persaingan yang ketat antar usaha bimbingan belajar lain, diperlukan sebuah analisis kepuasan wali murid terhadap hasil belajar anaknya yang sedang mengikuti bimbingan belajar di tempat ini. Analisis ini untuk membantu PT. Bimbingan Belajar Lesinaja mengetahui hal apa yang membuat wali murid puas dan yang membuat wali murid tidak puas dari wali murid selama anaknya mengikuti bimbingan belajar di sini. Sehingga dengan adanya Analisis ini PT. Lesinaja Edukasi Indonesia dapat meningkatkan apa yang dirasa kurang dalam pembelajaran selama bimbingan belajar[2]. Beberapa variabel yang digunakan untuk menganalisis kepuasan wali murid adalah tingkat penguasaan materi oleh guru, sikap guru bimbingan belajar, kualitas materi relevan atau tidak dengan pelajaran di sekolah, durasi mengajar sesuai dengan harga yang ditawarkan oleh bimbingan belajar dan hasil belajar anak ketika di sekolah.

Penelitian yang akan saya lakukan menggunakan metode algoritma C5.0 yang merupakan regenerasi dari algoritma C4.5 dan ID3. Pada algoritma C5.0 tingkat penghitungan keakuratan lebih tinggi dibanding dengan algoritma lain. Pada algoritma C5.0 terdapat fase terbaru dari algoritma sebelumnya, yaitu fase boosting dengan menghitung gain ratio. Penerapan teknik boosting dalam algoritma C5.0 dapat meningkatkan akurasi dari decision tree yang dihasilkan. Hal ini terjadi karena dalam fase boosting, bobot pada data latihan terus diperbarui, yang digunakan sebagai acuan dalam membangun decision tree. Dengan demikian, proses ini membantu meningkatkan keoptimalan decision tree yang dihasilkan oleh algoritma C5.0 melalui pembaruan berkelanjutan terhadap data latihan yang digunakan [3]. Data training yang sudah diperbarui akan diolah dengan perhitungan algoritma decision tree, sehingga model dari pohon keputusan akan terbaharui.

Pembuatan pohon keputusan dan perhitungan akurasi Algoritma C5.0 akan saya implementasikan ke dalam program Python. Selain menggunakan program Python, saya akan membuat sebuah sistem yang dapat digunakan oleh user untuk

mengambil keputusan apakah puas atau tidak puas dengan layanan PT.Lesinaja Edukasi Indonesia terhadap hasil belajar anaknya.

Variabel yang digunakan dalam mengambil data kepuasan wali murid di PT.Lesinaja Edukasi Indonesia diantaranya yaitu tingkat penguasaan materi oleh guru, sikap guru bimbeler, kualitas materi relevan atau tidak dengan pelajaran disekolah, durasi mengajar sesuai dengan harga yang ditawarkan oleh bimbeler dan hasil belajar anak ketika disekolah[3].

II. METODOLOGI PENELITIAN

Berikut adalah flowchart dari metodologi penelitian :



Gbr 1. Flowchart Algoritma C5.0

A. Algoritma C5.0

C5.0 atau See5 adalah versi yang diperbarui dari C4.5, yang pada awalnya mengadopsi aturan yang digunakan oleh ID3. C5 juga mengikuti prinsip-prinsip yang ada dalam C4.5, sehingga memiliki peraturan yang hampir serupa dengan algoritma C4.5. Seperti halnya C4.5, algoritma C5 juga menyediakan fitur seleksi, validasi silang, dan pemangkasan kesalahan untuk mengurangi kesalahan [5]. C5.0 memiliki empat fitur yang mencakup: (a) kemampuan untuk melihat decision tree yang kompleks sebagai aturan-aturan yang mudah dimengerti, (b) kemampuan untuk menghadapi kebisingan dan data yang hilang, (c) penyelesaian masalah overfitting dan error pruning yang ditingkatkan oleh algoritma C5.0, dan (d) kemampuan pengklasifikasian C5.0 untuk

memprediksi atribut yang relevan dan tidak relevan. Berikut ini adalah tabel perbandingan antara algoritma C4.5 dan C5.0 [7]:

Tabel I
Perbedaan Algoritma C4.5 dan Algoritma C5.0

Kategori	Algoritma C4.5	Algoritma C5.0
Akurasi	Memiliki tingkat error pruning yang tinggi	Memiliki error pruning lebih rendah karena terdapat fase menghitung gain ratio
Speed	Membutuhkan waktu yang lama dalam menemukan rules tree, yaitu 8 jam	Membutuhkan waktu yang cepat dalam menemukan rules tree, yaitu 3 menit
Memory	Memori yang digunakan mencapai 3 gb	Memori yang digunakan hanya 200 mb

Tahapan dalam membuat pohon keputusan algoritma C5.0 yaitu :

- 1) Mempersiapkan data training yang diambil dari kuisioner dan dikelompokkan dalam kelas-kelas tertentu
- 2) Menghitung akar pohon. Akar akan diambil dari atribut yang akan dipilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi akan menjadi akar pertama.
- 3) Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy[5]. Untuk menghitung nilai entropy digunakan rumus:

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

Keterangan :

S : Himpunan Kelas

n : Jumlah partisi S

pi : proporsi dari Si terhadap S

- 4) Setelah nilai entropy masing-masing dari atribut telah diperoleh, selanjutnya menghitung information gain digunakan rumus :

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

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

- 5) Apabila pada algoritma C4.5 berhenti pada perhitungan information Gain, maka algoritma C5.0 ini dilanjutkan pada tahap perhitungan gain ratio digunakan rumus :

$$Gain\ Ratio = \frac{Information\ Gain\ (S,A)}{\sum_{i=1}^n Entropy\ (S_i)}$$

- 6) Ulangi proses perhitungan gain ratio sampai seluruh cabang memiliki kelas masing-masing. Atribut yang telah dipilih tidak dimasukkan lagi kedalam perhitungan nilai gain ratio pada iterasi berikutnya.

B. Decision Tree

Decision Tree merupakan metode yang sangat umum digunakan dalam klasifikasi untuk memecahkan berbagai masalah. Decision Tree adalah struktur yang mirip dengan flowchart di mana setiap node internal (bukan leaf atau node terluar) mewakili pengujian pada atribut variabel. Setiap cabang dari node tersebut mencerminkan hasil pengujian, sementara node terluar atau leaf menjadi labelnya [8].

Sebuah decision tree terdiri dari beberapa node, termasuk akar pohon (root), node internal, dan daun (leafs). Konsep entropi digunakan untuk menentukan bagaimana pohon akan dibagi (split) berdasarkan atribut tertentu. Dalam decision tree, setiap node internal membagi ruang menjadi dua atau lebih berdasarkan nilai atribut input yang diskrit. Dalam kasus yang paling sederhana dan umum, setiap tes menganggap atribut tunggal dan mempartisi ruang sesuai dengan nilai atribut yang ada. Klasifikasi menggunakan decision tree dilakukan dengan mengarahkan rute dari akar pohon hingga mencapai daun yang sesuai. Beberapa algoritma decision tree yang umum digunakan termasuk ID3, C4.5, C5.0, dan CART.

C. Teknik Kategorisasi Kepuasan

Penyekoran kepuasan dibagi menjadi dua jenjang, puas dan tidak puas. Perhitungan kategorisasi dengan dua jenjang diperoleh dari rumus [6] :

- 4 = Baik
- 3 = Cukup Baik
- 2 = Kurang Baik
- 1 = Tidak Baik

Tabel II
 Kategorisasi Kepuasan Dua Jenjang

Rumus	Kategori
$X \geq \mu$	Puas
$X < \mu$	Tidak Puas

Keterangan :
 X : Skor mentah sampel
 μ : rata-rata distribusi dalam populasi

D. Python

Python merupakan sebuah bahasa pemrograman berbasis objek yang sering digunakan sebagai bahasa skrip. Python sangat fleksibel dalam pengembangan perangkat lunak dan dapat digunakan pada berbagai sistem operasi yang berbeda. Selain itu, Python disediakan secara gratis atau freeware, sehingga dapat diakses dan digunakan tanpa dikenakan biaya, yang berarti tidak ada batasan dalam mengkopi atau mendistribusikannya. Python dilengkapi dengan source code, debugger, dan profiler, serta menyediakan antarmuka untuk layanan sistem, antarmuka pengguna grafis (GUI), dan basis data [9]. Beberapa fitur yang dimiliki Python adalah:

- 1) Python memiliki koleksi pustaka yang luas; dalam distribusi Python, terdapat modul-modul yang tersedia.
- 2) Python memiliki struktur bahasa yang jelas dan mudah dipelajari.
- 3) Python memiliki aturan tata letak kode sumber yang mempermudah pemeriksaan, membaca ulang, dan penulisan ulang kode sumber.
- 4) Python berbasis objek.
- 5) Python dapat dikembangkan menggunakan bahasa Python maupun C/C++

E. Evaluasi Model

Evaluasi model sangat penting dilakukan untuk meng cross check data yang sudah dihitung valid atau tidak. Evaluasi model menggunakan metode confusion matrix atau yang disebut dengan cross validation. Proses evaluasi model dengan metode confusion matrix adalah dengan membagi data set ke dalam beberapa persen di data training dan beberapa persen di data testing [3]. Di dalam confusion matrix terdapat 4 (empat) istilah representasi hasil klasifikasi yaitu :

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gbr 2. Confussion Matrix

1. True Positive (TP)
Memprediksi positif dan itu benar
2. True Negative (TN)
Memprediksi negative dan itu benar
3. False Positive (FP)
Memprediksi Positif dan itu salah
4. False Negative (FN)
Memprediksi Negatif dan itu salah

III. HASIL DAN PEMBAHASAN

Terdapat 4 bagian yang akan dijabarkan pada hasil dan pembahasan yaitu pemodelan matematis algoritma C5.0, implementasi decision tree algoritma C5.0 pada python, implementasi sistem kepuasan berbasis website, dan hasil pengujian website sistem kepuasan wali murid PT.Lesinaja Edukasi Indonesia..

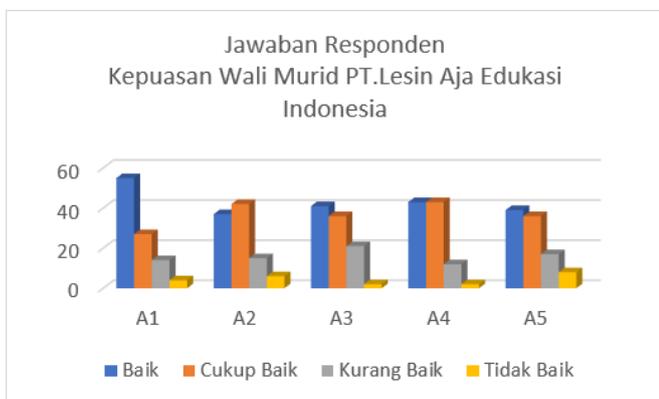
A. Pemodelan Matematis Algoritma C5.0

Penelitian ini didukung dengan data yang didapat dari kuisioner yang disebarakan kepada wali murid dari anak didik PT.Lesinaja Edukasi Indonesia. Gbr 1. adalah data hasil kuisioner kepuasan wali murid.



Gbr 3. Grafik Hasil Kuisioner

Gbr 1. merupakan grafik dari data hasil kuisioner yang telah disebarakan. Terdapat 100 responden yang mengisi data dan setiap responden mengisi atribut A1,A2,A3,A4, dan A5 secara lengkap.



Gbr 4. Grafik Jawaban Responden

Gbr 2. ditunjukkan bahwa terdapat jumlah dari masing-masing responden yang mengisi setiap pertanyaan dengan parameter penilaian yang berbeda-beda. Tabel III merupakan jumlah responden yang mengisi kuisioner dengan parameter yang berbeda.

Tabel III
 Rekapitulasi Jawaban Responden

	Baik	Cukup Baik	Kurang Baik	Tidak Baik
A1	55	27	14	4
A2	37	42	15	6
A3	41	36	21	2
A4	43	43	12	2
A5	39	36	17	8

Dari gbr 3. dapat dilihat bahwa jumlah responden yang puas terdapat 83 orang dan yang tidak puas terdapat 17 orang. Cuplikan dataset dapat dilihat pada tabel II.



Gbr 5. Grafik Pelabelan Data

Tabel IV
 Cuplikan Dataset 10 dari 100 Responden

No	A1	A2	A3	A4	A5	Hasil
1.	Baik	Baik	Baik	Baik	Baik	Puas
2.	Baik	Cukup Baik	Baik	Baik	Baik	Puas
3.	Baik	Kurang Baik	Cukup Baik	Baik	Cukup Baik	Puas
4.	Cukup Baik	Cukup Baik	Baik	Baik	Baik	Puas
5.	Baik	Baik	Kurang Baik	Baik	Kurang Baik	Puas
6.	Cukup Baik	Cukup Baik	Baik	Baik	Baik	Puas
7.	Baik	Baik	Cukup Baik	Baik	Cukup Baik	Puas
8.	Cukup Baik	Kurang Baik	Kurang Baik	Baik	Kurang Baik	Puas

9.	Kurang Baik	Kurang Baik	Kurang Baik	Cukup Baik	Kurang Baik	Tidak Puas
10.	Baik	Baik	Cukup Baik	Cukup Baik	Cukup Baik	Puas

Sebelum dilakukan perhitungan algoritma C5.0, dilakukan split data untuk mengetahui pembagian data yang memiliki tingkat akurasi paling tinggi. Data yang memiliki akurasi paling tinggi akan digunakan untuk membentuk decision tree algoritma C5.0. Hasil dari perhitungan akurasi dapat dilihat pada tabel 3.

Tabel V
Hasil Analisis Akurasi Dengan Python

No.	Split Data	Hasil
1.	60% : 40%	91%
2.	65% : 35%	90%
3.	70% : 30%	92%
4.	80% : 20%	95%
5.	85% : 15%	93%

Dari hasil analisis akurasi menggunakan python pada tabel III, split data dengan data training 80% dan data testing 20% memiliki hasil akurasi paling tinggi dibandingkan dengan split data yang lain. Hasil akurasi yang diperoleh adalah 95%.

Setelah diketahui split data yang paling tinggi akurasinya adalah 80% dan 20% maka dilakukan perhitungan algoritma C5.0 secara manual. Di bawah ini akan dijabarkan cara menghitung entropy, information gain, dan gain ratio untuk membentuk sebuah pohon keputusan.

1) Menghitung Entorpy masing-masing atribut

$$\text{Entropy Total} = \left(-\frac{63}{80} \times \log_2\left(\frac{63}{80}\right)\right) + \left(-\frac{17}{80} \times \log_2\left(\frac{17}{80}\right)\right)$$

$$= 0.746234301$$

Entropy A1 :

Nilai atribut "B" =

$$\left(-\frac{42}{47} \times \log_2\left(\frac{42}{47}\right)\right) + \left(-\frac{0}{47} \times \log_2\left(\frac{0}{47}\right)\right) = 0$$

Nilai atribut "CB" =

$$\left(-\frac{19}{21} \times \log_2\left(\frac{19}{21}\right)\right) + \left(-\frac{2}{21} \times \log_2\left(\frac{2}{21}\right)\right) = 0.453716339$$

Nilai atribut "KB" =

$$\left(-\frac{2}{12} \times \log_2\left(\frac{2}{12}\right)\right) + \left(-\frac{11}{12} \times \log_2\left(\frac{11}{12}\right)\right) = 0.619382195$$

Nilai atribut "TB" =

$$\left(-\frac{0}{4} \times \log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} \times \log_2\left(\frac{4}{4}\right)\right) = 0$$

Entropy A2 :

Nilai atribut "B" =

$$\left(-\frac{31}{31} \times \log_2\left(\frac{31}{31}\right)\right) + \left(-\frac{0}{31} \times \log_2\left(\frac{0}{31}\right)\right) = 0$$

Nilai atribut "CB" =

$$\left(-\frac{27}{30} \times \log_2\left(\frac{27}{30}\right)\right) + \left(-\frac{3}{30} \times \log_2\left(\frac{3}{30}\right)\right) = 0.468995594$$

Nilai atribut "KB" =

$$\left(-\frac{5}{13} \times \log_2\left(\frac{5}{13}\right)\right) + \left(-\frac{8}{13} \times \log_2\left(\frac{8}{13}\right)\right) = 0.961236605$$

Nilai atribut "TB" =

$$\left(-\frac{0}{6} \times \log_2\left(\frac{0}{6}\right)\right) + \left(-\frac{6}{6} \times \log_2\left(\frac{6}{6}\right)\right) = 0$$

Entropy A3 :

Nilai atribut "B" =

$$\left(-\frac{30}{30} \times \log_2\left(\frac{30}{30}\right)\right) + \left(-\frac{0}{30} \times \log_2\left(\frac{0}{30}\right)\right) = 0$$

Nilai atribut "CB" =

$$\left(-\frac{27}{28} \times \log_2\left(\frac{27}{28}\right)\right) + \left(-\frac{1}{28} \times \log_2\left(\frac{1}{28}\right)\right) = 0.222284831$$

Nilai atribut "KB" =

$$\left(-\frac{6}{20} \times \log_2\left(\frac{6}{20}\right)\right) + \left(-\frac{14}{20} \times \log_2\left(\frac{14}{20}\right)\right) = 0.881290899$$

Nilai atribut "TB" =

$$\left(-\frac{0}{2} \times \log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right)\right) = 0$$

Entropy A4 :

Nilai atribut "B" =

$$\left(-\frac{33}{33} \times \log_2\left(\frac{33}{33}\right)\right) + \left(-\frac{0}{33} \times \log_2\left(\frac{0}{33}\right)\right) = 0$$

Nilai atribut "CB" =

$$\left(-\frac{27}{35} \times \log_2\left(\frac{27}{35}\right)\right) + \left(-\frac{8}{35} \times \log_2\left(\frac{8}{35}\right)\right) = 0.775512658$$

Nilai atribut "KB" =

$$\left(-\frac{3}{10} \times \log_2\left(\frac{3}{10}\right)\right) + \left(-\frac{7}{10} \times \log_2\left(\frac{7}{10}\right)\right) = 0.881290899$$

Nilai atribut "TB" =

$$\left(-\frac{0}{2} \times \log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right)\right) = 0$$

Entropy A5 :

$$\text{Nilai atribut "B"} = \left(-\frac{32}{32} \times \log_2\left(\frac{32}{32}\right)\right) + \left(-\frac{0}{32} \times \log_2\left(\frac{0}{32}\right)\right) = 0$$

Nilai atribut "CB" =

$$\left(-\frac{25}{25} \times \log_2\left(\frac{25}{25}\right)\right) + \left(-\frac{0}{25} \times \log_2\left(\frac{0}{25}\right)\right) = 0$$

Nilai atribut "KB" =

$$\left(-\frac{6}{15} \times \log_2\left(\frac{6}{15}\right)\right) + \left(-\frac{9}{15} \times \log_2\left(\frac{9}{15}\right)\right) = 0.970950594$$

Nilai atribut "TB" =

$$\left(-\frac{0}{8} \times \log_2\left(\frac{0}{8}\right)\right) + \left(-\frac{8}{8} \times \log_2\left(\frac{8}{8}\right)\right) = 0$$

2) Menghitung information gain

Nilai Information Gain A1 =

$$0.746234301 - \left(\left(\frac{42}{80}\right) \times 0\right) + \left(\left(\frac{21}{80}\right) \times 0.453716339\right) + \left(\left(\frac{13}{80}\right) \times 0.619382195\right) + \left(\left(\frac{4}{80}\right) \times 0\right) = 0.526483001$$

Nilai Information Gain A2 =

$$0.746234301 - \left(\left(\frac{31}{80}\right) \times 0\right) + \left(\left(\frac{30}{80}\right) \times 0.468995594\right) + \left(\left(\frac{13}{80}\right) \times 0.961236605\right) + \left(\left(\frac{6}{80}\right) \times 0\right) = 0.414160301$$

Nilai Information Gain A3 =

$$0.746234301 - \left(\left(\frac{30}{80}\right) \times 0\right) + \left(\left(\frac{28}{80}\right) \times 0.222284831\right) + \left(\left(\frac{20}{80}\right) \times 0.881290899\right) + \left(\left(\frac{2}{80}\right) \times 0\right) = 0.448111601$$

Nilai Information Gain A4 =

$$0.746234301 - \left(\left(\frac{33}{80}\right) \times 0\right) + \left(\left(\frac{35}{80}\right) \times 0.775512658\right) + \left(\left(\frac{10}{80}\right) \times 0.881290899\right) + \left(\left(\frac{2}{80}\right) \times 0\right) = 0.296786301$$

Nilai Information Gain A5 =

$$0.746234301 - \left(\left(\frac{32}{80}\right) \times 0\right) + \left(\left(\frac{25}{80}\right) \times 0\right) + \left(\left(\frac{15}{80}\right) \times 0.970950594\right) + \left(\left(\frac{8}{80}\right) \times 0\right) = 0.564181301$$

3) Menghitung gain ratio

Nilai gain ratio A1

$$= \left(\frac{0.526483001}{1.073098534}\right) = 0.490619$$

Nilai gain ratio A2

$$= \left(\frac{0.414160301}{1.430232199}\right) = 0.289576$$

Nilai gain ratio A3

$$= \left(\frac{0.448111601}{1.10357573}\right) = 0.406054$$

Nilai gain ratio A4

$$= \left(\frac{0.296786301}{1.656803557}\right) = 0.179132$$

Nilai gain ratio A5

$$= \left(\frac{0.564181301}{0.970950594}\right) = 0.581061$$

Dari perhitungan diatas, Atribut A5 memiliki nilai gain ratio tertinggi, sehingga atribut A5 menjadi node I dari pohon keputusan. Selanjutnya dicari cabang dari setiap value dengan cara yang sama yaitu menghitung entropy, information gain, dan gain ratio.

Tabel VI
Rekapitulasi Hasil Node I

Node	Keterangan	Jumlah Kasus	Puas	Tidak Puas	Entropy	Information Gain	Gain Ratio
1	Total Kasus	80	63	17	0.746234301		
	A1					0.526483001	0.490619
	B	42	42	0	0		
	CB	21	19	2	0.453716339		
	KB	13	2	11	0.619382195		
A2	TB	4	0	4	0		
	B					0.414160301	0.289576
	B	31	31	0	0		
	CB	30	27	3	0.468995594		
	KB	13	5	8	0.961236605		
A3	TB	6	0	6	0		
	B					0.448111601	0.406054
	B	30	30	0	0		
	CB	28	27	1	0.222284831		
	KB	20	6	14	0.881290899		
A4	TB	2	0	2	0		
	B					0.296786301	0.179132
	B	33	33	0	0		
	CB	35	27	8	0.775512658		
	KB	10	3	7	0.881290899		
A5	TB	2	0	2	0		
	B					0.564181301	0.581061
	B	32	32	0	0		
	CB	25	25	0	0		
	KB	15	6	9	0.970950594		
	TB	8	0	8	0		

B. Implementasi Decision Tree Algoritma C5.0 Pada Python
Terdapat beberapa tahapan dalam membuat sebuah decision tree menggunakan python. Berikut adalah tahapan-tahapan yang harus dilakukan secara runtut :

1) Import libraries python

```
# import library
import pandas as pd
from sklearn.model_selection import
train_test_split
from sklearn.tree import
DecisionTreeClassifier, export_graphviz
import graphviz
```

Kode diatas merupakan penggunaan beberapa pustaka (libraries) dalam bahasa pemrograman Python untuk melakukan analisis data dan pembelajaran mesin (machine learning).Pustaka yang diimport adalah[6]:

- Pandas: digunakan untuk manipulasi dan analisis data.
- Numpy: digunakan untuk operasi matematika dan manipulasi array.
- Sklearn.tree: digunakan untuk membangun model pohon keputusan.
- Sklearn.model_selection.train_test_split: digunakan untuk membagi dataset menjadi data latih dan data uji untuk evaluasi model.
- Import graphviz : digunakan untuk menghasilkan gambar pohon keputusan

2) Import dataset dalam format “csv”

```
# membaca dataset
data = pd.read_csv('datasetfix2.csv')
```

Fungsi yang digunakan adalah read_csv() yang memungkinkan untuk membaca file CSV dan mengonversinya menjadi DataFrame, yaitu struktur data berbentuk tabel yang sering digunakan dalam analisis data[1].

3) Membagi data training dan data testing. Pada perhitungan ini data training diambil 80% dan data testing 20%[4].

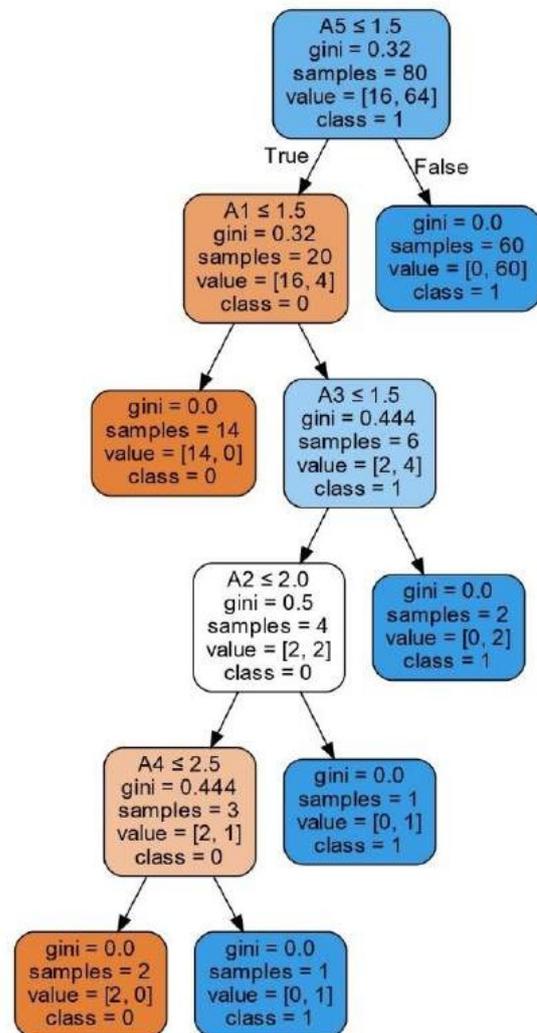
```
# membagi data menjadi training dan testing
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2,
random_state=465)
```

4) Membuat model pohon keputusan kepuasan walimurid dengan algoritma C5.0

```
# membuat model dengan algoritma C5.0
model = DecisionTreeClassifier(criterion='gini',
max_depth=5)
model.fit(X_train, y_train)
mymodel = model.fit(X_train, y_train)
```

5) Visualisasi pohon keputusan kepuasan wali murid Algoritma C5.0

Gbr 4. adalah bentuk pohon keputusan dari kepuasan wali murid PT.Lesinaja Edukasi Indonesia menggunakan python :



Gbr 6. Decision Tree

Pada gbr 4. Ditunjukkan bahwa atribut yang memiliki nilai gain ratio tertinggi adalah atribut A5, yaitu nilai anak ketika di sekolah setelah mengikuti bimbel. Hasil dari pohon keputusan ini akan di deploy di website sistem kepuasan.

6) Import library pickle untuk mendeploy decision tree ke dalam website

```
import pickle
filename = 'Kepuasan_walmur.sav'
pickle.dump(model, open(filename, 'wb'))
```

Mengimport library pickle untuk menyimpan objek hasil decision tree yang nantinya dapat di deploy ke dalam website. Setelah mengimport, simpan hasil decision tree dalam bentuk “sav”.

7) Evaluasi model algoritma C5.0

```
from sklearn.metrics import confusion_matrix

# melakukan prediksi pada data pengujian
y_pred = model.predict(X_test)

# menghitung matriks konfusi
cm = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:")
print(cm)
```

Berikut ini bentuk confusion matrix akurasi data :

```
Confusion Matrix:
[[ 2  0]
 [ 1 17]]
```

Gbr 7. Tabel confusion matrix

Tabel confusion matrix yang dihasilkan yaitu kondisi true positif(TP) 5, true negative(TN) 14, false positive(FP) 0, dan false negative(FN) 1. Setelah didapatkan hasil tabel confusion matrix, maka dilanjutkan perhitungan akurasi.

```
accuracy = (cm[0, 0] + cm[1, 1]) / cm.sum()
print("Akurasi:", accuracy)

Akurasi: 0.95
```

Gbr 8. Hasil Akurasi

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{5+14}{5+14+0+1}$$

$$\text{Akurasi} = \frac{19}{20} = 0.95$$

Akurasi mengukur seberapa banyak prediksi yang benar dibuat oleh model dalam hal ini model decision tree dengan kriteria gini. Semakin tinggi akurasi, semakin baik kinerja model[5]. Pada studi kasus klasifikasi kepuasan wali murid PT.Lesinaja Edukasi Indonesia menggunakan algoritma C5.0 mendapatkan tingkat akurasi sebesar 0.95. Maka, algoritma C5.0 cocok digunakan pada studi kasus ini.

Evaluasi model juga dilakukan menggunakan K-Fold Cross Validation. Berikut ini adalah hasil akurasi menggunakan K-Fold Cross Validation :

```
from sklearn.model_selection import
cross_val_score

#menghitung k-fold cross validation
score =
cross_val_score(mymodel,X_train,y_train,cv = 12)
score
```

Pada percobaan diatas dilakukan sebanyak 12 kali perhitungan akurasi. Tabel V merupakan hasil akurasi menggunakan k12-fold cross validation :

Tabel VII
 K12-Fold Cross Validation

K1	1
K2	1
K3	1
K4	1
K5	0.85714286
K6	0.85714286
K7	1
K8	0.85714286
K9	0.83333333
K10	1
K11	1
K12	1

Setelah ditemukan hasil akurasi tersebut dihitung rata-ratanya seperti dibawah ini :

```
import numpy as np

#menghitung rata-rata k12-fold cross validation
print(np.mean(score))

0.9503968253968255
```

Gbr 9. Mean K12-Fold Cross Validation

C. Implementasi Sistem Kepuasan Berbasis Website

Implementasi sistem kepuasan wali murid PT.Lesinaja Edukasi Indonesia dibuat dalam sebuah website sederhana yang mudah digunakan untuk user. Dibawah ini adalah bentuk website yang telah dibuat :

- 1) Mendeploy rules decision tree ke website menggunakan fungsi pickle

```
model=pickle.load(open('Kepuasan_walmur.sav
```

2) Halaman Sistem Kepuasan Wali Murid



Gbr 10. Halaman Sistem Kepuasan Wali Murid

Pada halaman ini wali murid diwajibkan untuk mengisi nama. Setelah mengisi nama, wali murid harus mengisi 5 aspek untuk mengukur kepuasan terhadap hasil belajar anak. Instruksi telah dijelaskan yaitu: 1 = Tidak Baik, 2= Kurang Baik, 3= Cukup Baik, 4= Baik. Setelah ke 5 aspek telah di isi, lalu klik tombol “Lihat hasil analisis”. Ketika tombol “Lihat hasil analisis” di klik, akan ditampilkan hasil kepuasan wali murid Puas/Tidak Puas.

3) Hasil Kepuasan Wali Murid



Gbr 11. Halaman hasil puas wali murid

Setelah wali murid mengisi ke 5 aspek, maka akan keluar hasil puas/tidak puas. Contoh seperti gbr 5. yaitu, Wali Murid Bernama Angga mengisi A1 dengan nilai 3, A2 dengan nilai 4, A3 dengan nilai 4, A4 dengan nilai 4, dan A5 dengan nilai 3, maka hasil akhir yang di dapat adalah wali murid puas dengan hasil belajar anak di PT.LesinAja Edukasi Indonesia.



Gbr 12. Halaman hasil tidak puas wali murid

Setelah wali murid mengisi ke 5 aspek, maka akan keluar hasil puas/tidak puas. Contoh seperti gbr 6. yaitu, Wali Murid Bernama Linda mengisi A1 dengan nilai 2, A2 dengan nilai 1, A3 dengan nilai 2, A4 dengan nilai 1, dan A5 dengan nilai 2, maka hasil akhir yang di dapat adalah wali murid tidak puas dengan hasil belajar anak di PT.LesinAja Edukasi Indonesia.

D. Hasil Pengujian Website Sistem Kepuasan Wali Murid PT.Lesinaja Edukasi Indonesia

Setelah website siap digunakan, dilakukan pengujian menggunakan blackbox testing. Blackbox testing dilakukan secara internal oleh pembuat sistem serta 5 user lainnya. Hasil pengujian dapat dilihat pada tabel VI :

Tabel VIII
 Hasil Pengujian Sistem

No.	Indikator capaian keberhasilan	Berhasil	Tidak
1.	Dapat menampilkan pertanyaan survey	√	
2.	Dapat menampilkan pilihan jawaban survey	√	
3.	Dapat memilih salah satu jawaban survey	√	
4.	Dapat menampilkan jawaban yang sudah di pilih	√	
5.	Dapat menganalisis jawaban yang sudah di pilih	√	
6.	Dapat menampilkan hasil klasifikasi tingkat kepuasan wali murid	√	

Pengujian berjalan dengan baik, semua fungsi berjalan dengan semestinya. User dapat mendapatkan hasil puas atau tidak puas setelah mengisi seluruh pertanyaan

IV. KESIMPULAN

Dari penelitian diatas dapat disimpulkan bahwa :

1. Atribut dengan urutan yang paling dominan dalam mempengaruhi kepuasan wali murid terhadap belajar anak di PT.Lesinaja Edukasi Indonesia yaitu, nilai siswa baik/tidak setelah mengikuti bimbingan, tingkat penguasaan mentor Ketika mengajar, materi yang disampaikan mentor relevan/tidak dengan pembelajaran di sekolah, sikap guru bimbingan Ketika

- mengajar, dan durasi mengajar sesuai dengan harga yang ditawarkan.
2. Performa algoritma C5.0 dalam studi kasus ini sangat bagus dengan akurasi 0.95. Sehingga dapat disimpulkan bahwa metode ini cocok untuk menghitung klasifikasi kepuasan seperti penelitian-penelitian sebelumnya.
 3. Implementasi website sistem kepuasan membuat proses survey kepuasan menjadi lebih efektif karena tidak perlu lagi menyebarkan selebaran kertas door to door.

V. SARAN

Penelitian ini jauh dari kata sempurna, sehingga penulis menyarankan hal berikut :

1. Melakukan penelitian di tempat yang sama dengan menggunakan metode yang berbeda.
2. Melakukan penelitian dengan objek yang berbeda dengan menggunakan algoritma C5.0.
3. Menambahkan atribut penelitian agar proses dalam klasifikasi lebih mendetail.
4. Mengembangkan website menjadi beberapa fitur yang relevan.

UCAPAN TERIMA KASIH

Dengan menyebut nama Allah SWT yang maha pengasih lagi maha penyayang, kuucapkan terima kasih kepada Bapak/Ibu yang sudah mendukung karya ini sampai selesai:

1. Ayahandaku Sunaryo dan ibuku Lilik Widyaningsih, Semangatnya yang selalu membara, Doa yang tidak pernah putus sepanjang masa dan orang yang selalu support saya dari segi moral dan materil demi kesuksesanku ini.
2. Terima kasih kepada Bapak Dr. Ricky Eka Putra, S.Kom., M.Kom. selaku dosen pembimbing skripsi saya, yang telah membimbing saya dari nol sampai terbentuknya laporan skripsi ini.

3. Terima kasih kepada Ibu Anita Qoiriah, S.Kom., M.Kom. selaku dosen penguji I yang telah memberikan saran dan masukan dalam skripsi ini.
4. Terimakasih kepada Ibu Dr. Yuni Yamasari, S.Kom., M.Kom. selaku dosen Penguji II skripsi yang telah memberikan saran dan masukan.

REFERENSI

- [1] Aldi, F., & Ade Rahma, A. (2019). University Student Satisfaction Analysis on Academic Services by Using Decision Tree C4.5 Algorithm (Case Study : Universitas Putra Indonesia "yPTK" Padang). *Journal of Physics: Conference Series*, 1339(1). <https://doi.org/10.1088/1742-6596/1339/1/012051>
- [2] Nasrullah, A. H. (2021). IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS. 7(2).
- [3] Sutoyo, I. (2018). IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI DATA PESERTA DIDIK. 14(2).
- [4] Wibawa, A. P., Guntur, M., Purnama, A., Fathony Akbar, M., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1).
- [5] Sihombing, J. H. (2022). Student Satisfaction Level Analysis Of Online Learning During Pandemic Covid 19 Using C5. 0 Algorithm. *Journal of Artificial Intelligence and Engineering ...*, 1(3). Retrieved from <https://www.ioinformatic.org/index.php/JAIEA/article/view/90%0Ahttps://www.ioinformatic.org/index.php/JAIEA/article/download/90/121>.
- [6] Widiyanti, R., Suhery, C., & Hidayati, R. (2022). Implementasi Algoritma C5.0 Untuk Klasifikasi Kepuasan Masyarakat Terhadap Pelayanan Kantor Kecamatan. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 1200. <https://doi.org/10.30865/jurikom.v9i4.4632>.
- [7] Pandya, Rutvija & Pandya, Jayati. (2019). C5. 0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning. *International Journal of Computer Applications*. 117. 18-21. 10.5120/20639-3318.
- [8] Harani, N. H., & Damayanti, F. S. (2021). Implementasi Algoritma C5.0 Untuk Menentukan Pelanggan Potensial Di Kantor Pos Cimahi. *Jurnal SITECH : Sistem Informasi Dan Teknologi*, 4(1), 69–76. <https://doi.org/10.24176/sitech.v4i1.6281>.
- [9] Triayudi, A. (2022). Penerapan Algoritma C5 . 0 Data Mining Untuk Mengetahui Pola Kepuasan Mahasiswa Terhadap Pelayanan Akademik. 6, 2361–2366. <https://doi.org/10.30865/mib.v6i4.4961>.
- [10] Bayu Febriyanto, D., Handoko, L., & Aisyah, H. (2018). Implementasi Algoritma C4.5 Untuk Klasifikasi Tingkat Kepuasan Pembeli Online Shop. *Jurnal Riset Komputer (JURIKOM)*, 5(6), 569–575. Retrieved from <http://ejournal.stmik-budidarma.ac.id/index.php/jurikom%7CPage%7C569>.