

## Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Mengukur Tingkat Kepuasan Mahasiswa Dalam Penggunaan Edlink

Dafwen Toresa<sup>a\*</sup>, Ikhsan Hidayat<sup>b</sup>, Edriyansyah<sup>c</sup>, Rometdo Muzawi<sup>d</sup>, Taslim<sup>e</sup>, Lisnawita<sup>f</sup>, Febi Yanto<sup>g</sup>

<sup>a</sup> Fakultas Ilmu Komputer, Universitas Lancang Kuning, dafwen@unilak.ac.id\*

<sup>b</sup> Fakultas Ilmu Komputer, Universitas Lancang Kuning, ikhsanhidayat2008@gmail.com

<sup>c</sup> Fakultas Ilmu Komputer, Universitas Hang Tuah Pekanbaru, sinksonk@gmail.com

<sup>d</sup> STMIK-AMIK Riau, rometdomuzawi@stmik-amik-riau.ac.id

<sup>e</sup> Fakultas Ilmu Komputer, Universitas Lancang Kuning, taslim@unilak.ac.id

<sup>f</sup> Fakultas Ilmu Komputer, Universitas Lancang Kuning, lisnawita@unilak.ac.id

<sup>g</sup> Fakultas Ilmu Komputer, Universitas Islam Negeri SUSKA, febiyanto@uin-suska.ac.id

Submitted: 08-06-2023, Reviewed: 20-06-2023, Accepted 02-07-2023

<https://doi.org/10.47233/jteksis.v5i3.855>

### Abstract

The Faculty of Computer Science, Lancang Kuning University, as a private university in the city of Pekanbaru, uses the Sevima Edlink platform as a media for academic information systems and online learning. According to some students, there are still some obstacles encountered in understanding, using and functioning this edlink application. The purpose of this study was to measure the level of satisfaction of students of the Faculty of Computer Science in using Edlink using the C4.5 and Naïve Bayes algorithms. To measure the level of accuracy of the C4.5 and Naïve Bayes algorithms in order to measure the level of student satisfaction, the indicators used are the Servqual testing model, namely Tangible, Reability, Responsiveness, Assurance, and Empathy. Based on the level of accuracy of the two methods. In the dataset used there were 91 student respondents who had filled out the questionnaire. From the questionnaire data, it was then processed using both methods and 9 comparisons of the different Training Data and Testing Data were carried out. In general, students are satisfied and understand the use of the edlink application. This satisfaction was tested using the C4.5 Decision Tree Algorithm and the Naïve Bayes Classifier. Based on the comparison that has been carried out using the C4.5 Decision Tree Algorithm, it produces an average accuracy value of 77.78%, which is slightly more accurate than the Naïve Bayes Classifier which produces an average accuracy value of 71.11%.

**Keywords:** Measuring satisfaction, Edlink, C4.5, Naïve Bayes, Accuracy

### Abstrak

Fakultas Ilmu Komputer Universitas Lancang Kuning sebagai salah satu perguruan tinggi swasta di kota Pekanbaru menggunakan platform Sevima Edlink sebagai media sistem informasi akademik dan pembelajaran online. Menurut beberapa mahasiswa masih terdapat beberapa kendala yang dihadapi dalam pemahaman, penggunaan dan fungsi aplikasi edlink ini. Tujuan penelitian ini adalah mengukur tingkat kepuasan mahasiswa Fakultas Ilmu Komputer dalam penggunaan Edlink menggunakan algoritma C4.5 dan Naïve Bayes. Untuk mengukur tingkat akurasi dari algoritma C4.5 dan Naïve Bayes dalam rangka mengukur tingkat kepuasan mahasiswa ini indikator yang digunakan adalah model pengujian Servqual yaitu Tangible, Reability, Responsiveness, Assurance, dan Empathy. Berdasarkan tingkat akurasi dari kedua metode. Pada dataset yang digunakan terdapat 91 responden mahasiswa yang telah mengisi kuisioner. Dari data kuisioner tersebut selanjutnya diolah menggunakan kedua metode dan telah dilakukan percobaan sebanyak 9 kali perbandingan dari Data Training dan Data Testing yang berbeda. Secara umum mahasiswa puas dan memahami penggunaan aplikasi edlink. Kepuasan tersebut diuji dengan menggunakan Algoritma C4.5 Decision Tree dan Naïve Bayes Classifier. Berdasarkan perbandingan yang telah dilakukan klasifikasi dengan menggunakan Algoritma C4.5 Decision Tree menghasilkan rata-rata nilai akurasi sebesar 77,78% sedikit lebih akurat jika dibandingkan dengan Naïve Bayes Classifier yang menghasilkan rata-rata nilai akurasi sebesar 71,11%.

**Keywords:** Mengukur kepuasan, Edlink, C4.5, Naïve Bayes, Akurasi

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



### PENDAHULUAN

Di Indonesia konsep teknologi informasi saat ini sudah terjadi peningkatan untuk memenuhi kebutuhan banyak masyarakat di berbagai macam bidang. Termasuk pada bidang pendidikan di perguruan tinggi di Indonesia, perkembangan teknologi informasi memberikan pengaruh yang besar bagi dunia pendidikan Indonesia [1], [2]. Salah

satu bentuk pemanfaatan teknologi informasi yang sudah diterapkan di bidang pendidikan ialah *E-Learning*[3]. Kelebihan yang ditawarkan *E-Learning* tak lain sebab *E-Learning* adalah sebuah media pembelajaran yang didukung oleh teknologi komputer serta jaringan internet sehingga memungkinkan konten pembelajaran bisa diakses tanpa adanya keterbatasan berupa jeda serta waktu

[4], [5]. Ada beberapa jenis aplikasi yang digunakan sebagai E-Learning yaitu Google Classroom, Moodle, Sevima Edlink, Schoology dan yang lainnya.

Salah satu instansi pendidikan yang telah memanfaatkan teknologi dalam pendidikan terutama dalam penggunaan E-Learning adalah Fakultas Ilmu Komputer Universitas Lancang Kuning. Fakultas Ilmu Komputer Universitas Lancang Kuning menggunakan platform Sevima Edlink sebagai media pembelajaran. EdLink adalah produk sistem E-Learning yang dirancang guna membantu mahasiswa serta Dosen dalam kegiatan pembelajaran. EdLink saat ini tersedia dalam platform android, IOS, serta web browser [6].

Permasalahannya di sistem E-Learning Edlink yang digunakan di Fakultas Ilmu Komputer Universitas Lancang Kuning belum adanya pengukuran tingkat kepuasan mahasiswa terhadap sistem E-Learning Edlink yang ada pada saat ini dan terdapat kendala yang dihadapi dalam pemahaman mahasiswa terhadap aplikasi. Pada saat dilakukannya pelatihan ada mahasiswa atau mahasiswi yang tidak bisa hadir karena berhalangan atau sakit, dikarenakan kegiatannya yang dibarengi dengan kegiatan PMB (Pengenalan Mahasiswa Baru) sehingga kurang efektif. maka akan berdampak pada tingkat keberhasilan sebuah sistem pelayanan E-Learning Edlink. Peneliti memberi solusi dengan melakukan suatu pengukuran tingkat kepuasan mahasiswa. Adapun indikator yang digunakan dalam pengukuran tingkat kepuasan terhadap sistem E-Learning Edlink ini adalah model pengujian Servqual yaitu *Tangible* (bukti langsung), *Reability* (keandalan), *Responsiveness* (daya tanggap), *Assurance* (jaminan), dan *Empathy* (empati) [7].

Pada penelitian-penelitian sebelumnya yang berkaitan dengan kepuasan dan penggunaan algoritma C4.5 dan *Naïve Bayes*, dapat dilihat pada analisis kepuasan penumpang pada maskapai penerbangan menggunakan algoritma C4.5 dan *Naïve Bayes*. memperoleh hasil akurasi 81,19% dan 76,24% untuk algoritma *Naïve Bayes* [8]. Hasil yang tidak berbeda didapat dalam perbandingan metode klasifikasi C4.5 dan *Naïve Bayes* untuk mengukur kepuasan pelanggan hasil yang diperoleh adalah C4.5 mempunyai nilai *accuracy* sebesar 94,17%, lebih besar dibandingkan *Naïve Bayes* dengan nilai *accuracy* 85,83% [9]. Namun pada penelitian mengenai perbandingan algoritma C4.5 dan *Naïve Bayes* dalam klasifikasi tingkat kepuasan mahasiswa terhadap pembelajaran daring. Pada hasil perbandingan kedua algoritma bahwa tingkat akurasi dari algoritma *Naïve Bayes* lebih baik dibandingkan dengan C4.5 dengan selisih 11,77%. Nilai akurasi Algoritma *Naïve Bayes* dengan nilai 70,59%. Sedangkan tingkat akurasi dari algoritma

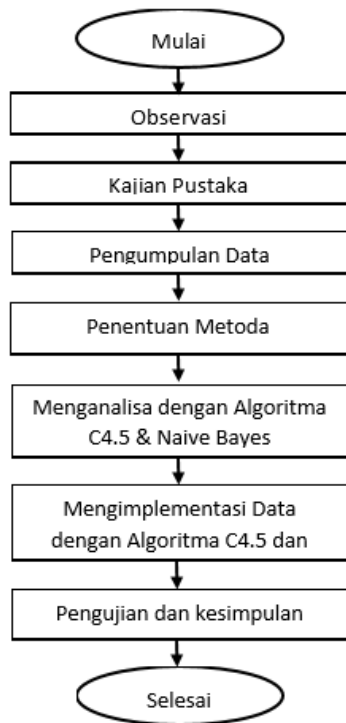
C4.5 adalah 58,82% [10]. Hasil yang tidak berbeda didapat dalam prediksi kepuasan pelayanan publik menggunakan algoritma C4.5 dan *Naïve Bayes*. Dari penelitian yang telah dilakukan, algoritma *Naïve Bayes* dengan pembagian data *training* 70% dan *testing* 30% mampu memberikan hasil prediksi yang lebih baik dibandingkan dengan algoritma C4.5 terbukti dengan nilai akurasi = 96,89%, *Presisi* = 95,50%, *recall*= 95,00% dan *f-measure*= 94,60% [11].

*Data Mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya untuk menemukan, menggali atau menambang mengetahui dari data atau informasi yang sudah ada. *Data Mining*, atau sering dikenal dengan nama *Knowledge Discovery in Database* (KDD). KDD merupakan serangkaian kegiatan seperti pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan pada set data berukuran besar [12]. Analisa sistem perlu dilakukan untuk mengevaluasi sistem yang sedang berjalan dan kebutuhan untuk pengembangan ketahap selanjutnya dengan melibatkan pengguna [13], [14]. Metoda C4.5 juga digunakan untuk klasifikasi data aroma the dengan hidung elektronik dengan akurasi 94,27% [15].

Hal yang melatar belakangi penelitian ini adalah (1) belum pernah dilakukan evaluasi kepada mahasiswa fasilkom unilak untuk mengukur kepuasan penggunaan aplikasi edlink menggunakan kedua algoritma tersebut diatas, (2) karena beberapa mahasiswa melaporkan ada yang belum memahami aplikasi edlink ini maka perlu mengetahui serta mengukur tingkat kepuasan mahasiswa Fakultas Ilmu Komputer Universitas Lancang Kuning dalam penggunaan Edlink sebagai media pembelajaran bagi mahasiswa dan media temu antara dosen dengan mahasiswa secara online. Pengukuran dilakukan dengan menggunakan metode Algoritma C4.5 dan *Naïve Bayes* sebagai alat bantu dalam mengklasifikasi kepuasan mahasiswa terhadap penggunaan Edlink di Fakultas Ilmu Komputer Universitas Lancang Kuning dan mengukur tingkat akurasi dari kedua Algoritma tersebut. Penggunaan metode Algoritma C4.5 dan *Naïve Bayes* pada penelitian ini karena secara umum kedua algoritma ini paling banyak digunakan untuk mengukur tingkat kepuasan pengguna. Hasil penelitian ini akan menghasilkan nilai kepuasan mahasiswa terhadap penggunaan aplikasi edlink ini dan menjadi masukan bagi pengelola Edlink di Universitas Lancang Kuning untuk pengembangan selanjutnya aplikasi edlink. Pada akhirnya penelitian ini tidak hanya mengukur tingkat kepuasan mahasiswa saja, akan tetapi juga membandingkan hasil akurasi dengan dua algoritma yang berbeda.

## METODE PENELITIAN

Dalam penyusunan penelitian ini terdapat tahapan-tahapan yang dilakukan dalam sebuah penelitian dapat dilihat pada gambar berikut :



Gambar 1. Alur Metodologi Penelitian

### Pengumpulan Data

Teknik pengumpulan data :

2.1.1 Observasi yaitu teknik pengumpulan data dengan pengamatan langsung ke tempat penelitian untuk memperoleh data yang dibutuhkan dalam melakukan penelitian.

2.1.2 Studi literatur adalah teknik dalam pengumpulan data dengan mengumpulkan data melalui media digital, dokumen cetak, jurnal, artikel ilmiah atau sumber lainnya yang berkaitan dengan judul dan permasalahan yang akan diteliti.

2.1.3 Kuesioner adalah teknik pada pengumpulan data dengan memberikan pertanyaan-pertanyaan yang akan dijawab oleh responden penelitian. Pada penelitian ini kuesioner dibuat dengan menggunakan *Google Form* kemudian link pengisian kuesioner dibagikan kepada responden yaitu mahasiswa Fakultas Ilmu Komputer. Kuesioner ini berisikan mengenai seputar pertanyaan yang berhubungan dengan penelitian ini.

### Data Mining

Data mining ialah proses mengekstraksi dan mengidentifikasi informasi yang bermanfaat serta pengetahuan yang terkait dari berbagai *database* besar

dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* [12].

*Knowledge Discovery in Databases* (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD [16].

### Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma untuk mengubah fakta yang besar menjadi pohon keputusan (*decision tree*) yang merepresentasikan aturan (*rule*) [17]. Tujuan dari pembentukan pohon keputusan dalam algoritma C4.5 adalah untuk mempermudah dalam penyelesaian permasalahan [18] [19] [20].

Tahapan dalam penggunaan pemodelan Algoritma C4.5 adalah sebagai berikut :

- Menyiapkan data *training*, data ini didapat dari hasil kuesioner yang didapat menggunakan metode Skala Likert.
- Menentukan akar dari pohon keputusan yang akan dibangun dengan menghitung nilai *gain* yang paling tinggi dari masing masing atribut berdasarkan nilai index *entropy* paling rendah. Berikut rumus untuk menghitung atau mencari nilai index *entropy* :

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

Keterangan:

S : Mahasiswa pengguna Edlink

n : Jumlah respondent pengguna Edlink

$p_i$  : Proporsi dari  $S_i$  terhadap S

- Selanjutnya menghitung nilai *Gain*, berikut rumus untuk menghitung nilai *Gain* :

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

Keterangan:

S : Mahasiswa pengguna Edlink

A : Atribut butir pertanyaan

n : Jumlah partisi atribut A

$|S_i|$  : Jumlah kasus pada partisi ke i

$|S|$  : Jumlah kasus dalam S

- Langkah kedua harus terus diulangi hingga setiap cabang terpenuhi.
- Proses partisipasi *decision tree* berhenti saat semua cabang dalam node N berada pada *class* yang sama, tidak ada atribut dalam cabang yg bisa dibagi lagi serta tidak ada cabang yg kosong.
- Informasi yang dihasilkan terbagi menjadi dua kelompok yaitu Puas dan Tidak Puas.

### Algoritma Naïve Bayes

*Naïve Bayes Classifier* adalah metode statistik bayesian sederhana [21]. Disebut *Naïve* karena diasumsikan semua variabel berkontribusi terhadap klasifikasi dan saling berkorelasi, asumsi ini disebut *class conditional independence*. Ini didasarkan dengan teorema bayes tentang kemungkinan kondisional/*conditional probabilities* [22] [8] [23]:

Rumus algoritma *Naïve Bayes* didefinisikan sebagai berikut :

$$P(C_i|X) = \frac{P(X|C_i) P(C_i)}{P(X)} \quad (3)$$

Keterangan :

$P(C_i|X)$  : peluang tingkat kepuasan mahasiswa (posterior).

$P(C_i)$  : peluang puas atau tidak (priori propability)

$P(X|C_i)$  : sebagai kemungkinan salah satu kelas

$P(X)$  : adalah propabilitas prediktor kelas sebelumnya

### Teknik Pengambilan Sampel

Pada teknik pengambilan sampel untuk penelitian ini menggunakan teknik *probability sampling*. Teknik *probability sampling* adalah teknik pengambilan sampel yang memberikan peluang yang sama bagi setiap anggota populasi untuk dipilih menjadi anggota sampel.

#### a. Populasi

Pada penelitian ini yang menjadi populasi untuk penelitian adalah mahasiswa aktif Fakultas Ilmu Komputer Universitas Lancang Kuning dari angkatan 2019 - 2022 yang terdiri dari tiga prodi yaitu, Teknik Informatika, Sistem Informasi dan Bisnis Digital yang berjumlah 991 mahasiswa.

#### b. Sampel

Sampel merupakan sebagian dari populasi anggota populasi yang diambil dengan cara menggunakan teknik sampling. Pada penelitian ini menggunakan teknik *simple random sampling* untuk responden mahasiswa di Fakultas Ilmu Komputer Universitas Lancang Kuning sebanyak 991 populasi. Berikut langkah untuk pengambilan sampel menggunakan *simple random sampling*. Menggunakan rumus slovin (Usman, 2007) dalam [24] :

$$n = \frac{N}{1 + Ne^2} \quad (4)$$

Keterangan :

$n$  : Ukuran sampel

$N$  : Jumlah populasi

$E$  : Batas toleransi kesalahan (Slovin dengan *error* 10%)

Perhitungan rumus :

$$n = \frac{991}{1 + 991 (0,1)^2}$$

$$n = \frac{991}{1 + 991 \times 0,01}$$

$$n = \frac{991}{10,91}$$

$$n = 90,83$$

Dari hasil perhitungan rumus di atas maka total sampel yang di dapat dalam penelitian ini adalah 90,83 dan dibulatkan menjadi 91 responden Mahasiswa Fakultas Ilmu Komputer Universitas Lancang Kuning.

## HASIL DAN PEMBAHASAN

Analisis data pada penelitian ini menggunakan Algoritma C4.5 *Decision Tree* dan *Naïve Bayes Classifier*, dengan menggunakan data yang sebelumnya telah di tarik melalui kuesioner yang peneliti sebarakan hingga tanggal 16 Januari 2023 kepada mahasiswa Fasilkom Unilak dari angkatan 2019 hingga 2022. Pada penelitian ini juga dilakukan perhitungan langsung menggunakan bantuan *Tools Google Colaboratory* dengan bahasa pemrograman *Python*.

### Pengumpulan Data

Data yang dikumpulkan oleh peneliti secara langsung dengan cara menyebarkan kuesioner kepada mahasiswa Fasilkom Unilak dari angkatan 2019 hingga 2022 dengan bantuan media *Google Form*. Kuesioner berisikan pertanyaan mengenai tingkat kepuasan mahasiswa dalam menggunakan aplikasi Edlink Sevima. Sebanyak 91 Responden yang telah mengisikan form kuesioner tersebut telah dilakukan perekapan.

### Praproses Data

Pada praproses data ini dilakukan dengan mengambil data mentah dari Tabel 5.1 lalu direkap dalam bentuk data yang tersimpan dengan format *Excel (.xls)*, yang nantinya akan dijadikan sebuah data baru yang terdiri dari kode Responden serta Jawaban dari masing-masing pertanyaan variabel Model Servqual yang sudah di konversikan menjadi angka.

### Pembersihan Data

Pembersihan data adalah tahap yang berfungsi untuk menghilangkan data duplikat, tidak konsisten, *missing values*, dan data yang bersifat *outlier* dari data mentah yang akan diklasifikasikan. Beberapa langkah yang dilakukan untuk proses pembersihan data adalah dengan cara menghilangkan *missing values* dan menghapus data yang tidak diperlukan. Dalam hal ini pembersihan tidak dilakukan karena data yang diberikan dari tempat penelitian tidak terdapat *missing values* dan untuk data identitas seperti nama nim dan prodi perlu dihapuskan dan dibuat menjadi kode Responden. Untuk kategori jawaban dikonversikan menjadi angka sesuai dengan keterangan berikut :

Jawaban "Sangat Puas" bernilai 5

Jawaban "Puas" bernilai 4

Jawaban "Cukup Puas" bernilai 3

Jawaban "Kurang Puas" bernilai 2

Jawaban "(*missing values*)" 1

### Transformasi Data

Dari data yang sudah melewati tahapan Pembersihan data selanjutnya dilakukan perhitungan dan membuat sebuah sheet baru pada MS.Excel dengan isian data yang berisikan data Kode Responden, Jawaban Kusioner dari 5 Variabel Model Servqual : *Tangible* (Var.A) dengan pertanyaan Q1 hingga Q4, *Reliability*



(Var.B) dengan pertanyaan Q5 hingga Q7, *Responsiveness* (Var.C) dengan pertanyaan Q8 hingga Q11, *Assurance* (Var.D) dengan pertanyaan Q12 hingga Q13 dan *Emphaty* (Var.E) dengan pertanyaan Q14 hingga Q15 serta Label dengan pertanyaan Q16 dengan total sebanyak 91 Responden. Untuk pelabelan pada pertanyaan Q16 jika jawaban bernilai 4 dan 5 maka masuk ke kategori label “Puas” sedangkan jawaban selain itu dengan kategori label “Tidak Puas”. Selanjutnya data disimpan untuk digunakan pada perhitungan pada *Tools Google Colaboratory*.

#### Data Hasil Praproses

Adapun hasil praproses data dapat diuraikan pada tabel berikut :

**Tabel 1.** Data Olah

RES	A1	A2	A3	A4	B1	B2	B3	C1	C2	C3	C4	D1	D2	E1	E2	LABEL
R001	4	5	4	4	5	5	4	4	4	4	5	4	4	5	4	Puas
R002	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	Puas
R003	3	3	3	3	4	4	3	4	4	3	4	5	5	4	5	Puas
R004	3	3	4	3	3	3	3	3	3	3	3	3	3	3	3	Tidak Puas
R005	3	3	3	3	3	3	4	3	3	3	3	3	3	3	3	Tidak Puas
R006	3	3	4	4	5	5	4	3	3	3	3	4	4	3	3	Puas
R007	3	5	5	5	5	5	5	3	3	4	5	5	3	3	3	Puas
R008	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	Puas
R009	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	Puas
R010	4	3	4	3	4	4	4	4	4	4	4	4	4	4	4	Puas

**Tabel 2.** Data Olah

Kode	Keterangan
Kode R-n	Kode Responden
A	Pertanyaan Variabel <i>Tangible</i>
B	Pertanyaan Variabel <i>Reliability</i>
C	Pertanyaan Variabel <i>Responsiveness</i>
D	Pertanyaan Variabel <i>Assurance</i>
E	Pertanyaan Variabel <i>Emphaty</i>
Label	Kategori Label

#### Implementasi dan Analisis Hasil

Setelah melakukan persiapan data selanjutnya melakukan pengujian dengan *Tools Google Colaboratory* untuk melihat hasil klasifikasi tingkat kepuasan mahasiswa Fakultas Ilmu Komputer Universitas Lancang Kuning dalam penggunaan Aplikasi Edlink Sevima dengan metode Algoritma C4.5 Decision Tree dan Naïve Bayes Classifier. Adapun tahapannya adalah sebagai berikut :

#### Menyiapkan Data

Data yang disiapkan merupakan Data Olah yang terdapat pada Tabel 5.2 yang disimpan dalam format .xlsx. data ini nantinya digunakan sebagai bahan perhitungan analisis yang diunggah ke halaman *Google Colaboratory*

#### Pengujian Dengan Tools Google Colaboratory

- Unggah Data Ke *Google Drive*
- Mengimport *Library*  
 Pada implementasi *Google Colaboratory* memiliki kelas yaitu kelas proses. Selain itu,

perangkat lunak ini menggunakan *library python*

- Memanggil Data Dari *Google Drive*  
 Pada tahap ini melakukan pemanggilan data yang sudah diunggah ke *Google Drive* sebelumnya
- Membagi *Data Training* dan *Data Testing* Dataset, selanjutnya dilakukan proses pembagian *data training* dan *data testing*, dalam hal ini peneliti mencoba menggunakan data responden R001 hingga R081 sebagai *data training*, sementara *data testing* yang digunakan adalah data responden R082 hingga R091.

#### Klasifikasi Dengan Algoritma C4.5 Decision Tree

Adapun tahapannya adalah sebagai berikut :  
 Memproses data training dengan Algoritma C4.5 Decision Tree

Pada tahap ini analisis dilakukan dengan melakukan perhitungan dengan menggunakan metode Algoritma C4.5 *Decision Tree*.

Membuat Visualisasi *Decision Tree*

Selanjutnya menjalankan *cell code* dan menghasilkan visualisasi pohon keputusan

#### Hasil Klasifikasi Berdasarkan Rules *Decision Tree*

Pada tahap ini *data testing* dapat diklasifikasikan berdasarkan rules atau aturan yang telah dihasilkan dari *Decision Tree* sebelumnya. Selanjutnya menjalankan *cell code* dan menghasilkan klasifikasi dan tingkat akurasi dari Algoritma C4.5 *Decision Tree*

Pada hasil klasifikasi diatas, untuk *data testing* dengan label “0” menyatakan “PUAS”, sedangkan label dengan “1” menyatakan “TIDAK PUAS”. Baris pertama adalah data label aktual atau data label yang sebenarnya, sementara baris kedua merupakan label hasil prediksi. Dapat diketahui pada baris ketiga akurasi dari Algoritma C4.5 *Decision Tree* dengan menggunakan 10 *data testing* sebesar 70%.

#### Klasifikasi Dengan Naïve Bayes Classifier

Adapun tahapannya adalah sebagai berikut  
 Memproses *data training* dengan *Naïve Bayes Classifier*.

Pada tahap ini analisis dilakukan dengan melakukan perhitungan dengan menggunakan metode *Naïve Bayes Classifier*.

Hasil Klasifikasi dengan *Naïve Bayes Classifier*.

Pada tahap ini *data testing* dapat diklasifikasikan berdasarkan dari *data training* yang telah diolah menggunakan *Naïve Bayes Classifier* sebelumnya. Selanjutnya menjalankan *cell code* dan menghasilkan klasifikasi dan tingkat akurasi dari *Naïve Bayes Classifier*

#### Perbandingan Metode Klasifikasi

Berdasarkan tingkat akurasi dari kedua metode diatas, didapat bahwa tingkat akurasi metode Klasifikasi Algoritma C.45 *Decision Tree* sebesar 70% lebih besar dari Metode Klasifikasi *Naïve Bayes Classifier* yaitu sebesar 50%, artinya metode Klasifikasi Algoritma C.45 *Decision Tree* lebih baik digunakan dalam mencari pengklasifikasian data dengan menggunakan 10 *data*

testing responden pada analisis ini. Adapun tampilan klasifikasinya dapat dilihat pada tabel 3.

Tabel 3. Perbandingan Hasil Klasifikasi Kedua Metode

Resp	Label Aktual (Sebenarnya)	Hasil Klasifikasi	
		ALGO C4.5 DT	NBC
R082	Tidak Puas	Puas	Tidak Puas
R083	Puas	Puas	Tidak Puas
R084	Puas	Tidak Puas	Tidak Puas
R085	Puas	Puas	Puas
R086	Tidak Puas	Puas	Puas
R087	Tidak Puas	Tidak Puas	Tidak Puas
R088	Tidak Puas	Tidak Puas	Puas
R089	Puas	Puas	Tidak Puas
R090	Puas	Puas	Puas
R091	Puas	Puas	Puas

Selanjutnya jika dibandingkan dengan *Data Training* dan *Data Testing* yang berbeda dan dilakukan percobaan sebanyak 9 kali maka menghasilkan tingkat akurasi dengan total rata-rata seperti tabel 4.

Tabel 4. Perbandingan Tingkat Akurasi Dengan Dataset Kedua Metode Yang Berbeda

Perc Ke-	Dataset (Responden Ke-)		Tingkat Akurasi (%)	
	Data Training	Data Testing	Algo C4.5 DT	NBC
1	1-81	82-91	70%	50%
2	1-71, 82-91	72-81	60%	40%
3	1-61, 72-91	62-71	80%	70%
4	1-51, 62-91	52-61	70%	70%
5	1-41, 52-91	42-51	70%	70%
6	1-31, 42-91	32-41	80%	90%
7	1-21, 32-91	22-31	90%	80%
8	1-11, 22-91	12-21	90%	80%
9	1, 12-91	2-11	90%	90%
<b>Rata-rata :</b>			<b>77,78%</b>	<b>71,11%</b>

Pada tabel 4 diatas dapat dilihat perbedaan yang tampak bahwasannya klasifikasi dengan menggunakan Algoritma C4.5 *Decision Tree* menghasilkan rata-rata sebesar 77,78% sedikit lebih akurat jika dibandingkan dengan *Naive Bayes Classifier* yang menghasilkan rata-rata sebesar 71,11% dengan total sebanyak 9 kali percobaan dengan Data Training dan Data Testing yang berbeda.

## SIMPULAN

Pada dataset yang digunakan terdapat 91 responden mahasiswa yang telah mengisikan kuisioner. Responden 01 - 81 digunakan sebagai data *training*, Responden 82 - 91 digunakan sebagai data *testing*, berdasarkan data tersebut selanjutnya di olah menggunakan kedua metode dan menghasilkan nilai akurasi sebesar 70 % pada Algoritma C4.5 dan 50 % pada algoritma *Naive Bayes Classifier*, artinya metode Klasifikasi Algoritma C.45 *Decision Tree* lebih baik digunakan dalam mencari pengklasifikasian data dengan menggunakan 10 *data testing* responden pada analisis ini.

Pada perbandingan dari *Data Training* dan *Data Testing* yang berbeda, telah dilakukan percobaan sebanyak 9 kali dengan menggunakan kedua metode, kemudian data tersebut selanjutnya di olah. Berdasarkan perbandingan yang telah dilakukan klasifikasi dengan menggunakan Algoritma C4.5 *Decision Tree* menghasilkan rata-rata nilai akurasi sebesar 77,78% sedikit lebih akurat jika dibandingkan dengan *Naive Bayes Classifier* yang menghasilkan rata-rata nilai akurasi sebesar 71,11%.

Pada penelitian yang dilakukan telah berhasil melakukan analisa klasifikasi pengukuran tingkat kepuasan mahasiswa dalam penggunaan Edlink dengan menggunakan Algoritma C4.5 *Decision Tree* dan *Naive Bayes Classifier* pengujian dilakukan dilakukan dengan *tools Google Colaboratory* dengan bahasa pemograman *python*.

## DAFTAR PUSTAKA

- [1] S. Djujar, T. Taslim, and D. Toresa, "Pengaruh Kemudahan, Kebutuhan, Kemampuan, Minat terhadap Persepsi Guru dalam Menggunakan Aplikasi Google," *ETHOS J. Penelit. dan Pengabd. Kpd. Masy.*, vol. 8, no. 2, pp. 242–247, 2020, doi: 10.29313/ethos.v8i2.5649.
- [2] S. Sains, D. Toresa, S. Handayani, and R. Muzawi, "Digitalisasi Pengelolaan Pustaka Sekolah," 2023, doi: 10.33372/stn.v9i1.989.
- [3] D. Toresa, Taslim, Susi Handayani, and Musfawati, "Pelatihan Membuat Website Dan Jaringan Lan Komputer Bagi Siswa Sma Negeri 4 Tualang Kab. Siak," *J-COSCIS J. Comput. Sci. Community Serv.*, vol. 2, no. 2, pp. 37–42, 2022, doi: 10.31849/jcscis.v2i2.9198.
- [4] A. R. Dikananda, H. B. Santoso, R. D. Dana, and D. Sudrajat, "E-Learning Usability Evaluation Menggunakan Fuzzy Logic dan Usulan Alternatif Desain Interaktif Learning Management System (LMS) Chamilo," *J. ICT Inf. Commun. Technol.*, vol. 18, no. 1, pp. 65–70, 2019, doi: 10.36054/jict-ikmi.v18i1.56.
- [5] R. A. Indris, L. Lhaura, V. Fc, M. Kom, and D. Toresa, "Analisa Kepuasan Mahasiswa Terhadap Blended Learning Di Universitas Lancang Kuning," *SEMASTER Semin. Nas. Teknol. Inf. Ilmu Komput.*, vol. 1, no. 1, pp. 39–52, 2020.

- [6] A. A. Putro, A. Ambarwati, and E. Setiawan, "Analisa Manajemen Risiko E-Learning Edlink Menggunakan Metode NIST SP 800-30 Revisi 1," *J. Teknol. dan Inf.*, vol. 11, no. 2, pp. 125–136, 2021, doi: 10.34010/jati.v11i2.5314.
- [7] Gustientiedina, M. Siddik, and Y. Desnelita, "Penerapan Naïve Bayes untuk Memprediksi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademis," *J. Infomedia*, 2019.
- [8] Y. Hendra Kusuma, S. Suprapto, and Y. Setiawan, "SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Analysis of Passenger Satisfaction on Airlines Using C4.5 and Naïve Algorithm Analisis Kepuasan Penumpang pada Maskapai Penerbangan Menggunakan Algoritma C4.5 dan Naïve Bayes." pp. 162–171, 2022.
- [9] Yunita and I. H. Ikasari, "Perbandingan Metode Klasifikasi C4.5 dan Naïve Bayes untuk Mengukur Kepuasan Pelanggan," *Inform. Univ. Pamulang*, vol. 6, no. 3, pp. 2622–4615, 2021.
- [10] F. Narti, "JTIM: Jurnal Teknologi Informasi dan Multimedia Perbandingan Algoritma C4.5 dan Naive Bayes dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring," vol. 4, no. 1. 2022.
- [11] Y. Umaidah and U. Enri, "Prediction of Public Service Satisfaction Using C4.5 and Naïve Bayes Algorithm," *J. PILAR Nusa Mandiri*, vol. 17, no. 2, pp. 143–148, 2021.
- [12] Z. Gustiana, "Penerapan algoritma c 4.5 dalam sistem pendukung keputusan evaluasi kinerja fasilitator pamsimas (studi kasus di kabupaten kampar)," *J. Inf. Technol. Res.*, vol. 1, no. 1, pp. 20–28, 2020.
- [13] M. Jufri, "Analisis Faktor-Faktor Yang Mempengaruhi Tingkat," vol. 4, no. December 2020, pp. 107–115, 2021, doi: 10.52062/jakd.v14i2.1457.
- [14] G. Gushelmi, M. Neldi, and Y. Septiadi, "Analisa Kualitas Sistem Informasi Manajemen Menggunakan Framework Cobit 5 (Studi Kasus Pada Kantor Dprd Kabupaten Sijunjung)," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 4, no. 1, pp. 89–96, 2022, doi: 10.47233/jteksis.v4i1.376.
- [15] Darussalam and G. Arief, "Jurnal Resti," *Resti*, vol. 1, no. 1, pp. 19–25, 2017.
- [16] F. Riandari and A. Simangunsong, *PENERAPAN ALGORITMA C4.5 UNTUK MENGUKUR TINGKAT KEPUASAN MAHASISWA*. 2020.
- [17] V. No *et al.*, "Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer," vol. 5, no. 2, pp. 73–83, 2023.
- [18] S. Faisal, "Klasifikasi Data Mining Menggunakan Algoritma C4.5 Terhadap Kepuasan Pelanggan Sewa Kamera Cikarang," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2019, doi: 10.36805/technoxplore.v4i1.541.
- [19] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
- [20] "Increasing Accuracy of C4.5 Algorithm Using Information Gain Ratio and Adaboost for Classification of Chronic Kidney Disease," *Journal of Soft Computing Exploration*, vol. 1, no. 1. 2020, doi: 10.52465/josce.v1i1.6.
- [21] W. Ningsih, B. Budiman, and I. Umami, "Implementasi Algoritma Naïve Bayes Untuk Menentukan Calon Penerima Beasiswa Di SMK YPM 14 Sumobito Jombang," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 4, no. 2, pp. 446–454, 2022, doi: 10.47233/jteksis.v4i2.570.
- [22] D. Marutho, "Perbandingan Metode Naive Bayes, KNN, Decision Tree Pada Laporan Water Level Jakarta," *Manaj. Inform. AMIK JTC Semarang*, vol. 15, no. 2, pp. 90–97, 2019.
- [23] K. Khoirunnisa, L. Susanti, I. T. Rokhmah, and L. Stianingsih, "Prediksi Siswa Smk Al-Hidayah Yang Masuk Perguruan Tinggi Dengan Metode Klasifikasi," *J. Inform.*, vol. 8, no. 1, pp. 26–33, 2021, doi: 10.31294/ji.v8i1.9163.
- [24] K. Adi Kurniawan Saputra, P. Dian Pradnyanitasari, and N. I. Made Intan Priandani dan Gst B Ngr P Putra, "Praktek Akuntabilitas Dan Kompetensi Sumber Daya Manusia Untuk Pencegahan Fraud Dalam Pengelolaan Dana Desa," *J. KRISNA Kumpul. Ris. Akunt.*, vol. 10, no. 2, pp. 168–176, 2019.