

## PERBANDINGAN *NAIVE BAYES* DAN C45 DALAM KLASIFIKASI TES KESEHATAN MAHASISWA BARU AKBID AS-SYIFA

Feby Wulandari Sembiring<sup>1</sup>, Wiwin Handoko<sup>2\*</sup>,  
Firdha Agis Utami Batu Bara<sup>3</sup>, Sulaseh<sup>4</sup>  
<sup>1,2,3,4</sup>Sistem Informasi, STMIK Royal Kisaran  
\*email: win.van.handoko@gmail.com

**Abstract:** Medical test is important to determine the health of a person's body so they are often carried out and have even become one of the requirements for an institution to accept new members such as the As-Syifa Midwifery Academy which conducts medical tests for prospective new students. The problem is that so far the classification system for recapitulation of medical test results is still done manually, thus slowing the performance of the campus and it is feared that there will be damage to the data format. The problem solving technique above is carried out with a data mining process using Naïve Bayes and Decision Tree C45 where the two algorithms are compared to find the 1 best classification algorithm to be implemented in the system. The dataset uses data on the recapitulation of the results of the 2018 new student health tests sourced from the Administration (TU) of Akbid As-Syifa. The comparison uses 4 data testing models and the confusion matrix as the performance evaluation value of the modeling algorithm. The modeling results obtained that the Decision Tree C45 algorithm is superior and suitable to be implemented with an accuracy rate of 100% while Naïve Bayes has a maximum accuracy rate of 96%. The purpose of this study was to help Akbid As-Syifa classify the results of the health test of prospective new students.

**Keywords:** Medical test; Prospective new students; Naïve Bayes; C45.

**Abstrak:** Tes kesehatan merupakan hal penting untuk mengetahui kesehatan tubuh seseorang sehingga banyak dilakukan bahkan telah menjadi salah satu persyaratan bagi sebuah lembaga untuk menerima anggota baru seperti halnya Akademi Kebidanan (Akbid) As-syifa yang melakukan tes kesehatan bagi calon mahasiswa baru. Permasalahannya selama ini sistem klasifikasi rekapitulasi hasil tes kesehatan masih dilakukan secara manual sehingga memperlambat kinerja pihak kampus serta dikhawatirkan adanya kerusakan format data. Adapun teknik penyelesaian masalah diatas dilakukan dengan proses data *mining* menggunakan *Naïve Bayes* dan *Decision Tree* C45 dimana kedua algoritma tersebut dibandingkan untuk mencari 1 algoritma klasifikasi terbaik untuk diimplementasikan dalam sistem. Dataset menggunakan data rekapitulasi hasil tes kesehatan calon mahasiswa baru tahun 2018 bersumber dari Tata Usaha (TU) Akbid As-syifa. Perbandingan menggunakan 4 model pengujian data dan *confusion matrix* sebagai nilai evaluasi performa dari pemodelan algoritma tersebut. Hasil pemodelan diperoleh algoritma *Decision Tree* C45 lebih unggul dan cocok untuk diimplementasikan dengan tingkat akurasi 100% sedangkan *Naïve Bayes* tingkat akurasinya maksimal sebesar 96%. Tujuan penelitian ini adalah membantu pihak Akbid As-syifa mengklasifikasikan hasil tes kesehatan calon mahasiswa baru.

**Kata Kunci:** Tes kesehatan; Calon Mahasiswa Baru; Naïve Bayes; C45.

## **PENDAHULUAN**

Kesehatan merupakan sesuatu alamiah yang harus dimiliki oleh setiap makhluk hidup. Kesehatan merupakan sebuah keadaan yang baik secara mental, fisik, spiritual maupun sosial yang memungkinkan setiap orang untuk hidup produktif secara sosial, ekonomi maupun pendidikan [1]. Pada dunia pendidikan kesehatan menjadi hal yang vital karena dengan kondisi tubuh yang sehat seseorang mampu untuk mengikuti pelajaran dan dapat menerima ilmu yang diberikan dengan baik terlebih lagi pada lingkungan kampus dikarenakan sistem pembelajaran dan materi yang diberikan dua kali lebih berat daripada masa sekolah. Sehingga saat ini bukan hal yang baru apabila lembaga pendidikan ingin mengetahui kondisi kesehatan peserta didiknya sehingga diadakan tes kesehatan untuk penerimaan mahasiswa baru khususnya kampus – kampus berbasis kesehatan.

Tes kesehatan pada calon mahasiswa baru yang akan memasuki sebuah lingkungan pendidikan bertujuan untuk mengetahui kondisi tubuh calon mahasiswa tersebut apakah riwayat kesehatannya baik atau buruk sebelum diresmikan mahasiswa di lingkungan pendidikan tersebut [2]. Calon mahasiswa baru merupakan orang yang akan mendapat hak pendidikan ataupun kemampuan di suatu perguruan tinggi dengan baik berdasarkan syarat-syarat yang telah ditentukan dari suatu akademik atau perguruan tinggi [3]. Tes kesehatan dilakukan dengan mengukur sejumlah variabel yang meliputi kondisi fisik dari calon mahasiswa tersebut. Hasil tes kesehatan pada umumnya menghasilkan lulus/layak maupun tidak lulus/tidak layak dari orang yang sedang di tes kesehatannya. Salah satu kampus yang menerapkan tes kesehatan pada pemilihan calon mahasiswa baru di kabupaten Asahan adalah Akademi Kebidanan (Akbid) As-syifa Kisaran. Hal tersebut bertujuan agar mahasiswa dengan jurusan kesehatan tersebut memiliki kondisi yang sehat dan tanpa cacat karena dikhawatirkan apabila memiliki kondisi kesehatan buruk dan terdapat cacat akan membahayakan yaitu terjadi kesalahan ketika mereka melakukan pemeriksaan terhadap pasien sementara pada dunia kesehatan pemeriksaan terhadap pasien wajib dilakukan dengan baik dan benar serta ketelitian yang kuat [4]. Itulah sebabnya hasil tes kesehatan menjadi salah satu faktor penentu selain akademik bagi calon mahasiswa untuk diterima di kampus Akbid As-syifa tersebut. Oleh karena itu pihak kampus Akbid As-syifa Kisaran menyelenggarakan tes kesehatan pada setiap calon mahasiswanya yang berjumlah puluhan bahkan ratusan setiap tahunnya sehingga menghasilkan data rekapitulasi kesehatan mahasiswa baru yang jumlahnya sangat banyak.

Namun permasalahannya adalah selama ini dengan jumlah data yang sangat banyak namun rekapitulasi data hasil kesehatan calon mahasiswa baru masih diproses dan disimpan secara manual. Proses yang masih manual tersebut menyebabkan kesulitan untuk pihak kampus dalam mengklasifikasikan hasil tes kesehatan mahasiswa baru sehingga membutuhkan waktu yang cukup lama. Selain itu, dikarenakan belum adanya sistem yang digunakan untuk rekapitulasi data hasil rekapitulasi kesehatan calon mahasiswa baru tersebut dikhawatirkan data – data penting hilang dan dokumennya rusak sehingga dinilai bahwa hal demikian akan memperlambat proses kinerja pihak kampus untuk segera melihat hasil klasifikasi dari rekapitulasi calon mahasiswa yang lulus tes kesehatan serta belum adanya metode perhitungan yang digunakan untuk melakukan rekapitulasi klasifikasi data hasil tes kesehatan calon mahasiswa baru

tersebut dinilai kurang tepat karena dikhawatirkan hasil pengklasifikasiannya tidak tepat sasaran.

Berdasarkan permasalahan tersebut, pihak kampus Akbid As-syifa membutuhkan sebuah sistem yang dapat melakukan klasifikasi hasil tes kesehatan pada calon mahasiswa baru secara efisien. Permasalahan mengenai pengolahan data tersebut yang masih dilakukan secara manual dengan jumlah data yang banyak memerlukan penanganan berbasis *machine learning* dengan penerapan data *mining* menggunakan metode klasifikasi dengan Algoritma *Naïve Bayes Classifier* dan *Decision Tree C45 Classifier*. Penggunaan kedua metode tersebut karena metode tersebut memiliki tingkat akurasi yang cukup baik. Penerapan Algoritma *Naive Bayes* dalam studi kasus klasifikasi penentuan sebuah kelulusan menghasilkan tingkat akurasi diatas sebesar 90% seperti dalam penelitian [5][6]. Disamping itu, dalam pengklasifikasian algoritma *Decision Tree C45* juga memiliki tingkat akurasi diatas 90% juga seperti pada penelitian [7][8].

Penerapan algoritma *Naïve Bayes* dan C45 dalam tes kesehatan sudah sering dilakukan. Pada penelitian yang dilakukan [9] dengan judul “Rancang Bangun Sistem Pakar Jenis Penyakit Pada Hasil Tes Darah” menggunakan algoritma *Naïve Bayes Classifier* untuk menghitung hasil tes darah agar memperoleh hasil akurasi yang lebih akurat. Dalam penelitian tersebut menghasilkan tingkat akurasi sebesar 100%. Kemudian hasil dari perhitungan algoritma tersebut diimplementasikan ke dalam rancangan untuk direalisasikan penggunaannya. Klasifikasi Algoritma C.45 juga telah diimplementasikan dalam penelitian [10] yang berjudul “Simulasi Penerapan Metode *Decision Tree (C45)* Pada Penentuan Status Gizi Balita” dimana dalam penelitian tersebut algoritma C45 digunakan penentuan status gizi balita dengan *software RapidMiner* dari hasil perhitungan tersebut diperoleh nilai akurasi sebesar 100% sehingga penelitian tersebut dikategorikan sebagai *Good Classification*.

Berdasarkan pemaparan diatas, dikarenakan algoritma *naïve bayes* dan C45 sama – sama memiliki tingkat akurasi yang bagus untuk mengklasifikasikan sebuah tujuan dari permasalahan yang ada pada dunia tes kesehatan sehingga perlu dilakukan perbandingan algoritma antara *Naïve Bayes* dan C45. Perbandingan metode tersebut bertujuan untuk mengetahui algoritma yang paling manakah yang paling baik dengan memiliki tingkat akurasi paling tinggi. Sehingga pada penelitian ini, peneliti akan menerapkan proses *data mining* dengan membandingkan kedua algoritma tersebut untuk mencari algoritma yang paling baik agar diimplementasikan kedalam sistem klasifikasi rekapitulasi data tes kesehatan calon mahasiswa baru Akbid As-syifa Kisaran berbasis *machine learning*. Kedua algoritma tersebut akan diuji pada 4 model pengujian data yaitu pada rasio 60:40, 70:30, 80:20 dan 90:10. Selanjutnya untuk mengukur tingkat performa dari masing – masing algoritma menggunakan nilai evaluasi *confusion matrix* sehingga hasilnya benar – benar akurat. Output dari penelitian ini diharapkan dapat mengatasi permasalahan berkaitan dengan klasifikasi rekapitulasi data tes kesehatan calon mahasiswa baru sehingga proses rekapitulasi data tes kesehatan tersebut dapat dilakukan lebih mudah, efektif dan efisien serta dengan hasil yang akurat.

**METODE**

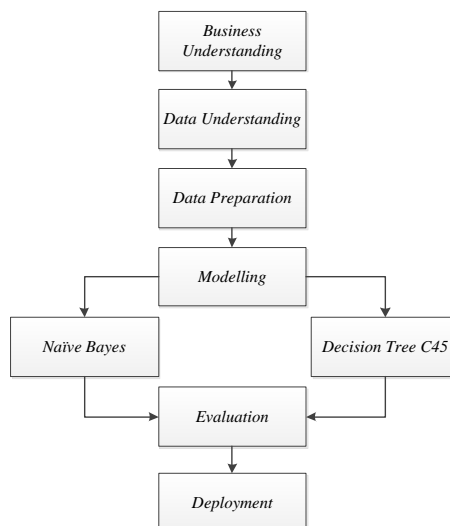
Metode yang digunakan dalam penelitian ini menggunakan penerapan konsep *data mining*. *Data mining* merupakan sebuah proses mengolah satu atau lebih teknik pembelajaran (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan secara otomatis terpaut berbagai sumber data besar [11]. *Data mining* proses pengetahuan untuk menemukan pola dan teknik statistik matematika, kecerdasan buatan untuk mengidentifikasi informasi pengetahuan dari database [12].

Tahapan penelitian ini berpedoman pada *Cross-Industry Standard Process* (CRISP-DM) dalam *data mining*. CRISP-DM adalah urutan proses *data mining* untuk memecahkan sebuah permasalahan yang bersifat universal dari unit penelitian ataupun bisnis [13]. Berikut ini merupakan tahapan – tahapan yang terdapat dalam CRISP-DM yakni :



Gambar 1. Tahapan CRISP-DM

Kemudian tahapan dalam membandingkan 2 algoritma klasifikasi divisualisasikan sebagai berikut:



Gambar 2. Tahapan Perbandingan Algoritma

***Bussiness Understanding***

Hasil tes kesehatan bagi calon mahasiswa baru merupakan salah satu faktor penentu kelulusan untuk calon mahasiswa baru di Akademi Kesehatan (Akbid) As-syifa

Kisaran. Selama ini proses pengklasifikasian rekapitulasi hasil tes kesehatan bagi mahasiswa baru masih dilakukan secara manual sehingga memperlambat proses kinerja pihak kampus dalam mengklasifikasikan rekapitulasi data tersebut. Fokus penelitian ini yaitu membandingkan algoritma *Naïve Bayes* dan *C45* dengan tujuan untuk mencari algoritma yang memiliki hasil akurasi paling baik untuk diimplementasikan pada sistem yang akan dibuat berdasarkan hasil perhitungan metode yang paling baik nantinya dalam mengolah data klasifikasi hasil tes kesehatan calon mahasiswa baru.

### Data Understanding

*Data understanding* merupakan proses identifikasi dan pemahaman data dari dataset yang telah diperoleh. Proses dalam data *understanding* meliputi pengambilan data, telaah data dasar berupa deskripsi dat dan visualisasi data. Data rekapitulasi hasil tes kesehatan diperoleh dari pihak Tata Usaha (TU) Akbid As-syifa yang beralamat di Kisaran Naga Kecamatan Kisaran Timur Sumatera Utara. Atribut yang menjadi variabel fitur (*x*) adalah tinggi badan, berat badan, identifikasi virus mata, kondisi buta warna, telinga, hidung, lidah, gigi, jantung, paru –paru dan reflek. Kemudian atribut yang menjadi variabel target (*class*) adalah Hasil Tes berupa Lulus atau Tidak Lulus. Berikut ini merupakan gambaran data *understanding* dataset yang digunakan:

```
In [4]: #menampilkan info dataset
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 62 entries, 0 to 61
Data columns (total 15 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   No               62 non-null    int64
 1   Nama Peserta    62 non-null    object
 2   No Tes          62 non-null    int64
 3   Tinggi Badan    62 non-null    int64
 4   Berat Badan     62 non-null    int64
 5   Virus Mata      62 non-null    object
 6   Buta Warna      62 non-null    object
 7   Telinga         62 non-null    object
 8   Hidung          62 non-null    object
 9   Lidah           62 non-null    object
10  Gigi            62 non-null    object
11  Jantung         62 non-null    object
12  Paru - Paru     62 non-null    object
13  Reflek          62 non-null    object
14  HasilTes       62 non-null    object
dtypes: int64(4), object(11)
memory usage: 7.4+ KB
```

Gambar 3. Informasi Dataset



Gambar 3. Visualisasi Data

### Data Preparation

Merupakan sebuah proses dalam data *mining* untuk mempersiapkan data agar data tersebut dapat diolah dengan maksimal. Tahapan yang dapat dilakukan dalam data *preparation* berupa pengecekan total *missing value*, memeriksa tipe data, melakukan *encoding data* yaitu aktivitas mengubah data dari format kategorik ke dalam format bertipe numerik maupun sebaliknya [14]. Tahapan – tahapan yang dilakukan dalam data *preparation* :

```
In [9]: #mengecek Missing Value pada Dataset
df.isnull().sum()

Out[9]: Tinggi Badan      0
        Berat Badan      0
        Virus Mata       0
        Buta Warna       0
        Telinga          0
        Hidung           0
        Lidah            0
        Gigi             0
        Jantung          0
        Paru - Paru     0
        Reflek           0
        HasilTes        0
        dtype: int64
```

Gambar 4. Pengecekan *Missing Value* Dataset

```
In [7]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 62 entries, 0 to 61
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  ---          -
0   Tinggi Badan    62 non-null    int64
1   Berat Badan    62 non-null    int64
2   Virus Mata     62 non-null    object
3   Buta Warna     62 non-null    object
4   Telinga        62 non-null    object
5   Hidung         62 non-null    object
6   Lidah          62 non-null    object
7   Gigi           62 non-null    object
8   Jantung        62 non-null    object
9   Paru - Paru    62 non-null    object
10  Reflek         62 non-null    object
11  HasilTes      62 non-null    object
dtypes: int64(2), object(10)
memory usage: 5.9+ KB
```

Gambar 5. Deskripsi Tipe Data

	Tinggi Badan	Berat Badan	Virus Mata	Buta Warna	Telings	Hidung	Lidah	Gigi	Jantung	Paru-Paru	Reflek	HasilTes
0	168	55	1	1	1	1	0	0	1	1	1	0
1	155	50	1	1	1	1	0	1	1	1	1	0
2	150	38	1	1	1	1	0	0	1	1	1	0
3	170	45	1	1	1	1	0	0	1	1	1	0
4	150	45	1	1	1	1	0	0	1	1	1	0

Gambar 6. Encoding Data

### Modelling

Pemodelan dalam penelitian ini menggunakan 2 model algoritma yakni *Naïve Bayes* dan C45 dikarenakan nantinya dari kedua model tersebut akan dilakukan perbandingan untuk memperoleh algoritma paling baik untuk klasifikasi rekapitulasi data kesehatan calon mahasiswa baru Akbid As-syifa. Selanjutnya kedua model tersebut akan dilakukan *split* data dalam 4 model besaran rasio yaitu 60:40, 70:30, 80:20 dan 90:10. Hal tersebut bertujuan untuk mengetahui algoritma yang paling banyak menghasilkan nilai akurasi yang tinggi di berdasarkan 4 rasio pembagian daa tersebut. Algoritma *Naïve Bayes* merupakan algoritma untuk klasifikasi *probabilistic* berdasarkan penerapan Teorema Bayes, dimana algoritma *Naïve Bayes* memprediksi peluang yang akan terjadi dimasa yang akan datang berdasarkan pengalaman yang telah terjadi di masa lalu (Parthiban et al., 2011; Sari et al., 2020) [15].

Persamaan dari teorema Bayes adalah sebagai berikut :

$$P(H|X) = \frac{p(X|H).p(h)}{p(x)} \quad (1)$$

Dimana :

X :Kelas data yang belum diketahui

H :Hipotesis merupakan suatu kelas yang lebih spesifik

P(H|X) :Probabilitas hipotesis nilai H berdasarkan kondisi nilai X (posteriori probabilitas

P(H) :Probabilitas hipotesis H (prior probabilitas)

P(X|H) :Probabilitas nilai X berdasarkan kondisi pada hipotesis nilai H

P(X) :Probabilitas X

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (*Decission Tree*). Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal dipahami dengan bahasa alami [16]. Dalam C45 terdapat nilai *Entropy* dan *Gain* dalam membuat pohon keputusan.

Persamaan untuk mencari nilai *Entropy* yaitu :

$$Entropy(S) = \sum_{i=0}^n -pi * \log_2 pi \quad (2)$$

Dimana :

S : Himpunan dari kasus

n : Jumlah dari banyaknya partisi S

pi : Jumlah kasus dari partisi ke-i

Setelah memperoleh nilai *Entropy*, selanjutnya menghitung nilai *Gain* dengan persamaan berikut ini :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \left[ \frac{|S_i|}{|S|} \right] * Entorpy(S_i) \quad (3)$$

Dimana :

- S : Himpunan dari kasus  
 n : Jumlah dari banyaknya partisi A  
 A : Atribut dataset  
 |S<sub>i</sub>| : Jumlah kasus dari partisi ke-i  
 |S| : Jumlah dari kasus S

Total data yang digunakan dalam penelitian ini sebanyak 62 data, kemudian data tersebut akan dibagi menjadi data *training* dan data *testing* dalam 4 rasio *split* data yaitu 60:40, 70:30, 80:20 dan 90:10.

Kemudian melakukan proses *training model* untuk menemukan model yang sesuai dengan dataset.

### Evaluation

Tahap selanjutnya adalah menghitung nilai evaluasi dari model algoritma yang sudah diimplementasikan kedalam data. Pengukuran nilai evaluasi dalam pengklasifikasian menggunakan *confusion matrix* yang pada umumnya memuat persentase nilai *precision*, *recall* dan *f1-score* [17].

### Deployment

Hasil dari penerapan algoritma terbaik dalam proses data *mining* kemudian diimplementasikan dan disajikan dalam visualisasi sebuah sistem aplikasi menggunakan bahasa pemrograman agar masyarakat khususnya pihak kampus Akbid As-syifa kirsan dapat dengan mudah menggunakannya.

## HASIL DAN PEMBAHASAN

Tahapan yang pertama kita akan melakukan analisis terhadap hasil akurasi dari kedua algoritma dalam dataset rekapitulasi hasil tes kesehatan bagi calon mahasiswa baru Akbid As-syifa.

Tabel 1. Perbandingan Tingkat Akurasi

Rasio Split Data	<i>Naïve Bayes</i>	<i>Decision Tree C45</i>
60:40	0.96 (96%)	100%
70:30	0.94 (94%)	100%
80:20	0.92 (92%)	100%
90:10	0.85 (85%)	100%

Setelah memperoleh hasil akurasi dari kedua algoritma diatas, diperoleh hasil bahwa tingkat akurasi dalam algoritma *Naïve Bayes* cenderung berubah besarnya disetiap rasio pengujian data yang berbeda dengan nilai evaluasi performa menggunakan *confusion matrix* pada rasio 60:40 sebesar 96%, 70:30 sebesar 95%, 80:20 sebesar 92%, 90:10 sebesar 86%. Nilai akurasi pada algoritma *Naïve Bayes*



semakin mengecil apabila data *trainingnya* semakin besar dibandingkan data *testing*. Namun berbeda dengan algoritma *Decision Tree* C45 tingkat akurasi cenderung stabil dan nilai evaluasi performa juga stabil karena nilai akurasi tetap sama di berbagai rasio pengujian data yakni sebesar 100%. Dengan memiliki tingkat akurasi dan performa yang rata – rata diatas 90% kedua algoritma tersebut dikategorikan sebagai *Good Classification*.

## SIMPULAN

Kesimpulan berdasarkan pembahasan diatas, Algoritma *Decision Tree* C45 dinilai lebih unggul daripada algoritma *Naïve Bayes* dengan tingkat akurasi sebesar 100% sehingga algoritma *Decision Tree* C45 lebih cocok diimplementasikan dalam aplikasi klasifikasi rekapitulasi tes kesehatan calon mahasiswa baru Akbid As-syifa untuk mempermudah pihak kampus Akbid As-syifa dalam melakukan klasifikasi rekapitulasi data hasil kesehatan bagi calon mahasiswa baru. Untuk hal tersebut diperkuat dengan tingkat akurasi dari algoritma *Decision Tree* C45 yang bernilai tetap dibandingkan algoritma *Naïve Bayes* meskipun sudah diuji dalam 4 model pengujian.

## DAFTAR PUSTAKA

- [1] “Mengenal Makna Kesehatan.” <https://pusatkrisis.kemkes.go.id/mengenal-makna-kesehatan#> (accessed Jun. 26, 2022).
- [2] M. N. Lestari, P. A. F. Islami, K. M. Moses, and A. P. Wibawa, “Implementasi metode fuzzy tsukamoto untuk menentukan hasil tes kesehatan pada penerimaan peserta didik baru di sekolah menengah kejuruan,” *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 4, no. 1, pp. 7–13, 2018, doi: 10.26594/register.v4i1.718.
- [3] F. Marisa and D. Purnomo, “Implementasi Metode Trend Moment ( Peramalan ) Mahasiswa Baru Universitas Widyagama Malang,” vol. 3, no. 2, 2018, doi: 10.31328/jointecs.v3i2.785.
- [4] D. Mauli, “Tanggung Jawab Hukum Dokter Terhadap Kesalahan Diagnosis Penyakit Kepada Pasien,” *Cepalo*, vol. 2, no. 1, p. 33, 2019, doi: 10.25041/cepalo.v2no1.1760.
- [5] M. F. Rifai, H. Jatnika, and B. Valentino, “Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist ( MOS ),” vol. 12, no. 2, pp. 131–144, 2019.
- [6] N. Hidayati, V. G. Utomo, and N. Wakhidah, “Penerapan Metode Naïve Bayes Dalam Pemeriksaan Kesehatan,” vol. 5, no. 2, pp. 174–178, 2019.
- [7] K. Handayani *et al.*, “Komparasi Algoritma C4 . 5 dan Naïve Bayes dalam Penentuan Status Kelayakan Donor Darah,” vol. 10, pp. 676–687, 2021.
- [8] I. S. R. R. F. Farid Ali Ma’ruf1, Aria Pratama2, “Penerapan Model Prediksi Menggunakan Algoritma C.45 Untuk Prediksi Kelulusan Siswa SMK Wahidin,” vol. 1, no. 1, pp. 16–20, 2021.
- [9] P. T. Wibowo and T. Informatika, “Rancang bangun sistem pakar jenis penyakit pada hasil tes darah,” vol. 2, no. 5, pp. 1–10, 2022.

- [10] A. Prasetio, M. H. Hasibuan, and P. Sitompul, “Simulasi Penerapan Metode Decision Tree ( C4 . 5 ) Pada Penentuan Status Gizi Balita,” vol. 4, no. 3, pp. 209–214, 2021.
- [11] J. Jurnal and S. Informasi, “PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN WALLPAPER MENGGUNAKAN ALGORITMA C4 . 5 STMIK Royal , Ksieran,” vol. 2, 2018.
- [12] D. P. Utomo, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” vol. 4, no. April, pp. 437–444, 2020, doi: 10.30865/mib.v4i2.2080.
- [13] A. Asri, A. Arifin, and W. Handoko, “IMPLEMENTASI METODE NAIVE BAYES UNTUK KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN,” vol. 2, no. 1, pp. 21–26, 2022.
- [14] S. R. Saragih and D. P. Utomo, “Penerapan Algoritma Prefix Code Dalam Kompresi Data Teks,” *KOMIK (Konferensi Nas. ...)*, vol. 4, pp. 249–252, 2020, doi: 10.30865/komik.v4i1.2691.
- [15] M. A. 1, \* A. S., 2, E. T. L. 3, and ., “Analisis Perbandingan Akurasi Algoritma Naïve Bayes dan C4.5 untuk Klasifikasi Diabetes,” vol. 5, no. 2, pp. 147–156, 2021, doi: 10.29408/edumatic.v5i2.3424.
- [16] P. Studi, T. Informatika, S. Tinggi, T. Dumai, and D. Mining, “Implementasi Metode C . 45 dalam Menganalisa Tingkat Kecemasan Mahasiswa Menyusun Tugas Akhir,” vol. 14, no. 1, pp. 2580–2582, 2021.
- [17] M. P. Endah Fauziningrum, “EVALUASI DAN PREDIKSI PENGUASAAN BAHASA INGGRIS MARITIM MENGGUNAKAN METODE DECISION TREE DAN CONFUSION MATRIX (STUDI KASUS DI UNIVERSITAS MARITIM AMNI),” pp. 217–231, 2021.