

Klasifikasi Tingkat Kemampuan Adaptasi Siswa pada Pembelajaran Online menggunakan Naïve Bayes

Dian Novitaningrum^{*1)}, La Ode Abd. El Hafizh Hidayat²⁾, Dea Angelia Kamil³⁾, Ahmad Rizki Maulana⁴⁾

1. Program Studi Teknik Informatika Universitas Selamat Sri, Kabupaten Kendal, Indonesia
2. Program Studi Magister Ilmu Komputer Departemen Ilmu Komputer dan Elektronika, Universitas Gadjah Mada, Yogyakarta, Indonesia
3. Program Studi Doktorat Ilmu Komputer Departemen Ilmu Komputer dan Elektronika, Universitas Gadjah Mada, Yogyakarta, Indonesia.
4. Program Studi Magister Ilmu Komputer Departemen Ilmu Komputer dan Elektronika, Universitas Gadjah Mada, Yogyakarta, Indonesia

Article Info

Kata Kunci: Adaptasi; *Classification*; *Features Selection*; *Naïve Bayes*; Pembelajaran Online

Keywords: *Adaptation*; *Classification*; *Feature Selection*; *Naïve Bayes*; *Online Learning*

Article history:

Received: 09 November 2025

Revised: 11 November 2025

Accepted: 19 November 2025

Available online: 01 Mei 2026

DOI :

[10.48144/suryainformatika.v16i1.2255](https://doi.org/10.48144/suryainformatika.v16i1.2255)

* Corresponding author.

Dian Novitaningrum

E-mail address: dnovitaningrum.uniss@gmail.com

ABSTRAK

Pandemi mengharuskan pihak sekolah untuk menerapkan sistem pembelajaran online. Perubahan cara kerja pembelajaran yang semula diterapkan melalui sistem pembelajaran tatap muka kemudian dilakukan secara online (daring) menyebabkan siswa dan para orang tua perlu melakukan adaptasi untuk perubahan tersebut. Tujuan dari penelitian ini adalah untuk melakukan klasifikasi terhadap kemampuan adaptasi siswa dengan diberlakukannya pembelajaran online (daring) dengan menggunakan bantuan Algoritma *Naive Bayes*. Pemilihan algoritma *Naive Bayes* dilakukan karena mudah dan cepat untuk diimplementasikan. Metode lain yang digunakan adalah metode untuk melakukan *feature selection*. Data *splitting* yang dipilih pada penelitian ini adalah 70% untuk data *training* dan 30% untuk data *testing*. Evaluasi model yang dilakukan dengan jumlah *Information Gain feature* yang berbeda yakni 3, 6, 9, dan 13 kombinasi *feature* yang memiliki nilai *gain* terbaik. Sedangkan, jumlah *information gain* untuk mendapatkan *feature* terbaik yaitu saat menghasilkan 6 *feature*. *Feature* terbaiknya yaitu *Class Duration*, *Financial Condition*, *Age*, *Institution Type*, *Location*, dan *Device*. Penelitian ini juga membandingkan penggunaan *library* dan yang menggunakan perhitungan manual. Pemilihan *feature selection* dipilih 6 *feature* maka didapatkan hasil *accuracy* tertinggi jika dihitung dengan perhitungan manual menghasilkan *accuracy* sejumlah 66.02%, sedangkan jika dihitung dengan perhitungan *library* nilai *accuracy* yang dihasilkan adalah 66.29%. Penggunaan *library* terpantau memiliki perbedaan hasil dibandingkan dengan perhitungan manual karena adanya perbedaan *step* yang digunakan.

ABSTRACT

The pandemic has forced schools to implement online learning systems. The change in teaching methods, which were originally implemented through face-to-face learning systems and then moved online, has required students and parents to adapt to these changes. The purpose of this study was to classify students' ability to adapt to online learning using the Naive Bayes algorithm. The Naive Bayes algorithm was chosen because it is easy and quick to implement. Another method used was feature selection. The data splitting chosen in this study was 70% for training data and 30% for testing data. Model evaluation was carried out with different Information Gain feature values,

namely 3, 6, 9, and 13 combinations of features with the best gain values. Meanwhile, the information gain value to obtain the best features was when producing 6 features. The best features were Class Duration, Financial Condition, Age, Institution Type, Location, and Device. This study also compares the use of libraries and manual calculations. Feature selection was chosen for 6 features, resulting in the highest accuracy when calculated manually, which was 66.02%, while when calculated using the library, the accuracy value was 66.29%. The use of libraries was found to produce different results compared to manual calculations due to differences in the steps used.

1. PENDAHULUAN

Pendidikan merupakan suatu proses menyeluruh yang bertujuan mengembangkan potensi, sikap, serta perilaku positif dalam diri seseorang. Melalui pendidikan, manusia diarahkan untuk melaksanakan peran kemanusiaannya sebagai hamba di hadapan Sang Pencipta. Keberadaan pendidikan sangatlah vital dalam kehidupan, terlebih di tengah pesatnya perkembangan dunia yang menuntut kemampuan adaptasi. Dengan demikian, kebutuhan akan pendidikan semakin mendesak seiring dengan kemajuan zaman. Secara esensial, pendidikan dapat dimaknai sebagai upaya pembinaan dan pendampingan yang diberikan secara berkesinambungan kepada peserta didik guna mewujudkan tujuan pendidikan.

Pandemi COVID-19 yang terjadi secara global, termasuk di Indonesia, menimbulkan dampak signifikan terhadap berbagai aspek kehidupan, salah satunya sektor pendidikan. Kondisi tersebut mengharuskan lembaga pendidikan melaksanakan pembelajaran jarak jauh, sehingga proses belajar siswa dan aktivitas mengajar guru tetap berlangsung meskipun peserta didik berada di rumah [1].

Secara luas, pembelajaran dapat dimaknai sebagai upaya guru dalam melakukan tindakan yang mampu memengaruhi sekaligus mengubah perilaku peserta didik. Suatu pembelajaran dapat dikatakan berhasil apabila prosesnya terlaksana sesuai dengan sistem kurikulum yang diterapkan di lembaga pendidikan tempat pembelajaran berlangsung [2].

Sesuai dengan Undang-Undang Sistem Pendidikan Nasional (UU Sisdiknas) No. 20 Tahun 2003, pendidikan nasional bertujuan meningkatkan kemampuan individu sekaligus membangun karakter serta peradaban bangsa yang bermartabat, guna mewujudkan kehidupan bangsa yang cerdas. Pendidikan juga diarahkan untuk mengembangkan potensi peserta didik secara optimal, sehingga dapat menjadi individu yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, kompeten, kreatif, mandiri, serta tumbuh menjadi warga negara yang demokratis dan bertanggung jawab [3].

Faktor terpenting untuk mendukung tujuan tersebut diperlukan tenaga pendidik yang kompeten dalam bidangnya masing-masing. Tidak hanya faktor tenaga pendidik saja yang diperlukan, namun juga faktor lain

seperti kesehatan antara tenaga pendidik dan peserta didik yang perlu diperhatikan. Kesehatan menjadi faktor penting lainnya untuk dapat melangsungkan kegiatan pembelajaran.

Pembelajaran Jarak Jauh (PJJ) sangat ditentukan oleh minat serta keseriusan peserta didik dalam mengikuti dan berpartisipasi aktif pada kegiatan belajar daring. Dalam hal ini, pendidik memiliki keterbatasan dalam memantau aktivitas belajar siswa melalui berbagai *platform online*, seperti Google Meet, Zoom, dan sejenisnya.

Selama masa pandemi COVID-19, proses pembelajaran dilaksanakan melalui sistem dalam jaringan (*online*) yang dikenal dengan istilah *daring*. Untuk mendukung pelaksanaannya, pemerintah memberikan subsidi kuota internet bagi pendidik dan peserta didik, mengingat semakin banyaknya masyarakat yang terinfeksi Covid-19. PJJ pada dasarnya merupakan model pembelajaran yang menekankan kemandirian, di mana guru tetap dapat menyampaikan materi kepada peserta didik tanpa harus melakukan pertemuan tatap muka secara langsung di ruang kelas yang sama [3].

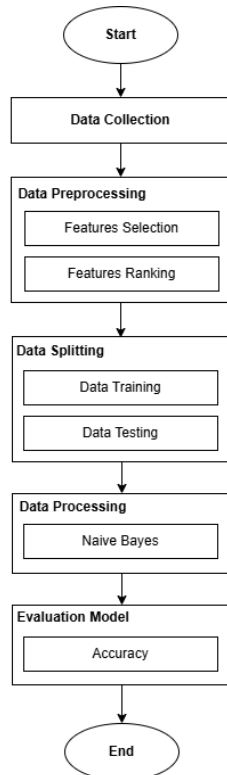
Pandemi COVID-19 memberikan pengaruh yang signifikan terhadap sektor Pendidikan. Pandemi mengharuskan pihak sekolah untuk menerapkan sistem pembelajaran *online*. Perubahan cara kerja pembelajaran yang semula diterapkan melalui sistem pembelajaran tatap muka kemudian dilakukan secara *online* menyebabkan siswa dan para orang tua perlu melakukan adaptasi untuk perubahan tersebut. Maka dari itu, penulis melakukan studi lanjut untuk klasifikasi kemampuan adaptasi siswa dengan diberlakukannya pembelajaran *online* menggunakan bantuan Algoritma Naive Bayes [4].

2. METODE PENELITIAN

Tujuan dari penelitian ini adalah untuk melakukan klasifikasi terhadap kemampuan adaptasi siswa dengan diberlakukannya pembelajaran *online* dengan menggunakan bantuan Algoritma Naive Bayes. Metode lain yang digunakan adalah metode untuk melakukan *feature selection*.

2.1 Kerangka Penelitian

Alur kerja dari sistem yang telah dibuat dijelaskan melalui *flowchart* pada Gambar 1. yaitu:



Gambar 1. *Flowchart* Sistem

Pada Gambar 1 menjelaskan bahwa setelah proses dimulai kemudian dilakukan *splitting data* yang dibagi menjadi data *training* dan data *testing*, dilanjutkan dengan *feature selection* dan *features ranking*. Kemudian dilakukan pemodelan dengan menggunakan algoritma Naïve Bayes. Setelah itu dilakukan evaluasi dan kemudian selesai.

2.2 Data Collection

Dalam penelitian ini, *dataset* yang digunakan bersumber dari *Kaggle* yang dapat diakses pada [link](#) berikut dengan judul “*Students Adaptability Level in Online Education*”. *Dataset* ini memiliki 13 *feature* : yaitu *Gender, Age, Education Level, Institution Type, IT Student, Location in Town, Load-shedding, Financial Condition, Internet Type, Network Type, Class Duration, Self LMS, dan Device*. Serta memiliki 1 target yaitu *Adaptivity level (Low, Moderate, High)* dengan jumlah total data sebanyak 1205.

2.3 Data Preprocessing

Penelitian ini memiliki tujuan untuk meningkatkan nilai *accuracy*, pada eksperimen ini dilakukan *feature selection* menggunakan metode *Information Gain*. dari metode tersebut, dapat diperoleh nilai *information gain* yang digunakan untuk perankingan. Semakin besar nilai *gain* dari suatu *feature* maka semakin berpengaruh *feature* tersebut untuk prediksi.

2.4 Data Splitting

Total jumlah baris pada *dataset* ini adalah 1205 baris. Dalam penelitian ini, *dataset* dibagi menjadi 70% data *training* (sejumlah 843 data) dan 30% data *testing* (sejumlah 342 data).

2.5 Data Processing

Data model yang dipergunakan adalah algoritma Naïve Bayes. *dataset* yang telah dipisahkan kedalam data *train* akan melewati fase pembelajaran pada dengan Algoritma klasifikasi Naive Bayes. Hasilnya yakni *predictive model* berupa *probabilitas prior* serta *probabilitas bersyarat* dari tiap-tiap *feature* yang dimiliki pada data *training*.

Setelah memastikan sistem telah menghasilkan *predictive model* dari Naive Bayes beserta *Features Ranking* dari *Information Gain* maka dapat dilakukan evaluasi terhadap data *testing* dengan 4 fase percobaan, dimana masing-masing percobaannya secara berturut-turut memakai *feature* sejumlah 3, 6, 9, dan 13 sesuai dengan ranking terbaik.

2.5.1. Algoritma Naïve Bayes

Algoritma Naive Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes [5], [6], [7], [8]. Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Probabilitas yang digunakan untuk mendapatkan estimasi akhir dihitung dari total frekuensi pada tabel Keputusan. Naive Bayes Classifier menunjukkan kinerja yang sangat baik dibandingkan dengan model klasifikasi lainnya. Persamaan dari teorema Bayes dijelaskan pada Persamaan (1) adalah : [9], [10]

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

Dimana :

X : Data dengan *class* atau label yang belum diketahui.

H : Hipotesis data X merupakan suatu *class* khusus.

P(H|X) : Untuk bukti (X), peluang hipotesis (H) benar (*posteriori probability*).

P(H) : Probabilitas awal dari hipotesis H (*prior probability*).

P(X|H) : Potensi data sampel (X) jika hipotesis (H) benar.

P(X) : Peluang data sampel X.

$$P(C|F1...Fn) = \frac{P(C).P(F1...Fn|C)}{P(F1...Fn)} \quad (2)$$

dengan

C : Kelas
 $F_1...F_n$: Karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi.

$$P(C|F_1, F_2, \dots, F_n) = P(C)P(F_1|C)P(F_2|C)P(F_3|C) \dots P(F_n|C) = P(C) \prod_{i=1}^n P(F_i|C) \quad (7)$$

Dengan demikian, rumus tersebut menggambarkan bahwa probabilitas suatu sampel dengan karakteristik tertentu untuk termasuk dalam kelas C (*posterior*) diperoleh dari hasil perkalian antara peluang awal kelas C sebelum sampel diproses (*prior*) dan peluang kemunculan karakteristik sampel dalam kelas C (*likelihood*), kemudian dibagi dengan peluang kemunculan karakteristik sampel secara keseluruhan (*evidence*). Oleh sebab itu, rumus tersebut dapat pula ditulis secara sederhana sebagai berikut :

$$Posterior = \frac{Prior\ likelihood}{evidence} \quad (3)$$

Nilai *evidence* bersifat konstan untuk setiap kelas pada satu sampel. Selanjutnya, nilai posterior tersebut akan dibandingkan dengan nilai posterior dari kelas lainnya untuk menentukan kategori kelas yang paling sesuai bagi sampel tersebut. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $(C|F_1, \dots, F_n)$ menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned} P(C|F_1...F_n) &= P(C) P(F_1...F_n|C) \\ &= P(C) P(F_1|C) P(F_2...F_n|C, F_1) \\ &= P(C) P(F_1|C) P(F_2|C, F_1) P(F_3...F_n|C, F_1, F_2) \\ &= P(C) P(F_1|C) P(F_2|C, F_1) P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned} \quad (4)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor – faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing – masing petunjuk (F_1, F_2, \dots, F_n) saling bebas (*independen*) satu sama lain.

Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut :

$$P(P_i|F_j) = \frac{P(F_i \wedge F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \quad (5)$$

untuk $i \neq j$, sehingga

$$P(F_i|C, F_j) = P(F_i|C) \quad (6)$$

Dari Persamaan (6) dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(C|F_1, F_2, \dots, F_n)$ dapat disederhanakan menjadi :

2.5.2. *Information Gain*

Information Gain menjadi sebuah metode seleksi fitur sederhana untuk menjadi sebuah bantuan dalam mengklasifikasikan data, salah satu tantangan dalam melakukan proses pembelajaran adalah banyaknya jumlah *feature* pada *dataset*. Banyaknya jumlah *feature* dapat berakibat pada mahalnya waktu serta proses komputasi dimana juga akan berpengaruh terhadap performansi pemodelan.

Oleh karena itu perlu adanya teknik untuk mengurangi jumlah *feature* yang ada. Dimana dalam pembahasan topik *Machine Learning* hal ini disebut sebagai Seleksi Fitur. Pada klasifikasi tujuan utama dari *Feature Selection* yakni untuk memilah dan memilih fitur-fitur mana saja yang berkontribusi besar dalam mengklasifikasikan data.

Seleksi fitur memiliki variasi yang berbeda-beda seperti *Chi-Squared*, *Likelihood-Ratio*, *DKM Criterion* dan masih banyak yang lainnya. Adapun dalam penelitian ini mengimplementasi Seleksi Fitur dengan metode *Information Gain*. *Information Gain* memanfaatkan komponen *Entropy*. *Entropy* dipergunakan untuk mengukur sebuah ketidakpastian kelas dengan menggunakan ketidakpastian kejadian akan suatu atau sebuah atribut tertentu [11].

$$Entropy(S) = \sum_i^c - p_i \log_2 p_i \quad (8)$$

Dimana :

- c : Total dari nilai yang terdapat pada kelas klasifikasi
- P_i : Total dari jumlah *sample* untuk kelas i .

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (9)$$

Dimana :

- A : Atribut / *features*
- v : Nilai untuk atribut A

- $Value(A)$: Kumpulan nilai atribut A
- $|S_v|$: Total *sample* untuk nilai v
- $|S|$: Total keseluruhan data *sample*
- $Entropy|S_v|$: *Entropy* untuk *sample* nilai v

Dalam bidang *Information Theory*, *Information Gain* merupakan reduksi terhadap *Information Entropy*. Dimana *Entropy* dapat diartikan sebagai ukuran keragaman suatu himpunan data. Pada persamaan (5) c menyatakan jumlah kelas sementara

P_i menyatakan porsi data sampel pada kelas i terhadap jumlah semua sampel di keseluruhan himpunan data.

Information Gain dapat digunakan sebagai indikator efektivitas suatu *feature* dalam mengklasifikasikan data. Pada persamaan (6) A adalah fitur, sementara $Values(A)$ adalah himpunan nilai yang mungkin pada fitur A . $|S_v|$ adalah jumlah sampel untuk nilai v , sementara $|S|$ adalah jumlah seluruh sampel data, dan $Entropy(S_v)$ menyatakan *entropy* untuk sampel-sampel yang memiliki nilai v [4].

Sebagai contoh pada dataset terdapat *feature Gender* d, maka perhitungannya sebagaimana langkah-langkah dijelaskan pada Gambar 2:

$$\begin{aligned}
 Values(Gender) &= Boy, Girl \\
 S &= [480_{low}, 625_{moderate}, 100_{High}], |S| = 1205 \\
 S_{Boy} &= [245_{low}, 347_{moderate}, 71_{High}], |S_{Boy}| = 663 \\
 S_{Girl} &= [235_{low}, 278_{moderate}, 29_{High}], |S_{Boy}| = 542 \\
 Entropy(S) &= -\left(\frac{480}{1205}\right) \log_2 \left(\frac{480}{1205}\right) - \left(\frac{625}{1205}\right) \log_2 \left(\frac{625}{1205}\right) - \left(\frac{100}{1205}\right) \log_2 \left(\frac{100}{1205}\right) \\
 &= 1,318208848 \\
 Entropy(S_{Boy}) &= -\left(\frac{245}{663}\right) \log_2 \left(\frac{245}{663}\right) - \left(\frac{347}{663}\right) \log_2 \left(\frac{347}{663}\right) - \left(\frac{71}{663}\right) \log_2 \left(\frac{71}{663}\right) \\
 &= 1,36476685 \\
 Entropy(S_{Girl}) &= -\left(\frac{235}{542}\right) \log_2 \left(\frac{235}{542}\right) - \left(\frac{278}{542}\right) \log_2 \left(\frac{278}{542}\right) - \left(\frac{29}{542}\right) \log_2 \left(\frac{29}{542}\right) \\
 &= 1,242797474 \\
 Gain(S, Gender) &= Entropy(S) - \sum_{v \in \{Boy, Girl\}} \frac{|S_v|}{|S|} Entropy(S_v) \\
 &= 1,318208848 - \left(\left(\frac{663}{1205}\right) \times 1,36476685 + \left(\frac{542}{1205}\right) \times 1,242797474\right) \\
 &= 1,318208848 - 1,309905936 \\
 &= 0,008302912899
 \end{aligned}$$

Gambar 2. Contoh Langkah Perhitungan pada *Features Gender*

2.6 Evaluation Model

Model evaluasi yang dipergunakan adalah nilai *accuracy* nya. *Accuracy* dilakukan dengan membandingkan nilai dengan perhitungan manual maupun hitungan *library*.

3. HASIL DAN PEMBAHASAN

Berdasarkan langkah-langkah dari Algoritma *Naive Bayes* yang telah dijabarkan pada metodologi penelitian, pada bagian pembahasan ini akan dijelaskan secara rinci mengenai perhitungan manual dari dataset tersebut. Langkah-langkah secara detailnya yaitu :

- a) Di tahap awal penulis melakukan perhitungan information gain di tiap fiturnya terhadap keseluruhan dataset. Berdasarkan komponen pada persamaan 8 dan 9 maka dibawah ini disajikan perhitungan manual untuk salah satu fitur yakni fitur Gender dengan nilai kategori Boy dan Girl. Berikut merupakan perhitungannya yang dijelaskan pada Gambar 3, yaitu:

$$Values(Gender) = Boy, Girl$$

$$S = [480_{low}, 625_{moderate}, 100_{High}], |S| = 1205$$

$$S_{Boy} = [245_{low}, 347_{moderate}, 71_{High}], |S_{Boy}| = 663$$

$$S_{Girl} = [235_{low}, 278_{moderate}, 29_{High}], |S_{Boy}| = 542$$

$$Entropy(S) = -\left(\frac{480}{1205}\right) \log_2 \left(\frac{480}{1205}\right) - \left(\frac{625}{1205}\right) \log_2 \left(\frac{625}{1205}\right) - \left(\frac{100}{1205}\right) \log_2 \left(\frac{100}{1205}\right) = 1,318208848$$

$$Entropy(S_{Boy}) = -\left(\frac{245}{663}\right) \log_2 \left(\frac{245}{663}\right) - \left(\frac{347}{663}\right) \log_2 \left(\frac{347}{663}\right) - \left(\frac{71}{663}\right) \log_2 \left(\frac{71}{663}\right) = 1,36476685$$

$$Entropy(S_{Girl}) = -\left(\frac{235}{542}\right) \log_2 \left(\frac{235}{542}\right) - \left(\frac{278}{542}\right) \log_2 \left(\frac{278}{542}\right) - \left(\frac{29}{542}\right) \log_2 \left(\frac{29}{542}\right) = 1,242797474$$

$$\begin{aligned}
 Gain(S, Gender) &= Entropy(S) - \sum_{v \in \{Boy, Girl\}} \frac{|S_v|}{|S|} Entropy(S_v) \\
 &= 1,318208848 - \left(\left(\frac{663}{1205}\right) \times 1,36476685 + \left(\frac{542}{1205}\right) \times 1,242797474\right) \\
 &= 1,318208848 - 1,309905936 \\
 &= 0,008302912899
 \end{aligned}$$

Gambar 3. *Information Gain* pada setiap *features*

- b) Membagi dataset menjadi 70 % data training sejumlah 843 data dan 30% data testing sebanyak 342 data.
- c) Menghitung jumlah data dan probabilitas pada tiap class target pada data training.

Penghitungan manual dihitung dengan memanfaatkan microsoft excel untuk membantu penghitungan. Kemudian dilakukan penghitungan untuk masing-masing kategori yang berada dalam *class target*.

Selanjutnya dilakukan penghitungan nilai Probabilitas yang dijelaskan pada Gambar 4:

$$Probabilitas = \frac{\text{nilai kategori dari class target}}{\text{jumlah total setiap kategori dalam class target}} \quad (10)$$

Dari rumus pada persamaan (10) menghasilkan nilai yang telah dituliskan pada Tabel 1.

Tabel 1 Penghitungan Manual Output (Class Target)

Class Target	Low	Prob (Low)	Mod	Prob (Mod)	High	Prob (High)	Jumlah class target
Jumlah	342	0,405	432	0,512	69	0,081	843

- d) Menghitung probabilitas pada tiap kelas *feature* pada data *training*.

Perhitungan manual dihitung dengan memanfaatkan microsoft excel untuk membantu penghitungan. Kemudian dilakukan penghitungan untuk masing-masing kategori yang berada dalam kelas *feature*.

Kemudian dilakukan penghitungan nilai Probabilitas dengan cara:

$$Probabilitas = \frac{\text{jumlah kemunculan kategori } x \text{ pada kategori target } y}{\text{jumlah kemunculan kategori target } y} \quad (11)$$

Berdasarkan hasil perhitungan *Information Gain*, kami menuliskan 6 probabilitas *feature* terbaik.

Tabel 2. menjelaskan mengenai probabilitas mengenai *feature* “Class Duration”.

Tabel 2. Probabilitas *Feature* ‘Class Duration’

Class Target	Low	Moderate	High
0	0.942	0.057	0.0
1-3	0.362	0.545	0.092
3-6	0.198	0.702	0.099

Tabel 3. menjelaskan mengenai probabilitas mengenai *feature* “Financial Condition”.

Tabel 3. Probabilitas *Feature* ‘Financial Condition’

Class Target	Low	Moderate	High
Poor	0.554	0.347	0.097
Mid	0.383	0.578	0.037
Rich	0.134	0.327	0.538

Tabel 4. menjelaskan mengenai probabilitas mengenai *feature* “Age”.

Tabel 4. Probabilitas *Feature* ‘Age’

Class Target	Low	Moderate	High
1-5	0.225	0.774	0.0
6-10	0.433	0.2	0.366
11-15	0.341	0.591	0.066
16-20	0.543	0.434	0.021
21-25	0.374	0.512	0.112
26-30	0.576	0.288	0.134

Tabel 5. menjelaskan mengenai probabilitas mengenai *feature* “Institution Type”.

Tabel 5. Probabilitas *Feature* ‘Institution Type’

Class Target	Low	Moderate	High
Government	0.624	0.327	0.048
Non Government	0.305	0.597	0.097

Tabel 6. menjelaskan mengenai probabilitas mengenai *feature* “Location”.

Tabel 6. Probabilitas *Feature* ‘Location’

Class Target	Low	Moderate	High
No	0.627	0.338	0.033
Yes	0.346	0.558	0.094

Tabel 7. menjelaskan mengenai probabilitas mengenai *feature* “Device”.

Tabel 7. Probabilitas *Feature* ‘Device’

Class Target	Low	Moderate	High
Mobile	0.439	0.495	0.064
Tab	0.117	0.823	0.058
Computer	0.236	0.570	0.192

e) Memprediksi *class target* pada data *testing* dengan menggunakan persamaan (4), dilakukan prediksi *class target* pada tiap *instance*.

Tabel 8. Prediksi terhadap *class target* setiap *instance*

No	Gender	Age	Education Level
x1	Boy	16-20	University
x2	Boy	11-15	School
x3	Boy	11-15	School
x4	Girl	21-25	University

No	Institution Type	IT Student	Location	Load-shedding
x1	Government	No	Yes	Low
x2	Non Gov	No	Yes	Low
x3	Non Gov	No	Yes	Low
x4	Non Gov	No	Yes	Low

No	Financial Condition	Internet Type	Network Type
x1	Mid	Wifi	3G
x2	Mid	Mobile Data	4G
x3	Mid	Mobile Data	4G
x4	Mid	Wifi	4G

No	Class Duration	Self Lms	Device
x1	1-3	Yes	Computer
x2	1-3	No	Mobile
x3	1-3	No	Mobile
x4	3-6	No	Tab

$$P(\text{Low} | \mathbf{x1}) = [P(\text{Boy} | \text{Low}) P(16-20 | \text{Low}) P(\text{University} | \text{Low}) P(\text{Government} | \text{Low}) P(\text{No} | \text{Low}) P(\text{Yes} | \text{Low}) P(\text{Low} | \text{Low}) P(\text{Mid} | \text{Low}) P(\text{Wifi} | \text{Low}) P(3G | \text{Low}) P(1-3 | \text{Low}) P(\text{Yes} | \text{Low}) P(\text{Computer} | \text{Low}) P(\text{Adaptivity Level} = \text{Low})]$$

$$P(\text{Moderate} | \mathbf{x1}) = [P(\text{Boy} | \text{Moderate}) P(16-20 | \text{Moderate}) P(\text{University} | \text{Moderate}) P(\text{Government} | \text{Moderate}) P(\text{No} | \text{Moderate}) P(\text{Yes} | \text{Moderate}) P(\text{Low} | \text{Moderate}) P(\text{Mid} | \text{Moderate}) P(\text{Wifi} | \text{Moderate}) P(3G | \text{Moderate}) P(1-3 | \text{Moderate}) P(\text{Yes} | \text{Moderate}) P(\text{Computer} | \text{Moderate}) P(\text{Adaptivity Level} = \text{Moderate})]$$

$$P(\text{High} | \mathbf{x}_1) = [P(\text{Boy} | \text{High}) P(16-20 | \text{High}) P(\text{University} | \text{High}) P(\text{Government} | \text{High}) P(\text{No} | \text{High}) P(\text{Yes} | \text{High}) P(\text{Low} | \text{High}) P(\text{Mid} | \text{High}) P(\text{Wifi} | \text{High}) P(3G | \text{High}) P(1-3 | \text{High}) P(\text{Yes} | \text{High}) P(\text{Computer} | \text{High}) P(\text{Adaptivity Level} = \text{High})]$$

- f) Hasil dari probabilitas tiap class target dibandingkan dan diambil kelas yang memiliki nilai paling besar.

Dari prediksi yang telah dilakukan dapat diperoleh *confusion matrix* yang nilainya digunakan untuk menghitung *accuracy* yang akan dijelaskan pada Tabel 9.

Tabel 9. Hasil penghitungan *Confusion Matrix*

		Actual Class		
		Low	Moderate	High
Predicted Class	Low	39	6	0
	Moderate	97	167	32
	High	0	0	0

Dari Tabel 9. diperoleh nilai *accuracy* sebesar 0.627. Sedangkan penjelasan mengenai nilai *information gain* untuk setiap features dijelaskan pada Tabel 10.

Tabel 10. Urutan nilai *information gain* untuk tiap *features*

No	Features	Information Gain
1	Class Duration	0.147921
2	Financial Condition	0.092548
3	Age	0.073862
4	Institution Type	0.063662
5	Location	0.049358
6	Device	0.031164
7	Education Level	0.026440
8	Network Type	0.018999
9	Self Lms	0.017757
10	Internet Type	0.012447
11	IT Student	0.011999
12	Gender	0.008303
13	Load-shedding	0.005881

Dari Tabel 10. Dihasilkan bahwa *features* yang ditampilkan dalam list tabel merupakan urutan *features* yang mempunyai nilai *information gain* dari yang terbesar sampai yang terkecil.

4. KESIMPULAN

Dilakukan percobaan dengan memilih 3 *features*, 6 *features*, 9 *features* dan 13 *features*. Dapat ditarik kesimpulan bahwa prediksi menggunakan 6 *feature* terbaik (*Class Duration*, *Financial Condition*, *Age*, *Institution Type*, *Location*, dan *Device*) memiliki hasil akurasi paling tinggi dibandingkan dengan menggunakan 3, 9, dan 13 kombinasi *feature*. Pemilihan *feature selection* dipilih 6 *feature* maka didapatkan hasil jika dihitung dengan perhitungan manual menghasilkan *accuracy* sejumlah 66.02%, sedangkan jika dihitung dengan perhitungan *library* nilai *accuracy* yang dihasilkan adalah 66.29%

Penggunaan *library* terpantau memiliki perbedaan hasil dibandingkan dengan perhitungan manual karena adanya perbedaan *step* yang digunakan. Namun, hasil akurasi tidak memiliki hasil yang jauh berbeda dengan kombinasi terbaik menggunakan 6 dan 9 kombinasi *feature* dengan *ranking* tertinggi.

REFERENSI

- [1] A. M. Basar, "Problematika Pembelajaran Jarak Jauh Pada Masa Pandemi Covid-19," *Edunesia : Jurnal Ilmiah Pendidikan*, vol. 2, no. 1, pp. 208–218, Jan. 2021, doi: 10.51276/edu.v2i1.112.
- [2] S. Sabaniah, D. F. Ramdhan, and S. K. Rohmah, "Peran Guru dalam Pelaksanaan Pembelajaran Jarak Jauh di Tengah Wabah Covid - 19," *Edunesia : Jurnal Ilmiah Pendidikan*, vol. 2, no. 1, pp. 43–54, Jan. 2021, doi: 10.51276/edu.v2i1.77.
- [3] P. W. Nugroho and M. A. Maulana, "Kemandirian Belajar dalam Pembelajaran Jarak Jauh," *Advice: Jurnal Bimbingan dan Konseling*, vol. 3, no. 1, pp. 10–16, Apr. 2021, doi: 10.32585/advice.v3i1.1429.
- [4] H. Susana, N. Suarna, F. Fathurrohman, and K. Kaslani, "Penerapan Model Klasifikasi Metode Naive Bayes terhadap Penggunaan Akses Internet," *Jurnal Sistem Informasi dan Teknologi Informasi*, vol. 4, no. 1, pp. 1–8, 2022, doi: <https://doi.org/10.52005/jursistekni.v4i1.96>.
- [5] H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *Jurnal Teknologi Informasi dan Komunikasi (TIK-omSiN)*, vol. 8, no. 2, Oct. 2020, doi: 10.30646/tik-omsin.v8i2.500.
- [6] N. H. Alfianty and S. Mulyati, "Penerapan Naive Bayes untuk Klasifikasi Data Penyakit Pada Anak," *Jurnal AUTOMATA*, vol. 3, no. 1, pp. 1–6, 2022.
- [7] J. S. Komputer, K. Buatan, and A. Ridwan, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," 2020.
- [8] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, p. 219, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.
- [9] U. A. Susilo, B. N. Sari, and A. Rizal, "Analisis Sentimen Opini Masyarakat Twitter terhadap Kasus Diabetes di Indonesia menggunakan Algoritma Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, pp. 682–689, Jul. 2025, doi: 10.23960/jitet.v13i3.6942.

- [10] M. R. A. Rifai, B. N. Sari, and G. Garno, "Analisis Sentimen terhadap Acara Ruangguru Clash Of Champions (COC) dengan Algoritma Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, pp. 990–1000, Jul. 2025, doi: 10.23960/jitet.v13i3.7112.
- [11] E. Rahmanita, Y. D. P. Negara, Y. Kustiyahningsih, V. Sasmeka, and B. K. Khotimah, "Implementasi Metode Naive Bayes dan Information Gain Untuk Klasifikasi Penyakit dan Hama Tanaman Jagung," *Teknika*, vol. 12, no. 3, pp. 198–204, Oct. 2023, doi: 10.34148/teknika.v12i3.684.