

## KLASIFIKASI TINGKAT KECANDUAN INTERNET TERHADAP REMAJA PEKANBARU MELALUI PENDEKATAN ALGORITMA NAÏVE BAYES

Mhd Ikhsanul Fikri <sup>1</sup>, Elvia Budianita <sup>2</sup>, Iwan Iskandar <sup>3</sup>, Eka Pandu Cynthia <sup>4</sup>

<sup>1,2,3,4</sup> Universitas Islam Negeri Sultan Syarif Kasim Riau

(Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau)

(Jl. HR. Soebrantas KM. 15 No. 155, Simpang Baru, Kota Pekanbaru, 28293, Indonesia)

email: <sup>1</sup>12050112674@students.uin-suska.ac.id, <sup>2</sup>elvia.budianita@uin-suska.ac.id,

<sup>3</sup>Iwan.iskandar@uin-suska.ac.id, <sup>4</sup>eka.pandu.cynthia@uin-suska.ac.id

### Abstrak

Penggunaan internet terus meningkat di kalangan remaja. Namun, kemampuan remaja dalam memilah aktivitas internet yang bermanfaat belum sepenuhnya terwujud. Menurut survei APJII 2022, penggunaan internet pada usia 13-18 tahun meningkat hingga 99,16%. Hal ini menunjukkan peningkatan signifikan terhadap kecanduan internet. Sehingga dilakukan penelitian untuk mengevaluasi akurasi klasifikasi kecanduan internet terhadap remaja Pekanbaru menggunakan data mining dengan algoritma Naïve Bayes. Data yang digunakan sebanyak 510 data melalui kusioner. Hasil penelitian dalam klasifikasi menerapkan pengujian 10-Fold Cross Validation dengan model data latih 459 data dan diuji pada 51 data untuk pengujian. Didapatkan bahwa nilai akurasi tertinggi yaitu pada fold ke-3 dengan nilai 98% memiliki nilai precision, recall, dan f1-score adalah 98%, 99%, dan 98%. Untuk nilai akurasi terendah yaitu pada fold ke-1 dengan nilai 86% memiliki nilai precision, recall, dan f1-score adalah 86%, 87%, dan 86%. Untuk performa rata-rata yang diperoleh melalui hasil 10-fold Cross Validation menunjukkan bahwa nilai accuracy, precision, recall, dan f1-score adalah 93%, 87,3%, 89,9%, dan 88,1%. Berdasarkan hasil rata-rata akurasi yang diperoleh sebesar 93% menunjukkan metode Naïve Bayes dapat mengklasifikasikan tingkat kecanduan internet yang terdiri atas 4 kelas yaitu normal, mild, moderate, severe.

**Kata kunci:** Klasifikasi; Data Mining; Naïve Bayes; Kecanduan internet

### Abstract

Internet use continues to increase among teenagers. However, the ability of teenagers to sort out useful internet activities has not been fully realized. According to the 2022 APJII survey, internet usage among 13-18 year olds increased to 99.16%. This shows a significant increase in internet addiction. So that research was conducted to evaluate the accuracy of the classification of internet addiction to Pekanbaru teenagers using data mining with the Naïve Bayes algorithm. The data used were 510 data through questionnaires. The results of the research in classification applied 10-Fold Cross Validation testing with a training data model of 459 data and tested on 51 data for testing. It was found that the highest accuracy value was in the 3rd fold with a value of 98% having precision, recall, and f1-score values of 98%, 99%, and 98%. For the lowest accuracy value, namely the 1st fold with a value of 86%, the precision, recall, and f1-score values are 86%, 87%, and 86%. The average performance obtained through 10-fold Cross Validation results shows that the accuracy, precision, recall, and f1-score values are 93%, 87.3%, 89.9%, and 88.1%. Based on the average accuracy obtained of 93%, it shows that the Naïve Bayes method can classify the level of internet addiction which consists of 4 classes, namely normal, mild, moderate, severe.

**Keywords:** Classification; Data Mining; Naïve Bayes; Internet addiction

## 1. PENDAHULUAN

Penggunaan teknologi informasi, seperti internet terus meningkat di kalangan berbagai kelompok terutama remaja. Meskipun demikian, remaja belum sepenuhnya memiliki kemampuan untuk memilah aktivitas internet yang membawa manfaat. Untuk itu masa remaja merupakan suatu tahap perkembangan yang dilalui manusia sebelum mencapai usia dewasa [1]. Berdasarkan data survei APJII tahun 2022, penggunaan internet mengalami peningkatan signifikan pada kelompok usia 13 hingga 18 tahun mengalami kenaikan sebesar 99,16% [2][3]. Usia ini termasuk dalam masa remaja, temuan ini mencerminkan bahwa penggunaan internet, terutama di kalangan remaja mengalami peningkatan yang signifikan, menunjukkan bahwa sebagian besar remaja telah mengalami kecanduan internet [4]. Salah satu konsekuensi dari tingginya penggunaan internet pada remaja adalah bahwa mereka cenderung menjadi terlalu bergantung pada internet untuk menghabiskan banyak waktu mereka, yang menyebabkan mereka rentan terhadap kecanduan internet [5]. Kecanduan internet pada remaja sering ditandai oleh kesulitan dalam mengendalikan keinginan untuk terus menggunakan internet [6]. Untuk itu, dilakukan penelitian untuk mengetahui akurasi klasifikasi tingkat kecanduan internet terhadap remaja Pekanbaru dengan proses data mining melalui pendekatan algoritma *Naïve Bayes*.

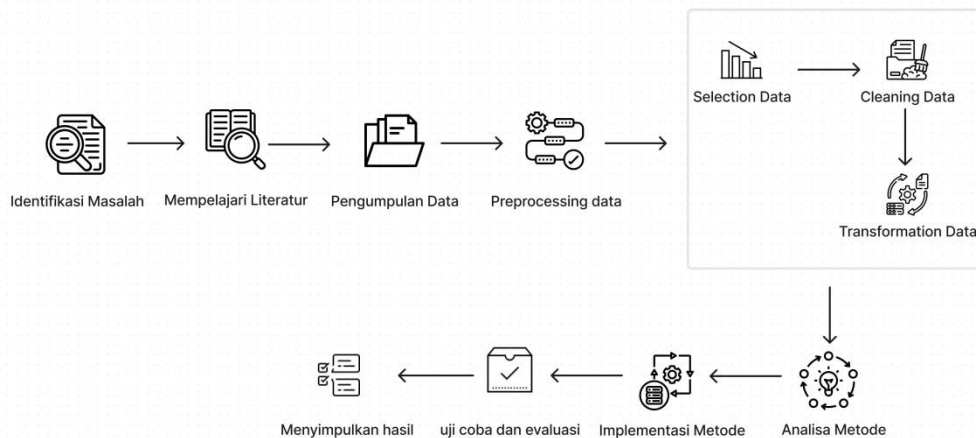
Data mining merupakan teknik dengan menggali pola-pola dari kumpulan data besar dengan tujuan mengolah data untuk menciptakan pengetahuan baru yang berakar dari informasi yang sudah ada sebelumnya. Hasil analisis tersebut dapat menjadi panduan untuk membuat pilihan yang lebih baik di kemudian hari [7]. Proses ini melibatkan langkah-langkah seperti *Selection Data*, *Cleaning Data*, dan *Transformation Data* [8]. Salah satu teknik yang digunakan untuk menganalisis kecanduan yaitu *Naïve Bayes*. Metode *Naïve Bayes* sebuah algoritma yang mengklasifikasikan berdasarkan probabilitas terjadinya di masa depan yang diprediksi. metode ini dinilai sederhana dan efektif untuk analisis data, salah satu algoritma untuk mengklasifikasikan statistik. Dengan menggunakan perhitungan probabilitas, pengklasifikasi algoritma *Naïve Bayes* dapat memprediksi seberapa mungkin suatu kumpulan data yang termasuk dalam kelas tertentu akan masuk ke dalam kelas tersebut [9]. *Naïve Bayes* terdapat pengelompokan dengan membagi data menjadi dua bagian, yaitu data latih dan data uji [10].

Dalam penelitian terkait melakukan penerapan algoritma *Naïve Bayes* untuk mengklasifikasi pengaruh negatif *game online* bagi remaja milenial. Penelitian tersebut dinyatakan Normal dengan akurasi keseluruhan 89.00%. *Class recall* kecanduan 77.27%, *Class recall* normal 92.31%, *Class precision* kecanduan 73.91%, dan *Class precision* normal 93.51% [11]. Penelitian lainnya, Penerapan Algoritma *Naïve Bayes Classifier* Dalam Klasifikasi Status Gizi Balita dengan Pengujian *K-Fold Cross Validation*. Penelitian ini mencapai akurasi tertinggi pada iterasi ke-5 sebesar 82,94%, sedangkan nilai akurasi terendah di iterasi ke-6 sebesar 65,88%. Rata-rata skor akurasi keseluruhan adalah 75,47%. Sedangkan nilai rata-rata presisi yang ditentukan sebesar 81,36% dan rata-rata *recall* sebesar 75,47% [12]. Penelitian implementasi metode *Naïve Bayes Classifier* pada evaluasi kepuasan mahasiswa terhadap pembelajaran daring. Penelitian ini mencapai akurasi sebesar 83,33%. Pada prediksi tidak memuaskan *class precision* memiliki nilai 50% sedangkan untuk prediksi memuaskan memiliki 100%. Selain itu *class recall* pada *true* tidak memuaskan memiliki nilai 50% sebaliknya *class recall* pada *true* memuaskan memiliki 100% [13].

Dalam penelitian terdahulu yang terkait dengan beberapa topik yang diatas terdapat perbedaan parameter yang di digunakan. Dalam penelitian ini parameter berasal dari pertanyaan kusioner IAT yaitu dari kusioner IAT Internet Young. Tujuan dari penelitian ini yaitu mengklasifikasi tingkat kecanduan internet terhadap remaja Pekanbaru melalui pendekatan algoritme *Naïve Bayes*. Manfaat penelitian ini diharapkan akan memberikan nilai akurasi untuk klasifikasi tingkat kecanduan internet terhadap remaja pekanbaru melalui pendekatan algoritme *Naïve Bayes*.

## 2. METODE PENELITIAN

Metode penelitian adalah serangkaian alur dalam suatu penelitian yang dilaksanakan secara teratur dan berpikir logis dengan tujuan mencapai hasil yang diinginkan. Alur konseptual diagram penelitian ini terdapat didalam Gambar 1.



Gambar 1. Alur Konseptual diagram

### 2.1. Identifikasi Masalah

Langkah awal yang penting dalam sebuah penelitian adalah mengidentifikasi masalah. Pada fase ini, kegiatan penelitian dilakukan untuk merinci dan memahami esensi permasalahan serta mencari solusi yang sesuai. Fokus permasalahan dalam penelitian ini adalah bagaimana menganalisis tingkat kecanduan internet terhadap remaja Pekanbaru melalui pendekatan algoritma *Naïve Bayes*.

### 2.2. Mempelajari Literatur

Tahap mempelajari literatur melibatkan pencarian dan analisis masalah yang dihadapi, kemudian menetapkan cakupan permasalahan, konteks latar belakang, serta memeriksa berbagai literatur terkait yang berkaitan dengan permasalahan tersebut. Dalam proses mempelajari literatur ini, teori-teori terkait penelitian ini pencarian dilakukan untuk menganalisis beberapa penelitian yang dianggap memiliki kesamaan sehingga dapat dijadikan patokan dan membentuk keterbaharuan untuk penelitian yang sedang berlangsung. Sumber jenis literatur seperti *ebook* dan jurnal.

### 2.3. Pengumpulan Data

Tahap pengumpulan data merupakan langkah untuk mencari informasi yang relevan dengan isu penelitian yang sedang difokuskan pada tahap penelitian. Data yang digunakan adalah data primer. Data dikumpulkan dengan menyebarkan form kusioner kepada remaja Pekanbaru untuk memperoleh informasi mengenai tingkat kecanduan internet di kalangan usia remaja. Pengambilan data melalui kusioner yang telah disebarakan mendapat 511 data responden.

### 2. 4. Preprocessing data

Preprocessing data adalah tahap awal yang krusial dalam menganalisis data, dimana informasi yang diperoleh akan diolah dan disiapkan untuk tahap klasifikasi selanjutnya. Tujuan untuk memastikan bahwa sampel data yang telah terkumpul melewati proses seleksi atribut data, di mana data akan disaring sesuai dengan kebutuhan analisis yang dilakukan [14]. Dalam konteks ini, pemrosesan data menjadi tahap yang krusial untuk mengoptimalkan performa algoritma *Naïve Bayes* dalam mengklasifikasikan tingkat kecanduan internet. Data Preprocessing ini melibatkan proses tahapan *selection*, *cleaning*, dan *transformation* dengan tujuan untuk proses memilih variabel atau atribut tertentu, data yang kurang lengkap, dan proses mengubah suatu data yang ada.

### 2.4.1 Selection Data

Penelitian ini memanfaatkan data yang dikumpulkan melalui Kuisisioner yang dilakukan terhadap remaja di Pekanbaru dengan menggunakan 20 pertanyaan berdasarkan kuisisioner internet IAT oleh Dr. Young. Penelitian ini menggunakan versi Indonesia dari IAT berdasarkan versi Bahasa Inggris IAT Dr. Young [15][16]. Pertanyaan kuisisioner IAT sudah divalidasi oleh pakar yang ahli dalam bidang yaitu seorang dosen psikologi. Data yang akan dipakai dalam studi ini yaitu hasil pengumpulan Kuisisioner IAT dari para responden. Kuisisioner IAT ditampilkan terdapat di dalam Tabel 1.

**Tabel 1.** Kuisisioner IAT

No	Pertanyaan
1	Seberapa sering Anda <i>online</i> lebih lama dari yang Anda rencanakan?
2	Seberapa sering Anda mengabaikan tugas rumah tangga demi bisa <i>online</i> lebih lama?
3	Seberapa sering Anda lebih memilih kesenangan main internet dibandingkan menghabiskan waktu bersama teman?
4	Seberapa sering Anda menjalin pertemanan baru dengan sesama pengguna internet?
5	Seberapa sering orang lain dalam hidup Anda mengeluh karena seringnya Anda menghabiskan waktu untuk <i>online</i> ?
6	Seberapa sering nilai Anda turun atau tugas sekolah Anda tercecer karena Anda sering <i>online</i> ?
7	Seberapa sering Anda memeriksa <i>email</i> dulu sebelum melakukan hal lain?
8	Seberapa sering kualitas pekerjaan atau produktivitas Anda merosot karena internet?
9	Seberapa sering Anda jadi mudah tersinggung atau tertutup jika ada orang yang bertanya apa yang Anda lakukan saat <i>online</i> ?
10	Seberapa sering Anda menutupi kecemasan Anda tentang kehidupan nyata dengan nyamannya hidup dunia internet?
11	Seberapa sering Anda merasa takut bahwa hidup tanpa internet itu akan membosankan, tidak bermakna, dan tidak menyenangkan?
12	Seberapa sering Anda merasa takut bahwa hidup tanpa internet itu akan membosankan, tidak bermakna, dan tidak menyenangkan?
13	Seberapa sering Anda menggerutu, membentak, atau merasa kesal ketika ada orang yang mengganggu ketika Anda sedang <i>online</i> ?
14	Seberapa sering Anda kehilangan jam tidur karena bergadang demi bisa <i>online</i> saat malam?
15	Seberapa sering Anda memikirkan internet ketika sedang offline dan berkhayal sedang <i>online</i> ?
16	Seberapa sering Anda berkata “sebentar lagi” ketika sedang <i>online</i> ?
17	Seberapa sering Anda berusaha mengurangi waktu untuk <i>online</i> tapi gagal?
18	Seberapa sering Anda merahasiakan sudah berapa lama Anda <i>online</i> ?
19	Seberapa sering Anda lebih memilih <i>online</i> daripada keluar dengan teman-teman?
20	Seberapa sering Anda merasa tertekan, tidak bersemangat, atau cemas ketika <i>offline</i> , dan rasa itu hilang begitu Anda <i>online</i> lagi?

Dari total 20 pertanyaan tersebut, kuisisioner IAT setiap pertanyaan memiliki rentang nilai antara 0 hingga 5. Penilaian keseluruhan diperoleh dengan menjumlahkan nilai-nilai tersebut. Pengelompokan nilai total dilakukan berdasarkan *class* kriteria interpretasi hasil terdapat pada Tabel 2.

**Tabel 2.** Class Kriteria Intepretasi Nilai Total

Skor	Keterangan
0-30	Tingkat penggunaan internet <i>normal</i>
31-49	Kecanduan ringan ( <i>mild</i> )
50-79	Kecanduan sedang ( <i>moderate</i> )
80-100	Kecanduan parah ( <i>severe</i> )

**2.4.2 Cleaning Data**

Pada tahap ini bertujuan untuk membersihkan data dari berbagai jenis ketidakakuratan, kesalahan, atau ketidaktepatan yang dapat mempengaruhi kualitas dan keandalan data yang digunakan dalam analisis. Data yang tidak berhubungan itu sebaiknya tidak digunakan. Data awal yang terkumpul mencapai 511 pada Tabel 3 akan diproses melalui tahap *cleaning*.

Untuk proses *cleaning* data yang kosong atau data hilang dilakukan penghapusan data. Terdapat data yang kosong yaitu berjumlah 1 data pada data no 128, sehingga data yang awalnya 511 menjadi 510 data yang digunakan. Data yang sudah melalui tahapan *cleaning* terdapat pada Tabel 4.

**Tabel 3.** Kusioner IAT Sebelum Melalui Tahapan *Cleaning*

No	Pertanyaan 1	Pertanyaan 2	Pertanyaan 3	Pertanyaan 4	Pertanyaan 5	...	Pertanyaan 12	Pertanyaan 13	...	Pertanyaan 20
1	Kadang-Kadang	Tidak Pernah	Kadang-Kadang	Sangat Sering	Jarang	...	Kadang - Kadang	Tidak Pernah	...	Kadang - Kadang
2	Kadang-Kadang	Kadang-Kadang	Sering	Kadang-Kadang	Jarang	...	Selalu	Kadang - Kadang	...	Selalu
3	Kadang-Kadang	Kadang-Kadang	Sering	Jarang	Kadang - Kadang	...	Sering	Kadang - Kadang	...	Sering
4	Jarang	Jarang	Jarang	Jarang	Jarang	...	Sering	Jarang	...	Sering
...	...	...	...	...	...	...	...	...	...	...
127	Kadang-Kadang	Jarang	Jarang	Tidak Pernah	Tidak Pernah	...	Kadang - Kadang	Tidak Pernah	...	Jarang
128	Sering	Tidak Pernah	Jarang	Sering	Kadang - Kadang	...	Jarang	-	...	Tidak Pernah
129	Kadang-Kadang	Tidak Pernah	Kadang-Kadang	Jarang	Kadang - Kadang	...	Kadang - Kadang	Sering	...	Sering
...	...	...	...	...	...	...	...	...	...	...
510	Kadang - Kadang	Kadang - Kadang	Jarang	Jarang	Jarang	...	Kadang - Kadang	Jarang	...	Kadang - Kadang
511	Selalu	Sering	Selalu	Selalu	Selalu	...	Sering	Selalu	...	Tidak Pernah

**Tabel 4.** Kusioner IAT Sebelum Melalui Tahapan *Cleaning*

No	Pertanyaan 1	Pertanyaan 2	Pertanyaan 3	Pertanyaan 4	Pertanyaan 5	...	Pertanyaan 12	Pertanyaan 13	...	Pertanyaan 20
1	Kadang-Kadang	Tidak Pernah	Kadang-Kadang	Sangat Sering	Jarang	...	Kadang - Kadang	Tidak Pernah	...	Kadang - Kadang

2	Kadang-Kadang	Kadang-Kadang	Sering	Kadang-Kadang	Jarang	...	Selalu	Kadang - Kadang	...	Selalu
3	Kadang-Kadang	Kadang-Kadang	Sering	Jarang	Kadang - Kadang	...	Sering	Kadang - Kadang	...	Sering
4	Jarang	Jarang	Jarang	Jarang	Jarang	...	Sering	Jarang	...	Sering
...	...	...	...	...	...	...	...	...	...	...
127	Kadang-Kadang	Jarang	Jarang	Tidak Pernah	Tidak Pernah	...	Kadang - Kadang	Tidak Pernah	...	Jarang
128	Kadang-Kadang	Tidak Pernah	Kadang-Kadang	Jarang	Kadang - Kadang	...	Kadang - Kadang	Sering	...	Sering
...	...	...	...	...	...	...	...	...	...	...
511	Selalu	Sering	Selalu	Selalu	Selalu	...	Sering	Selalu	...	Tidak Pernah

### 2.4.3 Transformation Data

Tahap ini meningkatkan kualitas data atau menyesuaikannya dengan kebutuhan, serta mengubah format data dari aslinya ke format lain. Tujuannya meningkatkan interpretasi, kualitas, atau keterbacaan data. Dalam proses metode *Naïve Bayes* data yang bisa diolah hanya informasi data yang berupa angka sedangkan data yang ada pada atribut pertanyaan berupa text sehingga data harus diubah kedalam angka sesuai dengan pembobotan kriteria memiliki skala penilaian yaitu nilai 0 sampai 5, Untuk dapat melanjutkan ke tahap proses *Naïve Bayes* seperti yang tercantum dalam Tabel 5.

**Tabel 5.** Pembobotan Nilai

Kuisisioner	Keterangan	Bobot Nilai
TP	Tidak pernah	0
J	Jarang	1
KK	Kadang-kadang	2
S	Sering	3
SS	Sangat sering	4
S	Selalu	5

Berikut ini merupakan data yang telah dilakukan *transformation data* sesuai pembobotan nilai dan menentukan jumlah setiap total nilai dari semua atribut untuk setiap individu. Sehingga mendapatkan pemilihan kategori berdasarkan nilai total dari semua atribut dalam baris tersebut yang terdapat pada Tabel 6.

**Tabel 6.** Transformation Data

No	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	...	P20	Jumlah	Range
1	2	0	2	4	1	1	0	1	1	3	...	4	38	Mild
2	2	2	3	2	1	2	1	2	0	2	...	1	37	Mild
3	2	2	3	1	2	2	3	1	2	2	...	3	44	Mild
4	1	1	1	1	1	1	2	3	1	1	...	3	32	Mild
5	5	3	2	1	1	3	5	4	1	4	...	5	63	Moderate
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

507	2	1	1	0	0	0	0	0	0	2	...	0	8	Normal
508	3	2	1	1	2	1	1	1	0	1	...	0	25	Normal
509	2	2	1	1	1	1	2	1	1	2	...	2	32	Mild
510	5	3	5	5	5	3	3	5	5	5	...	0	72	Moderate

**2.5. Analisa Metode**

Dalam penelitian ini menggunakan algoritma metode *Naïve Bayes* sebuah algoritma yang mengklasifikasikan data berdasarkan probabilitas terjadinya di masa depan yang diprediksi. Pengklasifikasi algoritma *Naïve Bayes* merupakan sebuah teknik klasifikasi yang dapat digunakan untuk menghitung tingkat kemungkinan keanggotaan suatu kelas [17]. *Naïve Bayes* juga sebuah Algoritma pembelajaran mesin probabilistik menggunakan *Teorema Bayes* untuk membuat prediksi dengan probabilitas [18]. Untuk menyelesaikan bentuk umum Persamaan untuk *Teorema Naïve Bayes* dapat dinyatakan dengan rumus sebagai berikut:

$$P(A/B) = \frac{P(B|A)*P(A)}{P(B)} \tag{1}$$

Rumus untuk menghitung probabilitas kejadian A ketika kejadian B telah terjadi. Ini dinyatakan sebagai P(A|B) dan dihitung dengan mengalikan probabilitas keadaan B ketika kejadian A telah terjadi (P(B|A)) dengan probabilitas terjadinya A (P(A)), dan kemudian dibagi dengan probabilitas terjadinya B (P(B)). Dalam konteks ini, P(A|B) menggambarkan seberapa mungkin kejadian A terjadi setelah kejadian B terjadi. Sehingga membantu dalam memahami hubungan antara dua kejadian yang mungkin saling mempengaruhi dengan memperhitungkan kemungkinan masing-masing terjadi.

Langkah-langkah menggunakan metode *Naïve Bayes* melibatkan berikut ini:

- 1) Membaca dan pengumpulan data *training*.
- 2) Perhitungan probabilitas (P(Xk|Ci)) untuk setiap kelas.
- 3) Perhitungan probabilitas (P(Xk|Ci)) kemunculan setiap nilai untuk tiap atribut. Untuk menghitung probabilitas tersebut setiap atribut Kuisisioner IAT dilakukan perhitungan 1 sampai 20 pertanyaan dalam kuisisioner IAT. Perhitungan pertanyaan 1 Kuisisioner terdapat pada Tabel 7.
- 4) Perhitungan probabilitas akhir untuk setiap kelas dengan melakukan perkalian dari probabilitas akhir fitur yang berhubungan dengan kelas yang sama.

**Tabel 7.** Tahapan Perhitungan Probabilitas Pada Pertanyaan 1

Pertanyaan 1	Normal	Mild	Moderate	Severe	P (Normal)	P (Mild)	P (Moderate)	P (Severe)
Tidak Pernah	18	1	0	0	0,08035714 3	0,00523560 2	0	0
Jarang	46	12	1	0	0,20535714 3	0,06282722 5	0,02439024 4	0
Kadang-kadang	88	60	1	0	0,39285714 3	0,31413612 6	0,02439024 4	0
Sering	63	80	19	0	0,28125	0,41884816 8	0,46341463 4	0
Sangat sering	9	26	15	0	0,04017857 1	0,13612565 4	0,36585365 9	0
Selalu	0	12	5	3	0	0,06282722 5	0,12195122	1
Jumlah	224	191	41	3				

## 2.6 Implementasi Metode

Tahap ini melibatkan menerapkan metode ke dalam kode menggunakan bahasa pemrograman Python di Google Colab.

### 2.6.1 Uji Coba

Tahap ini melakukan uji coba model klasifikasi menggunakan *Cross-validation* sebuah teknik validasi model yang dipakai untuk mengevaluasi kemampuan generalisasi hasil statistik terhadap data yang tidak tergantung. Salah satu metodenya adalah *K-fold Cross Validation*, yang melibatkan pembagian data menjadi K bagian dengan ukuran yang serupa [19]. Seleksi model terbaik sangat disarankan menggunakan *K-fold* dengan nilai K = 10 karena cenderung memberikan estimasi akurasi yang lebih baik dan dapat diukur dengan baik. Menentukan akurasi bisa dilakukan dengan cara kerja *K-fold cross validation*, total jumlah sampel dibagi menjadi N bagian. Pada *fold* ke-1, bagian ke-1 dijadikan data uji, sedangkan sisanya dijadikan data latih. Pada *fold* ke-2, bagian ke-2 digunakan sebagai data uji sementara bagian lainnya digunakan sebagai data latih. Proses selanjutnya adalah menghitung akurasi berdasarkan porsi data tersebut, Setiap *fold* berikutnya dilakukan serupa hingga mencapai *fold* ke-K. Dalam setiap iterasi, bagian ke-K diambil sebagai data uji, sementara yang lainnya digunakan sebagai data latih. Akurasi dihitung untuk setiap porsi data yang terlibat. Untuk itu rumus *k-fold cross-validation* sebagai berikut:

$$\text{Akurasi} = \frac{\text{klasifikasi benar}}{\text{data uji}} \times 100\% \quad (2)$$

### 2.6.2 Evaluasi

Pada tahap evaluasi ini, dilakukan pengecekan terhadap nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Akurasi mengevaluasi kinerja yang memperhatikan seberapa dekat prediksi dengan nilai sebenarnya [20]. Adapun persamaannya sebagai berikut:

1. *Accuracy* adalah ukuran keseluruhan seberapa banyak model berhasil mengklasifikasikan dengan tepat. Ini bisa dihitung dengan menggunakan persamaan 3.

$$\text{Accuracy} = \frac{N_{\text{benar}}}{N} \times 100\% \quad (3)$$

2. *Precision* mengukur seberapa sering model melakukan prediksi positif yang benar saat melakukan prediksi. Ini bisa dihitung menggunakan persamaan 4.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

3. *Recall* mengukur seberapa sering model memprediksi positif saat kelas aktualnya juga positif. Rumus dapat diekspresikan menggunakan persamaan 5.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

4. *F1-score* adalah hasil rata-rata gabungan dari *presisi* dan *recall*. Rumus dapat dinyatakan melalui persamaan 6.

$$\text{F1 Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (6)$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Implementasi Hasil Algoritma Naïve Bayes Menggunakan Python

Dalam menjalankan algoritma *Naïve Bayes* menggunakan bahasa pemrograman Python di Google Colab menggunakan beberapa pustaka dan modul Python. Terdapat di Gambar 2.



```

import numpy as np # For numerical computing and array manipulation.
import pandas as pd # For data manipulation and analysis.
import seaborn as sns # For statistical data visualization.
import matplotlib.pyplot as plt # For creating plots and visualizations.
from sklearn.model_selection import KFold # For model evaluation and validation.
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix # For evaluating model performance.
from sklearn.model_selection import train_test_split # For splitting data into training and testing sets.
from sklearn.preprocessing import StandardScaler # For robust feature scaling.
from sklearn.naive_bayes import GaussianNB # For Gaussian Naive Bayes classification.
    
```

Gambar 2. Modul Python

Untuk Gambar 2, terdapat pemodelan data seperti *NumPy*, *Pandas*, *Seaborn*, *Matplotlib.pyplot*, *sklearn*. Pustaka dan modul yang digunakan secara komprehensif mendukung analisis dan pemodelan data dalam kode tersebut. Selanjutnya menampilkan beberapa baris pertama dari dataset. Yang terdapat di Gambar 3.

▼ Exploratory data analysis

Sekarang, eksplorasi data untuk mendapatkan wawasan tentang data tersebut.

0d # preview the dataset  
data.head()

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	...	P13	P14	P15	P16	P17	P18	P19	P20	JUMLAH	RANGE
0	2	0	2	4	1	1	0	1	1	3	...	0	4	2	2	2	3	2	4	38	MILD
1	2	2	3	2	1	2	1	2	0	2	...	2	2	0	3	2	1	2	1	37	MILD
2	2	2	3	1	2	2	3	1	2	2	...	2	2	0	3	3	2	3	3	44	MILD
3	1	1	1	1	1	1	2	3	1	1	...	1	2	2	3	2	3	0	3	32	MILD
4	5	3	2	1	1	3	5	4	1	4	...	1	5	3	4	5	5	3	5	63	MODERATE

5 rows x 22 columns

Gambar 3. Hasil Pratinjau Data Awal

Di Gambar 3, Data yang sudah di *transformation* disimpan dalam 20 variabel data dengan total 510 data. Sehingga memberikan gambaran awal tentang struktur dan isinya. Pada penelitian ini terdapat kelas *normal* berjumlah 238 data, *mild* berjumlah 219 data, *moderate* berjumlah 50 data, dan *Severe* berjumlah 3 data sehingga total data keseluruhan berjumlah 510 data yang diklasifikasikan dengan metode *Naïve Bayes*. Pengujian dilakukan menggunakan validasi silang *K-fold*, model yang dibuat menggunakan teknik *K-Fold Cross Validation* memberikan nilai *accuracy*, *precision*, *recall* dan *F1-score* pada setiap lipatan. Proses pengolahan data dilakukan melalui bahasa pemrograman *Python* di *Google Colab* terdapat di Gambar 4.

```

+ Kode + Teks
✓ 0d from sklearn.metrics import classification_report


# Iterate over folds
for fold, result in enumerate(fold_results, 1):
    # Train classifier
    classifier = GaussianNB()
    classifier.fit(result['X_train'], result['y_train'])

    # Predict on test set
    y_pred = classifier.predict(result['X_test'])

    # Generate classification report
    report = classification_report(result['y_test'], y_pred)

    print(f"Fold {fold} Classification Report:")
    print(report)

```

 Fold 1 Classification Report:

	precision	recall	f1-score	support
MILD	0.86	0.89	0.88	28
MODERATE	0.82	1.00	0.90	9
NORMAL	0.91	0.71	0.80	14
accuracy			0.86	51
macro avg	0.86	0.87	0.86	51
weighted avg	0.87	0.86	0.86	51

**Gambar 4.** Hasil Pengolahan Data dengan *10-Fold Cross Validation*

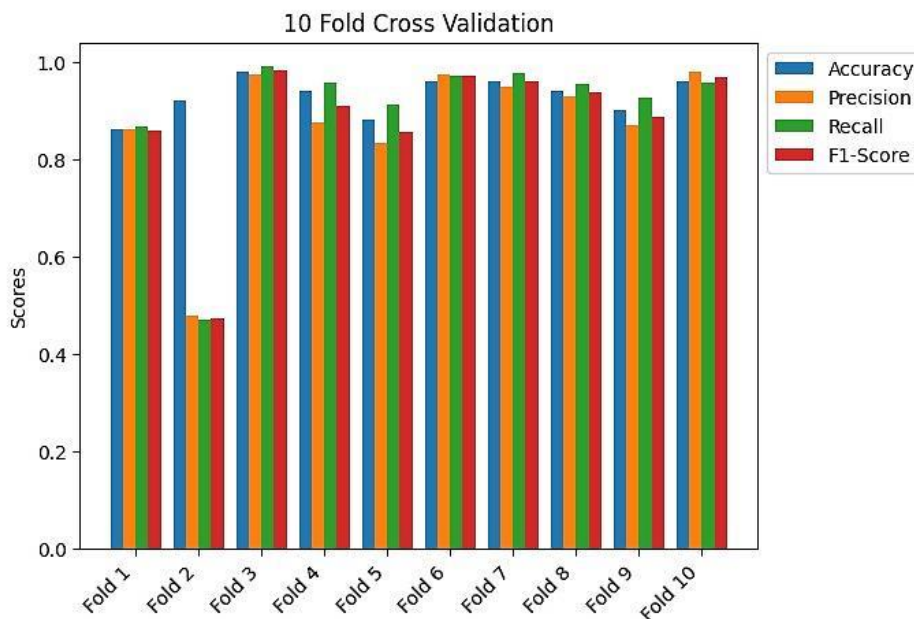
Di Gambar 4, menjelaskan langkah-langkah pembuatan model *Naïve Bayes*, penerapannya pada data, dan evaluasi kinerja melalui penilaian *accuracy*, *precision*, *recall* dan *F1-score*. Model ini dilakukan dengan *10-Fold Cross Validation* dengan mengetahui dari lipatan *fold 1* sampai *fold 10*, lipatan pertama, model dilatih menggunakan 459 data dan diuji pada 51 data lainnya, menghasilkan akurasi sebesar 0.86. Begitu pula dengan lipatan berikutnya, di mana model dilatih pada subset data latih dan diuji pada subset data uji yang berbeda. Tujuan evaluasi ini dilakukan menguji efektivitas algoritma *Naïve Bayes Classifier* dalam mengklasifikasikan tingkat kecanduan internet terhadap remaja di Pekanbaru. Hasil pengujian ini kemudian terdapat dalam Tabel 8.

**Tabel 8.** Hasil *10 Fold Cross Validation*

Nilai <i>K-fold</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Fold 1	86%	86%	87%	86%
Fold 2	92%	48%	47%	47%
Fold 3	98%	98%	99%	98%
Fold 4	94%	88%	96%	91%
Fold 5	88%	83%	91%	86%
Fold 6	96%	97%	97%	97%
Fold 7	96%	95%	98%	96%
Fold 8	94%	93%	95%	94%
Fold 9	90%	87%	93%	89%
Fold 10	96%	98%	96%	97%

Pada Tabel 8, terdapat akurasi tertinggi yaitu pada fold ke-3 dengan nilai 98%. Pada *fold* ke-3, memiliki nilai *precision*, *recall*, dan *f1-score* adalah 98%, 99%, dan 98%. Untuk nilai akurasi terendah yaitu pada *fold* ke-1 dengan nilai 86%. Pada *fold* ke-1 memiliki nilai *precision*, *recall*, dan *f1-score* adalah 86%, 87%, dan 86%. Performa rata-rata yang diperoleh melalui hasil *10-fold Cross Validation* menunjukkan bahwa nilai *accuracy*, *precision*, *recall*, dan *f1-score* adalah 93%, 87,3%, 89,9%, dan 88,1%. Untuk memudahkan pemahaman data yang terdapat dalam Tabel 8, informasi tersebut dapat

disajikan dalam bentuk *bar chart* grafik yang terlihat di Gambar 5.



**Gambar 5.** Grafik Kinerja Model Hasil 10 *Fold Cross Validation*

Pada Gambar 5, menjelaskan akurasi dievaluasi menggunakan validasi silang 10 kali lipat (rata-rata dan standar deviasi) untuk setiap *fold* nya sesuai dengan nilai *accuracy*, *presicion*, *recall*, dan *F1-Score*. Model algoritme dilatih dan diuji menggunakan semua fitur yang tersedia, dan pengaturan klasifikasi disesuaikan secara optimal untuk setiap tugas.

### 3.2 Pembahasan

Penelitian ini berfokus pada klasifikasi tingkat kecanduan remaja Pekanbaru terhadap penggunaan internet. Data yang digunakan pada penelitian ini adalah data primer yang diambil dari remaja Pekanbaru yang berdasarkan responden kuisioner tentang kecanduan internet. Metode algoritma yang digunakan dalam penelitian ini adalah algoritma *Naive Bayes* yang merupakan teknik pembelajaran mesin berbasis probabilitas.

Data dikumpulkan melalui pengumpulan kuisioner penelitian ini menggunakan IAT Dr.Young dengan class interpretasi hasil yaitu *normal*, *mild*, *moderate*, *severe*. Data tersebut kemudian diolah dan dianalisis menggunakan algoritma *Naive Bayes* untuk mengklasifikasikan kecanduan internet terhadap remaja Pekanbaru ke dalam kelompok dengan tingkat kecanduan yang sesuai berdasarkan pola yang terdapat pada data tersebut. Untuk tahapan pertamanya yaitu dengan identifikasi masalah, mempelajari literatur, pengumpulan data untuk mencari informasi yang relevan dengan isu penelitian yang sedang difokuskan pada tahap penelitian, melakukan preprocessing data yang melibatkan tahapan (*selection data*, *cleaning data*, *transformation data*), lalu analisa metode, implementasi metode, uji coba dan evaluasi, lalu menyimpulkan hasil penelitian.

Adapun Tujuan dari penelitian ini yaitu mengklasifikasi tingkat kecanduan internet terhadap remaja Pekanbaru melalui pendekatan algoritme *Naive Bayes*. Manfaat penelitian ini diharapkan akan memberikan nilai akurasi untuk klasifikasi tingkat kecanduan internet terhadap remaja pekanbaru melalui pendekatan algoritme *Naive Bayes*.

Pada penelitian ini berbeda dengan penelitian sebelumnya. Penggunaan kuesioner IAT Internet Young sebagai parameter memberikan pendekatan yang lebih terstruktur dan spesifik dalam mengidentifikasi tingkat kecanduan internet. Ini berbeda dengan penelitian lain yang mungkin menggunakan parameter yang lebih umum atau berbeda dalam konteksnya. Hasil rata-rata akurasi yang dilakukan dengan pengujian *10-fold Cross Validation* sebesar 93% yang dilakukan menggunakan

bahasa pemrograman Python di Google Colab yang menunjukkan bahwa kuesioner IAT adalah alat yang efektif dalam mengukur kecanduan internet dan bahwa algoritme *Naïve Bayes* dapat mengklasifikasikan class yaitu *normal, mild, moderate, severe*.

#### 4. KESIMPULAN

Berdasarkan penelitian yang dilakukan, kesimpulan sebagai berikut dapat diambil:

1. Klasifikasi tingkat kecanduan internet pada remaja di Pekanbaru menggunakan algoritme *Naïve Bayes* dan pengujian 10-Fold Cross Validation dengan 459 data latih dan 51 data uji.
2. Akurasi tertinggi tercapai pada *fold* ke-3 dengan nilai 98%, dengan *precision, recall, dan f1-score* masing-masing 98%, 99%, dan 98%. Akurasi terendah pada *fold* ke-1 adalah 86%, dengan *precision, recall, dan f1-score* masing-masing 86%, 87%, dan 86%.
3. Rata-rata performa dari *10-fold Cross Validation* menunjukkan nilai *accuracy* 93%, *precision* 87,3%, *recall* 89,9%, dan *f1-score* 88,1%. Sehingga rata-rata akurasi sebesar 93%.

Oleh karena itu untuk pengembangan penelitian selanjutnya peneliti memberikan saran menggunakan metode alternatif lain untuk klasifikasi guna memperoleh perspektif dan hasil akurasi yang berbeda dalam mengukur tingkat kecanduan internet pada remaja di Pekanbaru.

#### Daftar Pustaka

- [1] P. Anggarini, M. Manangkot, and O. A. Kamayani, "Hubungan kecanduan internet dengan kecerdasan emosional pada remaja," *J. Ilmu Keperawatan Jiwa*, vol. 5, no. 2, pp. 381–394, 2022, [Online]. Available: <https://journal.ppnijateng.org/index.php/jikj>
- [2] M. A. S. Putri and D. Suryadi, "Hubungan Antara Adiksi Internet Dengan Prokrastinasi Akademik Pada Siswa Smkn X Indramayu," *Provitae J. Psikol. Pendidik.*, vol. 16, no. 1, pp. 22–31, 2023, doi: 10.24912/provitae.v16i1.23539.
- [3] L. Marini\*, W. Hendriani, and P. Y. Wulandari, "Optimizing the Role of the Family in Prevention Teen Smartphone Addiction through Psychoeducation Smartphone Addiction to Parents," *Din. J. Pengabd. Kpd. Masy.*, vol. 7, no. 4, pp. 930–941, 2023, doi: 10.31849/dinamisia.v7i4.15217.
- [4] P. Hidayat, "Interaksi Sosial Online Dan Kecemasan Sosial Sebagai Prediktor Kecanduan Internet Pada Remaja," *J. Ilm. Psyche*, vol. 14, no. 2, pp. 83–92, 2020, doi: 10.33557/jpsyche.v14i2.1159.
- [5] L. U. Nadzifah, K. Bariyyah, and D. Permatasari, "Pengembangan inventori kecanduan internet untuk siswa sekolah menengah atas," *JPPI (Jurnal Penelit. Pendidik. Indones.)*, vol. 6, no. 2, p. 77, 2020, doi: 10.29210/02020494.
- [6] F. Fernandes, A. Y. Sari, and M. Mahathir, "Faktor-Faktor yang Berhubungan dengan Kecanduan Internet pada Siswa SMA N 'X' Padang," *NERS J. Keperawatan*, vol. 17, no. 1, p. 1, 2021, doi: 10.25077/njk.17.1.1-13.2021.
- [7] Y. Syafitri, R. Astika, and L. S. E. Esti Rahayu, "Pengembangan Aplikasi Pelanggaran Menggunakan Framework Codeigniter Berbasis Web," *J. Inf. dan Komput.*, vol. 10, no. 2, pp. 01–07, 2022, doi: 10.35959/jik.v10i2.381.
- [8] Y. Astuti, I. R. Wulandari, A. R. Putra, and N. Kharomadhona, "Naïve Bayes untuk Prediksi Tingkat Pemahaman Kuliah Online Terhadap Mata Kuliah Algoritma Struktur Data," *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 1, p. 28, 2022, doi: 10.26418/jp.v8i1.48848.
- [9] E. P. Saputra, M. Maulidah, N. Hidayati, and A. Saryoko, "Komparasi Evaluasi Kinerja Siswa Belajar dengan Menggunakan Algoritma Machine Learning," *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2239, 2022, doi: 10.30865/mib.v6i4.4786.
- [10] M. W. A. Putra, Susanti, Erlin, and Herwin, "Analisis Sentimen Dompok Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 72–86, 2020, doi: 10.25299/itjrd.2020.vol5(1).5159.
- [11] N. Asmiati, S. Informasi, S. Tinggi, M. Informatika, D. Komputer, and N. Mandiri, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Pengaruh Negatif Game Online Bagi Remaja Milenial (Application of the Naive Bayes Algorithm to Classify the Negative Effects of Online Games on Millennial Adolescents)," *J. Teknol. Inf. dan Multimed.*, vol. 2, no. 3, pp. 141–149, 2020.

- [12] N. Nurainun, E. Haerani, F. Syafria, and L. Oktavia, "Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation," *J. Comput. Syst. Informatics*, vol. 4, no. 3, pp. 578–586, 2023, doi: 10.47065/josyc.v4i3.3414.
- [13] B. Purba and R. Syahputra, "Implementasi metode Naive Bayes Classifier pada Evaluasi Kepuasan Mahasiswa terhadap Pembelajaran Daring," *InfoTekjar J. Nas. Inform. dan Teknol. Jar.*, vol. 6, no. 1, pp. 85–91, 2021, [Online]. Available: <https://doi.org/10.30743/infotekjar.v6i1.4352>
- [14] M. Rosanensi *et al.*, "Analisis Tingkat Kecanduan Game Online Terhadap Manajemen Waktu Dengan Menggunakan Metode Certainty Factor Dan Algoritma K-Nearest Neighbor Lombok Tengah," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 7, no. 2, pp. 473–482, 2023, doi: 10.52362/jisamar.v7i2.1040.
- [15] G. A. P. Ariani, S. S. Papuke, and R. Apriana, "Early Detection of Internet Addiction in Junior High School Teens in Gorontalo City," *Jambura J. Heal. Sci. Res.*, vol. 4, no. 3, pp. 784–797, 2022, doi: 10.35971/jjhsr.v4i3.13617.
- [16] R. Rismala, L. Novamizanti, K. N. Ramadhani, Y. S. Rohmah, S. Parjuangan, and D. Mahayana, "Kajian Ilmiah dan Deteksi Adiksi Internet dan Media Sosial di Indonesia Menggunakan XGBoost," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 1, p. 1, 2021, doi: 10.26418/jp.v7i1.43606.
- [17] E. Darnila, M. Maryana, and M. Azmi, "Aplikasi Klasifikasi Status Gizi Balita Menggunakan Metode Naïve Bayes Berbasis Android," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 5, no. 2, pp. 135–141, 2021, doi: 10.46880/jmika.vol5no2.pp135-141.
- [18] W. A. Firmansyah, U. Hayati, and Y. Arie Wijaya, "Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.
- [19] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [20] R. R. Adhitya, Wina Witanti, and Rezki Yuniarti, "Perbandingan Metode Cart Dan Naïve Bayes Untuk Klasifikasi Customer Churn," *INFOTECH J.*, vol. 9, no. 2, pp. 307–318, 2023, doi: 10.31949/infotech.v9i2.5641.



ZONAsi: Jurnal Sistem Informasi

is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)