



# Analisis Sentimen Terhadap Kesehatan Mental Remaja Menggunakan Metode Naive Bayes

Pii Syahputra\*, Rakhmat Kurniawan

Sains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Deli Serdang  
Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara, Indonesia

Email: <sup>1,\*</sup>piisyahputra1711@gmail.com, <sup>2</sup>rakhmat.kr@uinsu.ac.id

Email Penulis Korespondensi: piisyahputra1711@gmail.com

Submitted: 20/07/2024; Accepted: 29/07/2024; Published: 30/07/2024

**Abstrak**—Kehidupan sehari-hari, khususnya bagi remaja sekarang melibatkan teknologi internet. Twitter (sekarang dikenal sebagai X) adalah salah satu platform media sosial yang paling populer. Analisis sentimen data media sosial dapat meningkatkan pemahaman masyarakat tentang masalah kesehatan mental. Penelitian ini menggunakan algoritma Naive Bayes untuk menganalisis sentimen pengguna media sosial X terhadap kesehatan mental remaja. Tujuan lain dari penelitian ini adalah untuk mengukur seberapa efektif dan akurat teknik tersebut dalam mengidentifikasi sentimen dan menyajikan hasil analisis dalam bentuk cloud kata dan grafik. Data dikumpulkan dari awal tahun 2024 hingga saat ini dari tweet dengan hashtag Mental Health. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memiliki tingkat akurasi yang efektif dalam klasifikasi sentimen terhadap kesehatan menggunakan kamus InSetLexicon. Proses preprocessing data juga termasuk pembersihan, tokenizing, normalisasi, stowords, dan stemming. Selain itu, evaluasi kinerja dilakukan dengan menggunakan confusion matrix untuk menghitung precision, recall, F-1 Score, dan accuracy. Hasil klasifikasi yang diperoleh mendapatkan accuracy sebesar 0.8049792531120332 atau sekitar 80%, precision sebesar 83%, recall didapatkan 68% dan F-1 Score sebesar 74.9%.

**Kata Kunci:** Analisis Sentimen; Kesehatan Mental Remaja; Algoritma Naive Bayes; Teks Mining

**Abstract**—Everyday life, especially for teenagers, now involves internet technology. Twitter (now known as X) is one of the most popular social media platforms. Sentiment analysis of social media data can improve people's understanding of mental health problems. This research uses the Naive Bayes algorithm to analyze the sentiments of X social media users regarding adolescent mental health. Another goal of this research is to measure how effective and accurate the technique is in identifying sentiment and presenting analysis results in the form of word clouds and graphs. Data was collected from the beginning of 2024 to the present from tweets with the hashtag Mental Health. The research results show that the Naive Bayes algorithm has an effective level of accuracy in classifying sentiment towards health using the InSetLexicon dictionary. The data preprocessing process also includes cleaning, tokenizing, normalization, stockwords, and stemming. In addition, performance evaluation is carried out using confusion matrix to calculate precision, recall, F-1 Score, and accuracy. The classification results obtained an accuracy of 0.8049792531120332 or around 80%, precision of 83%, recall of 68% and F-1 Score of 74.9%.

**Keywords:** Sentiment Analysis; Adolescent Mental Health; Naive Bayes Algorithm; Text Mining

## 1. PENDAHULUAN

Saat ini teknologi internet sangatlah banyak digemari oleh seluruh kalangan khususnya para remaja, mereka menggunakan internet untuk beberapa kegiatan sehari-hari, mulai dari mengerjakan tugas sekolah atau kuliah hingga untuk dunia usaha [1]. Untuk dunia usaha internet merupakan suatu hal yang sangat dibutuhkan saat ini. Komunikasi antar karyawan dan berbagai kegiatan usaha sudah banyak menggunakan teknologi internet seperti jualan online. Perkembangan teknologi internet yang begitu pesat, kini banyak melahirkan berbagai jenis media sosial [2]. Hal ini didukung juga oleh perkembangan teknologi telepon seluler yang menyediakan fasilitas bagi para pengguna media sosial. Media sosial ini sangat populer karena memberikan kemudahan dalam melakukan komunikasi kepada sesama pengguna media sosial. Media sosial yang paling populer digunakan para remaja antara lain, Facebook, Twitter yang sudah berganti nama menjadi X, Instagram, Whatsapp, You tube dan masih banyak lagi [3].

Salah satu media sosial yang banyak digunakan oleh masyarakat adalah media sosial X. Berdasarkan buku yang ditulis oleh Hadi (2010 : 2) pengertian Twitter atau X adalah situs microblog yang memberikan fasilitas pengguna untuk mengirim sebuah teks dengan panjang maksimal 140 karakter melalui SMS, pengirim pesan instan, surat elektronik [4]. Twitter juga termasuk situs microblog terbaik didunia, Twitter atau X termasuk media sosial yang populer dikalangan remaja, karena mayoritas pengguna X adalah kalangan remaja, karena dari media sosial X pengguna dapat mengikuti orang, mendapatkan pengikut, berkomunikasi dan juga mengakses informasi dengan mudah [5].

Media sosial X juga membuat fitur pengguna untuk dapat memberikan pendapatnya dengan cara mempublikasikan tweet mereka atau sekedar memberikan komentar, dengan demikian para pengguna dapat

membahas apapun mengenai informasi maupun isu yang sedang diperbincangkan [6]. Seperti halnya mengenai kesehatan mental remaja yang pernah menjadi isu di X, banyaknya kasus bunuh diri, yang mana menurut data pusiknas polri sudah terjadi sebanyak 287 kasus bunuh diri pada tahun 2024 ini saja. Kekerasan seksual, kekerasan fisik maupun verbal, bullying dan juga hal-hal lainnya yang dapat mengakibatkan gangguan kejiwaan. Kesehatan mental dipengaruhi oleh beberapa peristiwa dalam kehidupan yang meninggalkan dampak yang sangat besar pada kepribadian dan perilaku seseorang [7]. Peristiwa tersebut dapat berupa kekerasan pada rumah tangga atau keluarga, pelecehan seksual, atau stres berat jangka panjang.

Berdasarkan dari data-data komentar dan juga tweet pengguna twitter berdasarkan dari pendapat masyarakat yang menggunakan twitter, dapat dilakukan analisis sentiment guna untuk menyaring tweet ataupun komentar yang terkait dengan isu kesehatan mental, yang nantinya akan dikelompokkan menjadi komentar negative, positif. Melakukan sentiment analis adalah untuk menganalisis, komentar, emosi, dan sikap dari setiap tweet yang diberikan mengenai kesehatan mental [8]. Dengan melakukan sentimen terhadap isu kesehatan mental ini, maka akan diketahui tanggapan apa yang paling dominan pada isu kesehatan mental ini, apakah masyarakat sudah paham mengenai kesehatan mental atau sebaliknya [9].

Dalam melakukan analisis sentimen, peneliti menggunakan algoritma naïve bayes dalam proses mengklasifikasi data twitter mengenai kesehatan mental. Peneliti memilih algoritma Naïve Bayes karena berdasarkan penelitian Dian Prajarini tentang klasifikasi data mining untuk prediksi penyakit kulit menggunakan algoritma C4.5, Naïve Bayes, KNN, SVM. Dari hasil menggunakan rasio 70:30, C4.5 bernilai 93.6%, Naive Bayes bernilai 98.2%, KNN bernilai 95.5%, dan SVM 98.2%. Jika dibandingkan berdasarkan nilai, maka Naïve Bayes dianggap sebagai algoritma klasifikasi yang lebih baik, tetapi jika dibandingkan secara global, maka Naïve Bayes dan SVM dianggap algoritma yang lebih baik dibandingkan dengan algoritma C4.5 dan juga KNN untuk prediksi penyakit kulit [10]. Berdasarkan dari penelitian tersebut menunjukkan bahwa algoritma naïve bayes memiliki nilai akurasi yang baik dibandingkan dengan algoritma yang lain, sehingga peneliti menggunakan algoritma naïve bayes dalam melakukan analisis sentimen media sosial X terhadap kesehatan mental [11].

## 2. METODOLOGI PENELITIAN

### 2.1 Kerangka Penelitian

Kerangka penelitian merupakan tahapan-tahapan yang akan diterapkan pada saat melakukan penelitian. Hal ini dilakukan agar tahapan penelitian terstruktur [12].

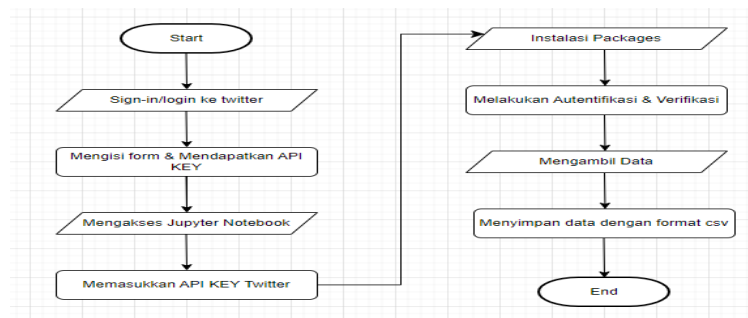


Gambar 1. Kerangka Penelitian

Gambar 1 adalah kerangka penelitian. Peneliti menggunakan model konseptual, yaitu suatu kerangka kerja konseptual, sistem atau skema yang menerapkan tentang serangkaian ide global tentang keterlibatan individu, kelompok, situasi atau kejadian, terhadap suatu ilmu dan pengembangannya. Penelitian yang penulis lakukan ini akan menghasilkan sebuah informasi mengenai kesehatan mental remaja atau mental health di Indonesia dengan menentukan atribut yang dibutuhkan dalam melakukan proses pengolahan data ini.

### 2.2 Pengumpulan Data

Pengumpulan data pada penelitian ini dibagi menjadi dua tahapan yaitu tahap mendapatkan API key dan tahap mengambil data media sosial X.



Gambar 2. Flowchart Pengumpulan Data

Gambar 2 adalah flowchart proses pengumpulan data. Pada tahap ini, yang pertama dilakukan adalah membuka tools jupyter notebook yang digunakan untuk melakukan proses crawling pengambilan data dari media sosial twitter [13]. Selain itu memasukkan API key twitter yang sudah didapat dari twitter berupa consumer key, consumer access, access token dan access secret token. Selanjutnya, dilakukan instalasi packages yang digunakan untuk bisa berkomunikasi dengan API twitter, kemudian melakukan autentifikasi dan verifikasi antara twitter dengan jupyter notebook dan setelah proses autentifikasi sukses maka dapat dilakukan pencarian data sesuai dengan keyword yang diinginkan. Setelah proses pengambilan data selesai, maka data akan tersimpan ke dalam bentuk format csv [14].

### 2.3 Algoritma Naive Bayes

Metode naive bayes merupakan metode yang memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggris Thomas Bayes. Naive bayes merupakan metode pengklasifikasian yang sangat sederhana dengan mengasumsikan klasifikasi atribut. Dengan metode naive bayes terlebih dahulu mencari nilai probabilitas dan likelihood maksimum dan setiap atribut untuk masing-masing kelas [15].

Metode naive bayes ini juga sudah banyak digunakan oleh penelitian lainnya, naive bayes merupakan sebuah pengklasifikasian probabilistik dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema bayes dan mengasumsikan semua atribut independent atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas [16]. Persamaan dari teorema bayes adalah [17]:

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

Dengan keterangan, H adalah hipotesis data merupakan suatu class spesifik, X adalah data dengan class yang belum diketahui, P(H|X) adalah probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas), P(X) adalah probabilitas X, P(H) adalah probabilitas H (prior probabilitas) dan P(X|H) adalah probabilitas X berdasarkan kondisi pada hipotesis H.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Pada penelitian ini penulis melakukan proses pengumpulan data diambil dari aplikasi X (Twitter). Data komentar diambil dengan menggunakan proses scrapping. Data yang diambil dengan proses scrapping akan disimpan kedalam file berekstensi .csv. Data yang diambil berjumlah 1200 data komentar, dimana komentar membahas tentang sentiment Kesehatan mental atau komentar para pengguna aplikasi X tentang Kesehatan mental. Komentar nantinya akan diberikan label positif dan negatif menggunakan kamus InsetLexicon dan Algoritma Naive Bayes untuk mencari accuracy, precision, recall dan F1-Score.

#### 3.1.1 Daftar Komentar Pengguna Terhadap Sentimen Kesehatan Mental

**Tabel 1.** Contoh Komentar Pengguna Aplikasi X Pada Sentimen Mental Health

No.	Username	CA	FC	Full_text
1	MiftahChoer	Wed Oct 25 11:45:45 +0000 2023	0	Mengungkap 4 Kebiasaan untuk Kesehatan Mental: 1. latihan fisik 2. tidur cukup 3. ibadah rutin meditasi dan 4. melakukan kegiatan sosial. #olahraga #ibadah #meditasi #kegiatanassosial #kesehatanmental @famouspersonprofiles <a href="https://t.co/pIUgRKzKf4">https://t.co/pIUgRKzKf4</a>
2	ekingsjuara2	Wed Oct 18 01:38:13 +0000 2023	0	Assalamu'alaikum Sobat..!! Jaga kesehatan mental merupakan aspek penting dalam kehidupan kita. Cobalah beberapa tips sederhana berikut untuk merawat kesehatan mental Anda yuk simak sama-sama!! #tips #tipsmenjagakesehatanmental #kesehatanmental <a href="https://t.co/WBaBEQytqy">https://t.co/WBaBEQytqy</a>
3	Madoenur	Wed Oct 11 13:22:16 +0000 2023	0	Pemerintah harusnya menyediakan layanan konsuling melalui kemendikbud untuk mahasiswa. Kesehatan mental mungkin lebih utama dari kesehatan fisik. Jika mentalnya baik maka fisiknya juga baik. #KesehatanMental
...	....	...	...	...
1202	Bibahjenner	Thu Jul 11 03:54:30 +0000 2024	0	@iladasilfar yuk belajar egois demi kesehatan mental kita



Tabel 1 adalah contoh komentar yang digunakan. Komentar yang digunakan didapatkan dari aplikasi X (Twitter) pada bulan Juni 2024 dan diakses melalui jupyter notebook dengan scrapper data. Ulasan diambil dari data komentar pengguna aplikasi X (Twitter).

### 3.2 Text Pre-Processing

#### 3.2.1 Text Clean and Case Folding

Text cleaning bertujuan untuk membersihkan data komentar yang sudah diperoleh. Mulai dari menghapus username dari X, menghapus link, menghapus tanda baca, mengubah ke huruf kecil dan lainnya. Komponen yang dibersihkan didefinisikan sebagai bagian yang tidak penting atau tidak relevan untuk proses pengklasifikasian data. Mengubah komentar menjadi huruf kecil dan menghapus semua tautan atau emotikon [18].

#### 3.2.2 Text Tokens

Teks dapat dipecah menjadi kata-kata atau token tertentu, seperti "Kesehatan" (kesehatan), "diperlukan" (diperlukan), "untuk" (untuk), dll. Tokenisasi teks adalah proses memecah teks menjadi unit-unit individu yang disebut token. Token adalah blok bangunan dasar teks, dan dapat berupa kata, frase, karakter, atau bahkan subkata (bagian dari kata). Tokenisasi memungkinkan kita untuk menganalisis teks pada tingkat yang lebih mendalam, yang memungkinkan kita untuk mengekstrak pola dan wawasan yang berguna [19].

#### 3.2.3 Text Normalization

Tabel 2. Normalisasi Kata atau Slang Words

Sebelum	Sesudah
Woww	wow
Met	Selamat
Yg	Yang
Hallo	halo
Kk	kakak

Tabel 2. merupakan hasil normalisasi kata atau slang words. Perubahan kata seperti "yg" menjadi "yang", "ga" menjadi "tidak", dan sebagainya dapat dilakukan untuk mengubah teks menjadi bahasa Indonesia yang baik menurut KBBI. Dataset "kamusGaul.txt" yang diambil dari github digunakan untuk penelitian.

#### 3.2.4 Stopwords

Karena kata stop tidak memiliki arti khusus atau penting, mereka dapat dianggap sebagai "suara" dalam teks. Stopwords dapat menyebabkan kekeliruan dalam analisis teks atau pencarian informasi, karena mereka dapat membuat model atau algoritma menjadi bingung. Menghapus stopwords dapat mengurangi jumlah data yang perlu diproses, sehingga meningkatkan efisiensi analisis teks atau pencarian informasi [20]. Pada penelitian menggunakan dataset "stopwordbahasa.csv" yang diakses dari github. Contoh dengan menghapus kata "ada", "adalah", "adanya", "Adapun", "agak", "akhir", dan lainnya dengan jumlah 758 kata.

#### 3.2.5 Stemming

Stemming digunakan untuk mengurangi dimensi data dan meningkatkan efisiensi dalam analisis teks dan pencarian informasi dengan mengurangi kata-kata ke bentuk dasar atau akar kata, sehingga kata-kata yang berbeda tetapi memiliki makna yang sama dapat dianggap sama [21].

#### 3.2.6 Detokenized Text

Tabel 3. Detokenized Text

No.	Stemming	Detokenized
1	['ungkap', 'biasa', 'sehat', 'mental', 'latih', 'fisik', 'tidur', 'ibadah', 'rutin', 'meditasi', 'giat', 'sosial', 'olahraga', 'ibadah', 'meditasi', 'kegiatassosial', 'kesehatanmental']	ungkap biasa sehat mental latih fisik tidur ibadah rutin meditasi giat sosial olahraga ibadah meditasi kegiatassosial kesehatanmental
2	['perintah', 'sedia', 'layan', 'konsuling', 'kemendikbud', 'mahasiswa', 'sehat', 'mental', 'utama', 'sehat', 'fisik', 'mental', 'fisik', 'kesehatanmental']	perintah sedia layanan konsuling kemendikbud mahasiswa sehat mental utama sehat fisik mental fisik kesehatanmental
3	['dukung', 'sehat', 'mental', 'pegawai', 'lupa', 'konsultasi', 'ahli', 'ya', 'biotizen', 'source', 'mentalhealth', 'kesehatanmental', 'worklifebalance', 'biofarmaid']	dukung sehat mental pegawai lupa konsultasi ahli ya biotizen source mentalhealth kesehatanmental worklifebalance biofarmaid
...	...	...
1202	['yuk', 'ajar', 'egois', 'sehat', 'mental']	yuk ajar egois sehat mental

Tabel 3. merupakan detokenized text. Proses detokenized merupakan proses yang menghilangkan tanda “[’]” didalam steaming, dimana proses ini menjadi acuan untuk menjadi hasil dari pre-processing yang telah dilakukan.

### 3.3 Pemberian Label Menggunakan Inset Lexicon

Langkah selanjutnya adalah proses melabelkan data dimana dilakukan secara otomatis dengan mengimplementasikan kamus InSetLexicon. Nilai polaritas ditentukan dalam kamus InSetLexicon dengan kata-kata dan bobotnya. Untuk mengetahui nilai polaritas, anda harus menghitung semua bobot kata dalam teks ulasan. Hasil perhitungannya adalah ulasan untuk setiap elemen dikategorikan sebagai sentimen positif atau negatif. Ulasan dianggap positif jika nilai polaritasnya lebih besar dari nol, dan negatif jika nilai polaritasnya lebih rendah dari nol.

**Tabel 4.** Positive Lexicon

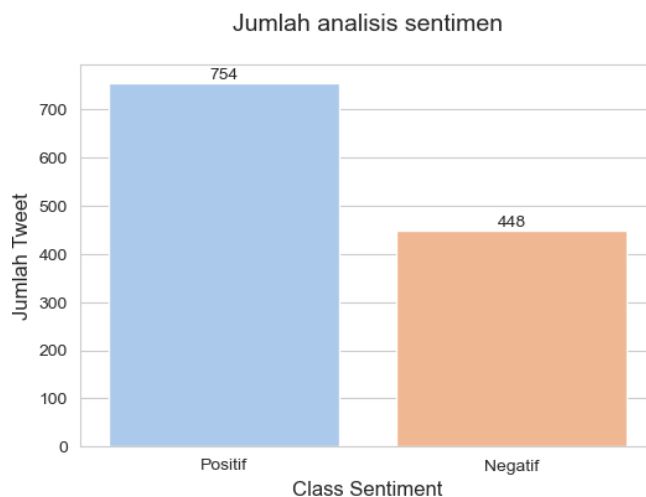
No.	Word	Weight
1	Hai	3
2	Tetap	3
3	Detail	2
4	Bagus	2
...	...	...
3610	Orisinal	3

Tabel 4 merupakan positive lexicon atau ulasan yang dianggap positif.

**Tabel 5.** Negative Lexicon

No.	Word	Weight
1	Isak	-5
2	Sakit	-5
3	Mulu	-1
4	Gamau	-4
...	...	...
6610	Mencoreng	-4

Tabel 5 merupakan negative lexicon atau ulasan yang dianggap negatif.



**Gambar 3.** Jumlah Analisis Sentimen

Gambar 3. merupakan hasil jumlah analisis sentimen. Jumlah komentar positif dan negatif pada dataset sentiment kesehatan mental yang diambil dari aplikasi X (Twitter) terdapat pada gambar 3, dimana proses dilakukan menggunakan pemrograman python di jupyter notebook.

### 3.4 Perhitungan Model Naïve Bayes

#### 3.4.1 Naïve Bayes Classification

Jumlah pada penelitian analisis sentiment mengenai Kesehatan mental menggunakan Algoritma Naïve Bayes adalah dengan total 1202 data. Data latih berjumlah 961 data, sedangkan data uji berjumlah 241 data. Proses

pelatihan akan menghasilkan bobot setiap kata pada setiap kelas dengan menggunakan metode pembobotan TF-IDF. Dengan menggunakan 2 data sampel sebelumnya sebagai data latih, maka ditentukan 1 data uji sebagai berikut.

**Tabel 6.** Sample Data Training

Latih Komentar	Label
latih yoga bukti ampuh atasi sehat mental	Negatif
biar capek ya sehat mental dijaga	Positif

Tabel 6 merupakan sample data training. Komentar dibuat dengan menghitung kemungkinan dokumen data uji menggunakan kemungkinan kata pada data latih. Algoritma Naive Bayes melakukan klasifikasi sentimen secara otomatis dengan menggunakan fungsi MultinomialNB. Tingkat probabilitas positif meningkatkan nilai sentimen yang diklasifikasi, sedangkan tingkat probabilitas negatif meningkatkan nilai sentimen yang diklasifikasi.

**Tabel 7.** Pembobotan Kata Pada Sampel Data Latih

No.	Kosa Kata	TF (+)	TF (-)
1.	Latih	1	0
2.	Yoga	1	0
3.	Bukti	1	0
4.	Ampuh	1	0
5.	Atasi	1	0
6.	Sehat	1	0
7.	Mental	1	0
8.	Biar	0	1
9.	Capek	0	1
10.	Ya	0	1
11.	Sehat	0	1
12.	Mental	0	1
15.	dijaga	0	1
<b>Jumlah Term</b>		<b>7</b>	<b>6</b>

Tabel 7 merupakan pembobotan kata pada sampel data latih. Untuk memulai klasifikasi kelas, probabilitas sebelumnya, probabilitas bersyarat, dan probabilitas posterior dihitung. Proses klasifikasi menggunakan algoritma Naive Bayes pada uji data adalah tahapan berikut.

$$\text{Perhitungan nilai probabilitas } P(\text{Kelas} \mid \text{Komentar}) = \frac{\text{Jumlah Kelas } X}{\text{Jumlah Komentar}} \tag{2}$$

Dengan menggunakan persamaan di atas, kita akan memperoleh probabilitas setiap kelas dalam sentimen.

- $P(\text{Positif} \mid \text{Komentar}) = \frac{1}{2} = 0.5$

- $P(\text{Negatif} \mid \text{Komentar}) = \frac{1}{2} = 0.5$

a. Perhitungan Nilai Probabilitas Bersyarat :

$$P(\text{Term} \mid \text{Kelas}) = \frac{\text{Total bobot TF Term pada Kelas}+1}{\text{Bobot TF Kelas}+\text{Total Bobot TF}} \tag{3}$$

Dengan menggunakan persamaan di atas, kita akan memperoleh probabilitas suku-suku di setiap kelas sentimen.

- Probabilitas pada kata “Latih”

$$P(\text{Latih} \mid \text{Positif}) = \frac{1+1}{7+13} = \frac{2}{20} = 0.1$$

$$P(\text{Latih} \mid \text{Negatif}) = \frac{0+1}{5+13} = \frac{1}{18} = 0.056$$

- Probabilitas pada kata “Yoga”

$$P(\text{Latih} \mid \text{Positif}) = \frac{1+1}{7+13} = \frac{1}{20} = 0.05$$

$$P(\text{Latih} \mid \text{Negatif}) = \frac{0+1}{5+13} = \frac{1}{18} = 0.056$$

- Probabilitas pada kata “Bukti”

$$P(\text{Latih} \mid \text{Positif}) = \frac{1+1}{7+13} = \frac{1}{20} = 0.05$$

$$P(\text{Latih} \mid \text{Negatif}) = \frac{0+1}{5+13} = \frac{1}{18} = 0.056$$

- Probabilitas pada kata “Sehat”

$$P(\text{Latih} \mid \text{Positif}) = \frac{1+1}{7+13} = \frac{1}{20} = 0.05$$

$$P(\text{Latih} \mid \text{Negatif}) = \frac{1+1}{5+13} = \frac{1}{18} = 0.056$$

- Probabilitas pada kata “Mental”

$$P(\text{Latih} | \text{Positif}) = \frac{1+1}{7+13} = \frac{1}{20} = 0.05$$

$$P(\text{Latih} | \text{Negatif}) = \frac{1+1}{5+13} = \frac{1}{18} = 0.056$$

Sebelum mencari probabilitas posterior yaitu dengan menentukan 1 komentar dari data uji dengan contoh data uji = “enggak punya lingkungan hidup orang toxic kesehatan mental stop toxic”. Dari data uji yang ditentukan didapat dua kata yang termasuk ke dalam data latih yaitu adalah kata “Sehat” dan “Mental”.

Perhitungan nilai probabilitas posterior

$$P(\text{Komentar} | \text{Kelas}) = P_{\text{Term}_1} \times \dots \times P_{\text{Term}_n} \times P(\text{Kelas} | \text{Komentar}) \tag{4}$$

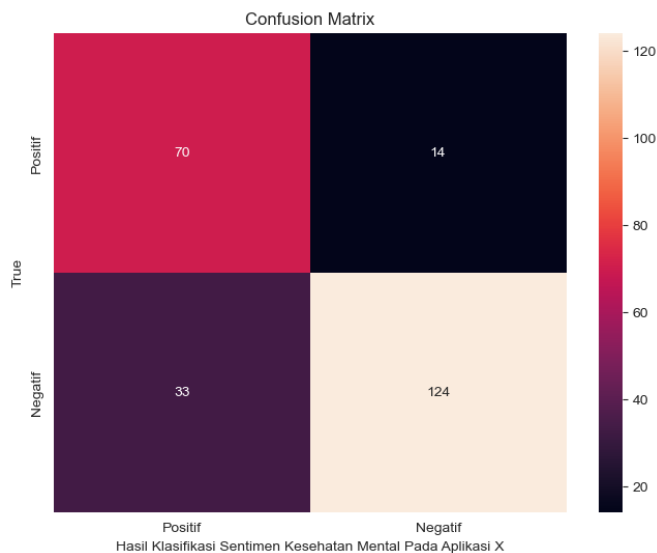
$$\begin{aligned} P(\text{DataUji} | \text{Positif}) &= P(\text{positif}) \times P(\text{sehat} | \text{positif}) \times P(\text{mental} | \text{positif}) \\ &= 0.5 \times 0.05 \times 0.05 \\ &= 0.00125 \end{aligned}$$

$$\begin{aligned} P(\text{DataUji} | \text{Negatif}) &= P(\text{negatif}) \times P(\text{sehat} | \text{negatif}) \times P(\text{mental} | \text{negatif}) \\ &= 0.5 \times 0.056 \times 0.056 \\ &= 0.001568 \end{aligned}$$

Hasil klasifikasi sentimen "sehat, mental" adalah Negatif, dengan nilai tertinggi dari data uji sentimen negatif dengan nilai 0.001568. Jadi nilai komentar ini diklasifikasikan negatif.

### 3.4.2 Mencari Akurasi

Gambar 4. merupakan confusion matrix. Langkah selanjutnya adalah dengan mencari akurasi pada algoritma naïve bayes dapat dilihat pada gambar dibawah.



**Gambar 4.** Confusion Matrix

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{70+124}{70+124+14+33} \times 100\% = 80\%$$

$$\text{precision} = \frac{TP}{TP+FP} = \frac{70}{70+14} \times 100\% = 83\%$$

$$\text{recall} = \frac{TP}{TP+FN} = \frac{70}{70+33} \times 100\% = 68\%$$

$$F1 - \text{Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = 2 \times \frac{0.833 \times 0.68}{0.833 + 0.68} \times 100\% = 74.9\%$$

Akurasi komentar untuk dataset terdapat pada gambar di bawah ini dimana didapat 0.73, precision yaitu 1.0, recall yaitu 0.114 dan F1-Score yaitu 0.205.

```

Accuracy: 0.8049792531120332
precision    recall  f1-score   support

Negatif      0.68    0.83    0.75        84
Positif      0.90    0.79    0.84       157

accuracy          0.80        241
macro avg         0.79        241
weighted avg      0.82        241
    
```

**Gambar 5.** Akurasi Dataset





- [15] R. Y. Hayuningtyas, “Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita,” *J. Inform.*, vol. 6, no. 1, pp. 18–22, 2019.
- [16] M. F. Rifai, H. Jatnika, and B. Valentino, “Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS),” 2019.
- [17] A. Karim, S. Esabella, K. Kusmanto, S. Suryadi, and E. Purba, “Penerapan Metode Teorema Bayes Dalam Mendiagnosa Penyakit Autoimun,” *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 254–263, 2023.
- [18] C. Nilam, “Analisis Sentimen Mengenai Program MSIB Pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes Classifier,” 2023.
- [19] V. P. Rantung and M. ST, *Teknik-Teknik Pemrosesan Bahasa Alami (NLP)*. Lakeisha, 2023.
- [20] N. L. Amalia, “Analisis sentimen pada perpindahan Ibukota Indonesia dengan algoritma support vector machine: evaluasi leksikon, metode ekstraksi fitur, dan kernel trick.” *Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah ...*, 2022.
- [21] F. Baehaqi and N. Cahyono, “Analisis Sentimen Terhadap Cyberbullying Pada Komentar Di Instagram Menggunakan Algoritma Naïve Bayes,” *Indones. J. Comput. Sci.*, vol. 13, no. 1, 2024.