

Analisis Sentimen Komentar Youtube terhadap Food Vlogger dengan Menggunakan Metode Naïve Bayes

¹Sulistia Maharani Harahap

Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara

Email : sulistiamaharani354@gmail.com

²Rakhamat Kurniawan

Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara

Email : rakhmat.Kr@uinsu.ac.id

ABSTRACT

YouTube is currently the media most frequently accessed by internet users in Indonesia with a percentage of 88%, then WhatsApp at 83%, Facebook at 81%, and Instagram at 80%. One of the themes that is often uploaded is about food reviews and people who upload them are called food vloggers. Food vloggers influence millennial audiences, namely before and after food consumption. The purpose of food review videos is to provide reviews of food in terms of taste, price, quality, atmosphere, etc. Other people also have the freedom to comment on food review videos. This research aims to analyze public sentiment towards food vloggers using data from YouTube social media. This research aims to analyze sentiment using the Naïve Bayes algorithm method for food vloggers used to analyze sentiment with an accuracy value of 44.8276%, precision of 16% or 0.16%, recall of 100% or 1.00%, and f1-score of 27% or 0.27%. This sentiment analysis can be a reference in understanding food vlogger preferences.

Keywords: *sentiment, food vlogger, YouTube, naive Bayes, text preprocessing*

PENDAHULUAN

Perkembangan teknologi informasi membawa suatu perubahan dalam masyarakat. Lahirnya media sosial menjadikan pola perilaku masyarakat mengalami perubahan baik budaya, etika, dan norma yang ada. Penggunaan platform media sosial (*medsos*) mengalami peningkatan pesat saat ini [1]. Platform seperti Instagram, Facebook, Twitter, dan YouTube menyediakan beragam informasi kepada pengguna dalam bentuk gambar, video, dan teks. Di antara ketiga media sosial tersebut. Di Indonesia, YouTube memiliki peningkatan yang sangat tinggi. Banyak *channel YouTube* dengan berbagai kategori konten video yang di upload oleh user atau pengguna *youtube* [2].

YouTube saat ini merupakan media yang paling sering diakses oleh pengguna internet di Indonesia dengan persentase sebesar 88%, kemudian WhatsApp sebesar 83%, Facebook sebesar 81%, dan Instagram sebesar 80%. Salah satu tema yang banyak diunggah adalah mengenai ulasan makanan (*food review*) dan orang yang mengunggahnya disebut dengan *food vlogger*. *Food vlogger* atau sebutan pada pembuat konten tentang

makanan ini mampu memberikan pengaruh dan meningkatkan penonton untuk membeli produk baik makanan maupun minuman yang ditampilkan [3]. Kolom komentar yang disediakan oleh youtube dapat dijadikan sebuah informasi Analisis Media Sosial yang diklasifikasikan menjadi opini positif dan opini negatif. *Opini public* terhadap tayangan di media sosial youtube dapat diketahui dengan analisis sentiment [4].

Model *Natural Language Processing* (NLP) telah mencapai kesuksesan fenomenal dalam tugas linguistik dan semantik seperti klasifikasi teks, terjemahan mesin, sistem dialog kognitif, pencarian informasi melalui *Natural Language Understanding* (NLU), dan *Natural Language Generation* (NLG) [5]. Pendekatan *text mining* dengan NLP (*Natural Language Processing*) dilakukan dalam menganalisa dan mencari makna dari setiap *opini* atau pendapat pengguna *Youtube*. Analisa sentimen memerlukan metode dalam mengklasifikasikan teks, salah satunya yakni metode *Gaussian naïve bayes* [7]. Algoritma *Naïve Bayes* merupakan suatu bentuk klasifikasi data dengan menggunakan metode probabilitas dan statistik. *Naïve bayes*

merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan bayes dengan asumsi ketidaktergantungan yang kuat. Algoritma *Naïve Bayes* ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes yang merupakan memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

Analisis sentimen adalah kegiatan yang digunakan untuk menganalisis pendapat atau opini seseorang tentang suatu topik. Hasil dari penelitian ini adalah klasifikasi komentar kedalam sentiment positif, dan sentiment negatif. Hasil pengujian algoritma yang digunakan diperoleh nilai akurasi yang menunjukkan seberapa persen ketepatan algoritma *Naïve Bayes Classifier* untuk melakukan klasifikasi sentiment [7].

Berdasarkan latar belakang diatas, maka penulis melakukan untuk mengambil judul “Analisis Sentimen Komentar YouTube Terhadap *Food Vlogger* Dengan Menggunakan Metode *Naïve Bayes*” yang bertujuan untuk menghasilkan klasifikasi dari komentar – komentar pengguna YouTube dengan kategori positif dan *negative*. Diharapkan sistem ini dapat menjadi bahan evaluasi para pemilik video meningkatkan kualitas dari saluran YouTube-nya.

METODE PENELITIAN

Terkait penelitian yang akan dilakukan, penulis berencana untuk membuat sistem yang dapat melakukan analisis sentimen agar dapat mengetahui sentimen yang diberikan oleh para pengguna *youtube* terkait mengenai *rivew – rivew* makanan yang diakukan oleh *food vlogger*, apakah sentimen tersebut bersifat positif atau negatif dengan menggunakan metode *naïve bayes classifier*. Adapun tahapan yang dilalui dalam penelitian ini meliputi :



Gambar 1 Kerangka Berfikir

1. Identifikasi Masalah

Prosedur penelitian yang digunakan dalam penelitian ini diawali dengan identifikasi suatu permasalahan atau permasalahan. Pada tahap identifikasi masalah dirumuskan latar belakang masalah yang menjadi alasan mengapa penelitian ini dilakukan.

2.Studi Literatur

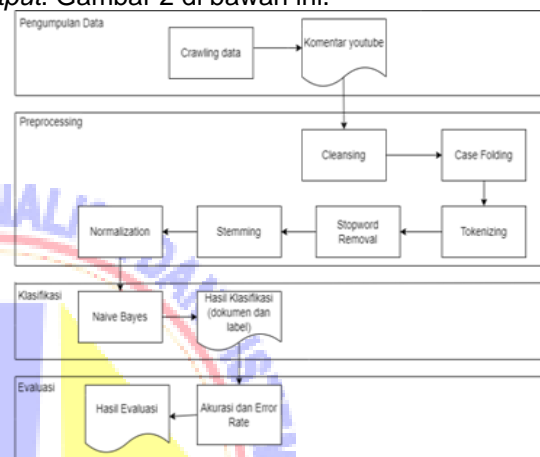
Langkah selanjutnya adalah tinjauan pustaka, yaitu penelusuran jurnal ilmiah, tesis, dan disertasi untuk dijadikan referensi judul penelitian.

3.Pengumpulan Data

Pengumpulan data merupakan tahapan yang paling penting dalam melakukan sebuah penelitian untuk mengumpulkan data yang diperlukan terkait penelitian yang akan dilakukan.

4.Disain Sistem

Diagram alur proses yang menjadi sebagai acuan bagi peneliti sebelum sistem tersebut dirancang dikenal sebagai diagram alur proses. Di dalam diagram alur proses ini mencakup tiga komponen utama yakni *input*, proses, dan *output*. Gambar 2 di bawah ini.



Gambar 2 Desain Sistem

5.Pembotan TF - IDF

Dalam langkah ini, dilakukan perhitungan bobot untuk pada suatu kata (*term*) yang terdapat dalam suatu dataset yang berasal dari komentar pengguna YouTube. Proses *TF-IDF* dimulai dengan menghitung nilai *term frequency* (TF) dan *document frequency* (DF). Selanjutnya, dihitung nilai invers *document frequency* (IDF) dan bobot akhir diperoleh dengan perkalian nilai *TF* dan *IDF* [7].

$$TF(d, t) = f(d, f) \dots\dots\dots(1)$$

$$IDF = \ln \frac{d + 1}{df + 1} \dots\dots\dots(2)$$

$$W(d, t) = TF(d, t) \times IDF \dots\dots\dots(3)$$

6.Algoritma Naïve Bayes

Algoritma *Naïve Bayes Classifier* adalah metode yang menerapkan konsep probabilitas bersyarat. Pada dasarnya algoritma NBC dapat dinotasikan pada persamaan sebagai berikut.

$$V_{map} = \underset{v_{jev}}{argmax} \prod_{i=1}^n P(x_i | V_j) P(V_j) \dots\dots\dots(4)$$

Untuk $P(x_i | V_j)$ dan $P(V_j)$ dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut :

$$P(V_j) = \frac{|docs_j|}{|contoh|} \dots \dots \dots (5)$$

$$P(x_i|V_j) \dots \dots \dots (6)$$

7. Evaluasi Hasil

Hasil pengujian menggunakan metode *Naïve Bayes* divisualisasikan dalam bentuk *confusion matrix*, yang dimana bentuk ini memberikan perbandingan hasil klasifikasi *Naïve Bayes* yang dibantu oleh metode pembobotan *TF-IDF*.

HASIL DAN PEMBAHASAN

1. Analisis Data

Pada penelitian ini penulis akan mengalisis data sentimen terhadap komentar youtube *food vlogger* dengan melakukan pengumpulan data. Proses pengumpulan data dilakukan dengan teknik *crawling* data komentar youtube *food vlogger* dengan menggunakan *library python*. Data yang dipakai didapatkan dari hasil *crawling* dan disimpan dan diproses dengan format.csv.

2. Pengumpulan Data

Pemrosesan *crawling* data komentar yuutube *food vlogger* menggunakan bahasa pemrograman python pada aplikasi *Google Colab*.

Gambar 3 Hasil Crawling Data

3. Preprocessing Data

Tahap ini memiliki beberapa tahapan yaitu *cleaning*, *tokenize*, *case folding*, *normalize*, *filtering*. Berikut adalah alur pada proses *preprocessing* :

Tabel 1 Data Hasil Cleaning

Hasil Komentar	Hasil Cleaning
Waduh...gratis lage...kebangeten...ja di seneng bgt tuh...öÿ~šöÿ~š	Waduhgratis lagekebangetenj adi seneng bgt tuh
Abang ownernya sangat friendly	Abang ownernya sangat friendly
Mirip bang evan hobby makan sihh ownerx	Mirip bang evan hoby makan sihh ownerx
Gimana nih kalau Bang Evan Hobby Makan Kulineran	Gimana nih kalau Bang Evan Hoby Makan Kulineran

Kesini.. Ayo like yang banyak biar Bang Evan Kesini.. öÿ~„öÿ~„	Kesini Ayo like yang banyak biar Bang Evan Kesini
Ayooo haha	Ayo haha

1. *Case Folding* : Pada tahap ini, seluruh huruf pada dataset diseragamkan menjadi huruf kecil (*Lowercase*).

Tabel 2 Hasil Casefolding

Hasil Casefolding
waduhgratis lagekebangetenjadi seneng bgt tuh
abang ownernya sangat friendly
mirip bang evan hobby makan sihh ownerx
gimana nih kalau bang evan hobby makan kulineran kesini ayo like yang banyak biar bang evan kesini
ayo haha

2. *Tokenizing*

Pada tahap ini dilakukan untuk memisahkan dokumen dalam dataset menjadi kata per kata.

Tabel 3 Data Hasil Tokenizing

Hasil Tokenizing
[waduhgratis, lagekebangetenjadi, seneng, bgt, tuh]
[abang, ownernya, sangat, friendly]
[mirip, bang, evan, hobby, makan, sihh, ownerx]
[gimana, nih, kalau, bang, evan, hobby, makan, kulineran, kesini, ayo, like, yang, banyak, biar, bang, evan, kesini]
[ayo, haha]

3. *Stopword Removal* : Tahap ini dilakukan menghapus kata dalam dokumen yang tidak memiliki arti seperti kata ganti, kata penghubung dan lainnya.

Tabel 4 Data Hasil Stopword Removal

Hasil Stopword Removal
[waduh, gratis, lage, kebangeten,jadi, seneng, bgt, tuh]
[abang, ownernya, friendly]
[bang, evan, hobby, makan, sihh, ownerx]
[gimana, nih, bang, evan, hobby, makan, kulineran, kesini, ayo, like, biar, bang, evan, kesini]
[ayo, haha]

4. *Stemming*

Tahap ini dilakukan untuk menghapus seluruh imbuhan.

Tabel 5 Data Hasil Stemming

Hasil Stemming

Waduh gratis lage kebangeten jadi neng banget tuh
abang ownernya friendly
bang evan hoby makan sih ownerx
gimana nih bang evan hoby makan kuliner kesini ayo like biar bang evan kesini
ayo haha

	kesini
5	ayo haha

Tabel 8 Kata – Kata Pada Kalimat diatas

waduh	bang	makan
gratis	evan	kuliner
lage	hoby	kesini
kebangetan	makan	ayo
jadi	sih	like
neng	ownernya	biar
banget	gimana	bang
tuh	nih	evan
abang	bang	kesini
ownernya	evan	ayo
friendly	hoby	haha

5. Normalization

Pada tahap Normalization dilakukan untuk memperbaiki, menghapus kata yang tidak baku, singkatan serta penulisan kata yang tidak sesuai bahasa Indonesia yang baku.

Tabel 6 Data Hasil Normalization

Hasil Normalization	
Waduh gratis lage kebangeten jadi neng banget tuh	
abang ownernya friendly	
bang evan hoby makan sih ownerx	
gimana nih bang evan hoby makan kuliner kesini ayo like biar bang evan kesini	
ayo haha	

Berdasarkan kata-kata seperti pada tabel 8 di atas, maka dihitunglah nilai term frequency. D1 mewakilkan kalimat nomor 1, D2 mewakilkan kalimat nomor 2, D3 mewakilkan kalimat nomor 3, D4 mewakilkan kalimat nomor 4, dan D5

Tabel 9 Perhitungan Nilai TF dan DF

Term	TF					DF
	D1	D2	D3	D4	D5	
abang	0	1	0	0	0	1
ayo	0	0	0	1	1	2
banget	1	0	0	0	0	1
bang	0	0	1	2	0	3
biar	0	0	0	1	0	1
evan	0	0	1	2	0	3
friendly	0	1	0	0	0	1
gratis	1	0	0	0	0	1
gimana	0	0	0	1	0	1
haha	0	0	0	0	1	1
hoby	0	0	1	1	0	2
jadi	1	0	0	0	0	1
kebangetan	1	0	0	0	0	1
kuliner	0	0	0	1	0	1
kesini	0	0	0	2	0	2
lage	1	0	0	0	0	1
like	0	0	0	1	0	1
makan	0	0	1	1	0	2
neng	1	0	0	0	0	1
nih	0	0	0	1	0	1
ownernya	0	1	0	0	0	1
ownerx	0	0	1	0	0	1
sih	0	0	1	0	0	1
tuh	1	0	0	0	0	1

4. Labelisasi Dataset

Pada tahap ini, daa akan dipresos secara otomatis yaitu melakukan perhitungan nilai skor sentimen dengan bahasa python. Pada penelitian ini penulis menggunakan google collaboratory dengan menggunakan bahasa phton.



Gambar 4 Pelabelan Data

5. Pembobotan TF-IDF

Tahap ini bertujuan untuk menghitung bobot tiap kata (term) yang terdapat didalam dokumen (N).

Berikut contoh perhitungan dari proses TF-IDF dari dokumen yang diberikan sebelumnya disajikan dalam tabel 7.

Tabel 7 Hasil Akhir Preprocessing

No	Hasil Normalization
1	waduh gratis lage kebangeten jadi neng banget tuh
2	abang ownernya friendly
3	bang evan hoby makan sih ownerx
4	gimana nih bang evan hoby makan kuliner kesini ayo like biar bang evan

berikut disajikan tabel hasil perhitungan *IDF* melanjutkan dari tahap perhitungan sebelumnya.

Berikut disajikan contoh perhitungan untuk *term* pertama yaitu "abang" yang mana telah diketahui jumlah $d = 5$ dan $df = 1$

$$IDF = \ln \frac{d + 1}{df + 1} + 1$$

$$IDF = \ln \frac{5+1}{1+1} + 1 = 2,098612$$

Tabel 10 Perhitungan Nilai *IDF*

Term	DF	Skor IDF
abang	1	2,09861
ayo	2	1,69315
banget	1	2,09861
bang	3	1,40546
biar	1	2,09861
evan	3	1,40546
frendly	1	2,09861
gratis	1	2,09861
gimana	1	2,09861
haha	1	2,09861
hoby	2	1,69315
jadi	1	2,09861
kebangetan	1	2,09861
kuliner	1	2,09861
kesini	2	1,69315
lage	1	2,09861
like	1	2,09861

Berikut disajikan perhitungan untuk *term* pertama yaitu "abang" yang mana telah diketahui nilai $tf = 1$ dan $idf = 2,098612$

$$W_{dt} = TF_{dt} \times IDF$$

$$W_{dt} = 1 \times 2,098612 = 2,098612$$

Tabel 11 Perhitungan Nilai *TF-IDF* (W_{dt})

Term	TF-IDF (W_{dt})				
	D1	D2	D3	D4	D5
abang	0	2,098612	0	0	0
ayo	0	0	0	2,098612	2,098612
banget	2,098612	0	0	0	0
bang	0	0	2,098612	2,81092	0
biar	0	0	0	2,098612	0
evan	0	0	2,098612	2,81092	0
frendly	0	2,098612	0	0	0
gratis	2,098612	0	0	0	0

gimana	0	0	0	2,098612	0
haha	0	0	0	0	2,098612
hoby	0	0	2,098612	2,098612	0
jadi	2,098612	0	0	0	0
kebangetan	2,098612	0	0	0	0
kuliner	0	0	0	2,098612	0
kesini	0	0	0	2	0
lage	2,098612	0	0	0	0
like	0	0	0	2,098612	0
makan	0	0	2,098612	2,098612	0
neng	2,098612	0	0	0	0
nih	0	0	0	2,098612	0
ownernya	0	2,098612	0	0	0
ownerx	0	0	2,098612	0	0
sih	0	0	2,098612	0	0
tuh	2,098612	0	0	0	0
waduh	2,098612	0	0	0	0

Gambar 5 Pembobotan *TF-IDF* pada Sistem

6. Split Dataset

Dalam penelitian ini akan diterapkan *split* data dengan perbandingan 8:2 yang artinya 80% dari total *dataset* akan digunakan sebagai data uji.

```

Size of X_train: (938,1173)
Size of y_train: (938,1173)
Size of X_test: (235,1173)
Size of y_test: (235,1173)
    
```

Gambar 6 Perbandingan Data Latih dan Data Uji

7. Klasifikasi *Naive Bayes*

Setelah tahapan pembobotan kata selesai, maka akan dilakukan tahap pengklasifikasian terhadap data uji dengan metode *naive bayes*.

Tabel 12 Contoh Data Latih

dok	Data Opini	Label
D1	Parah lu koh tidak bayar kasihan orang usaha	Negatif
D2	masakan enak goyang nasi padang	Positif
D3	sampayang enak minum udah mulai panas	Positif
D4	bayarlah bang abis makan	Negatif
D5	pantesan makin gemuk makan gratis terus	Negatif

1. Data latih sentiment positif

Tabel 13 Data Latih Positif

#	Term	Frekuensi (n_k)
1	masakan	1
2	enak	2
3	goyang	1
4	nasi	1
5	padang	1
6	sampayang	1
7	minum	1
8	udah	1
9	mulai	1
10	panas	1

Pada tabel diatas, dapat diketahui bahwa :

- Jumlah frekuensi keseluruhan sentiment positif (n) = 11
- Jumlah kosakata keseluruhan sentiment positif = 10

2. Data latih sentiment negatif

Tabel 14 Data Latih Sentimen Negatif

#	Term	Frekuensi (n_k)
1	parah	1
2	lu	1
3	koh	1
4	tidak	1
5	bayar	1
6	kasihan	1
7	orang	1
8	usaha	1
9	bayarlah	1
10	bang	1
11	abis	1
12	makan	2
13	pantesan	1
14	makin	1
15	gemuk	1
16	gratis	1
17	terus	1

Pada tabel diatas, dapat diketahui bahwa :

- Jumlah frekuensi keseluruhan sentiment negatif (n) = 18
- Jumlah kosakata keseluruhan sentiment negatif = 17

3. Data Uji

Tabel 15 Contoh Data Uji

Data Uji	Label
gila murah banget cinta nasi padang surga banget	?

Tahap dalam proses perhitungan klasifikasi kelas pada data uji dimulai dari perhitungan nilai *prior probability*, *conditiona probability*, dan *posterior probability*.

1. Perhitungan nilai *prior probability*

$$P(V_j) = \frac{|docs_j|}{|contoh_j|}$$

a. $P(Positive) = \frac{1}{2} = 0,5$

b. $P(Negative) = \frac{1}{2} = 0,5$

2. Perhitungan nilai *probability* setiap *term*

1. Data Latih

Probabilitas setiap *term* dapat dicari dengan menggunakan rumus :

$$P(X_i|V_j) = \frac{n_k + 1}{n + |kosakata|}$$

a. Probabilitas data latih sentiment positif

$P(masakan|positif) = \frac{1+1}{11+10} = 0,0952$

$P(enak|positif) = \frac{2+1}{11+10} = 0,1428$

$P(goyang|positif) = \frac{1+1}{11+10} = 0,0952$

$P(nasi|positif) = \frac{1+1}{11+10} = 0,0952$

$P(padang|positif) = \frac{1+1}{11+10} = 0,0952$

$P(sampayang|positif) = \frac{1+1}{11+10} = 0,0952$

$P(minum|positif) = \frac{1+1}{11+10} = 0,0952$

$P(udah|positif) = \frac{1+1}{11+10} = 0,0952$

$P(mulai|positif) = \frac{1+1}{11+10} = 0,0952$

$P(panas|positif) = \frac{1+1}{11+10} = 0,0952$

b. Probabilitas data latih sentiment negative

$P(parah|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(lu|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(koh|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(tidak|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(bayar|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(kasihan|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(orang|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(usaha|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(bayarlah|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(bang|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(abis|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(makan|negatif) = \frac{2+1}{18+17} = 0,0857$

$P(pantesan|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(makin|negatif) = \frac{1+1}{18+17} = 0,0571$

$P(gemuk|negatif) = \frac{1+1}{18+17} = 0,0571$

$$P(\text{gratis}|\text{negatif}) = \frac{1+1}{18+17} = 0,0571$$

$$P(\text{terus}|\text{negatif}) = \frac{1+1}{18+17} = 0,0571$$

Dengan menghitung nilai probabilitas dari semua kategori sentimen maka terdapat perubahan nilai probabilitas dari setiap *term* pada data latih. Sehingga perubahan probabilitas sebagai berikut :

Tabel 16 Perubahan Probabilitas *Term* Pada Sentimen Positif

#	Term	Frekuensi (n_k)	Probabilitas $P(x_j V_j)$
1	masakan	1	$(1+1)/(11+27) = 0,052$
2	enak	2	$(2+1)/(11+27) = 0,079$
3	goyang	1	$(1+1)/(11+27) = 0,052$
4	nasi	1	$(1+1)/(11+27) = 0,052$
5	padang	1	$(1+1)/(11+27) = 0,052$
6	sampayan	1	$(1+1)/(11+27) = 0,052$
7	minum	1	$(1+1)/(11+27) = 0,052$
8	udah	1	$(1+1)/(11+27) = 0,052$
9	mulai	1	$(1+1)/(11+27) = 0,052$
10	panas	1	$(1+1)/(11+27) = 0,052$

Tabel 17 Perubahan Probabilitas *Term* Pada Sentimen Negatif

#	Term	Frekuensi (n_k)	Probabilitas $P(x_j V_j)$
1	parah	1	$(1+1)/(18+27) = 0,044$
2	lu	1	$(1+1)/(18+27) = 0,044$
3	koh	1	$(1+1)/(18+27) = 0,044$
4	tidak	1	$(1+1)/(18+27) = 0,044$
5	bayar	1	$(1+1)/(18+27) = 0,044$
6	kasihan	1	$(1+1)/(18+27) = 0,044$
7	prang	1	$(1+1)/(18+27) = 0,044$
8	usaha	1	$(1+1)/(18+27) = 0,044$
9	bayarlah	1	$(1+1)/(18+27) = 0,044$
10	bang	1	$(1+1)/(18+27) = 0,044$

11	abis	1	$(1+1)/(18+27) = 0,044$
12	makan	2	$(2+1)/(18+27) = 0,066$
13	pantesan	1	$(1+1)/(18+27) = 0,044$
14	makin	1	$(1+1)/(18+27) = 0,044$
15	gemuk	1	$(1+1)/(18+27) = 0,044$
16	gratis	1	$(1+1)/(18+27) = 0,044$
17	terus	1	$(1+1)/(18+27) = 0,044$

2.Data Uji

Dalam tahap klasifikasi data uji, term yang terdaftar pada tabel 15 frekuensi data uji akan diklasifikasi dicari nilai probabilitasnya dengan membandingkan dan mencocokkan dengan *term* yang ada pada sentimen positif dan sentimen negatif data latih.

a. Kelas sentiment positif

$$P(\text{gila}|\text{positif}) = \frac{0+1}{11+27} = 0,026$$

$$P(\text{murah}|\text{positif}) = \frac{0+1}{11+27} = 0,026$$

$$P(\text{banget}|\text{positif}) = \frac{0+1}{11+27} = 0,026$$

$$P(\text{cinta}|\text{positif}) = \frac{0+1}{11+27} = 0,026$$

$$P(\text{nasi}|\text{positif}) = \frac{1+1}{11+27} = 0,052$$

$$P(\text{padang}|\text{positif}) = \frac{1+1}{11+27} = 0,052$$

$$P(\text{surga}|\text{positif}) = \frac{0+1}{11+27} = 0,026$$

Berdasarkan nilai probabilitas data uji pada sentiment positif diatas, nilai V_{map} untuk sentiment positif adalah sebagai berikut :

$$V_{map} = \text{argmax}_{(positif)} \prod_{i=1}^{13} P(x_i|V_j)P(V_j) = ((0,026)(0,026)(0,026)(0,026)(0,052)(0,052)(0,026)^* (0,05)) = 1,6063620352E-12$$

a. Kelas sentiment negative

$$P(\text{gila}|\text{negatif}) = \frac{0+1}{18+27} = 0,022$$

$$P(\text{murah}|\text{negatif}) = \frac{0+1}{18+27} = 0,022$$

$$P(\text{banget}|\text{negatif}) = \frac{0+1}{18+27} = 0,022$$

$$P(\text{cinta}|\text{negatif}) = \frac{0+1}{18+27} = 0,022$$

$$P(\text{nasi}|\text{negatif}) = \frac{18+27}{0+1} = 0,022$$

$$P(\text{padang}|\text{negatif}) = \frac{0+1}{18+27} = 0,022$$

$$P(\text{surga}|\text{negatif}) = \frac{0+1}{18+27} = 0,022$$

Berdasarkan nilai probabilitas data uji pada sentiment positif diatas, nilai untuk sentiment positif adalah sebagai berikut :

$$V_{map} = \underset{\text{armax}}{\text{(positif)}} \prod_{i=1}^{13} P(x_i|V_j)P(V_j)$$

$$= ((0,022)(0,022)(0,022)(0,022)(0,022)(0,022)(0,022)(0,022)(0,022)(0,022)(0,022)(0,022)(0,022)) * (0,05)$$

$$= 1,247178944E-13$$

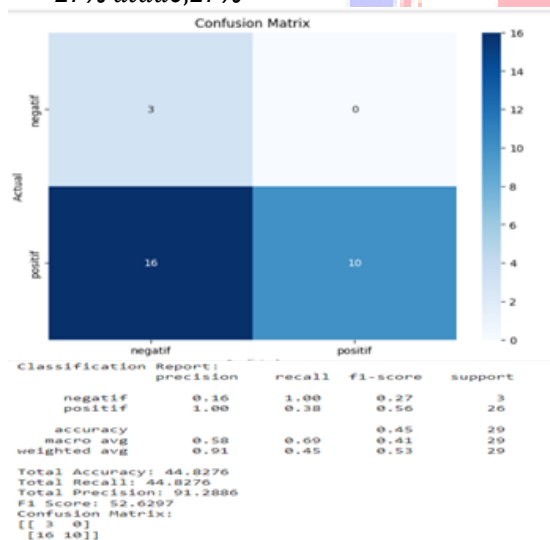
Evaluasi Hasil

Hasil pengujian menggunakan metode *Naïve Bayes* divisualisasikan dalam bentuk *confusion matrix*, yang dimana bentuk ini memberikan perbandingan hasil klasifikasi *Naïve Bayes* yang dibantu oleh metode pembobotan *TF-IDF*.

Tabel 11 *Confusion Matrix* Hasil Klasifikasi

ij		Kelas Prediksi (j)	
		Positif	Negatif
Kelas Aktual (i)	Positif	3	16
	Negatif	0	10

- a. Perhitungan *accuracy*:
 $Accuracy = \frac{3}{3+0+16+10} \times 100\% = 44,8276$
- b. Perhitungan *Precision*
 $Precision = \frac{3}{3+16} \times 100\% = 16\% \text{ atau } 0,16\%$
- c. Perhitungan *Recal*
 $Recall = \frac{3}{3+0} \times 100\% = 100\% \text{ atau } 1,00\%$
- d. Perhitungan *F1 score* = $\frac{2 \times 16 \times 100}{16+100} \times 100\% = 27\% \text{ atau } 0,27\%$



Gambar 7 *Classification Report*

KESIMPULAN

Berdasarkan hasil penelitian analisis sentimen komentar *youtube food vlogger* menggunakan data hasil *crawling* dengan kata kunci '*food vlogger*' diperoleh data sebanyak 1000 komentar kemudian data tersebut melalui tahap *preprocessing* yang dimana tahapan

didalamnya meliputi *cleaning, case folding, tokenizing, stopword removal, stemming, normaization*, agar data yang diperoleh dapat digunakan pada tahap pelabelan dengan metode *TF-IDF*. Selanjutnya data yang telah dilabeli akan diklasifikasi dengan memakai metode *Naïve Bayes*. Dari hasil *confusion matrix* diperoleh performa klasifikasi dengan metode *Naïve Bayes* yang diperkuat dengan pembobotan *TF-IDF* yang menghasilkan persentase akurasi sebesar 44.8276 % dari total 1000 komentar, dengan persentase sebesar 58% untuk untuk *precision*, 16% untuk nilai *recall* 100% untuk nilai *f1-cscore* 27%.

DAFTAR PUSTAKA

- [1] Jannah, F. R., Megiananta Aprilistya, A., & Khadijah, S. (2023). Analisis Framing Pemberitaan Perseteruan Farida Nurhan Dan Food Vlogger Codeblu Di Kompas.Com Dan Viva.Co.Id. *Prosiding Seminar Nasional*, 1040–1048.
- [2] Calvin Jonathan, Theresia Herlina Rochadiani, & Thamrin Sofian. (2023). Analisis Sentimen Komentar Video Youtube Flat Earth Theory Dengan Menggunakan Metode Unsupervised Dan Supervised Learning. *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2), 378–387. <https://doi.org/10.51454/decode.v3i2.210>
- [3] Afifah, N. U. R. (2022). PURCHASE INTENSITY ANALYSIS OF MUSLIM MILLENNIAL CONSUMERS WITH YOUTUBE FOOD CONCENT. 7(1), 57 – 72
- [4] Chely Aulia Misrun, Haerani, E., Fikry, M., & Budianita, E. (2023). Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1), 207–215. <https://doi.org/10.37859/coscitech.v4i1.4790>
- [5] Riset, J., Hasibuan, M.S. and Serdano, A. (2022) 'Analisis Sentimen Kebijakan Pembelajaran Tatap Muka Menggunakan Support Vector Machine dan Naive Bayes Policy Sentiment Analysis Face-to-face Learning Using Supports Vector and Naive Bayes Engines', 6(2), pp. 199–204.
- [6] Mujahidin, S., Prasetio, B., & Utomo, M. C. C. (2022). Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naïve bayes. *Voteteknika (Vocational Teknik*

Elektronika Dan Informatika), 10(3), 17.
<https://doi.org/10.24036/voteteknika.v10i3.118299>

- [7] Lubis, A.H. and Harahap, Y.F. (2023) 'Analisis Sentimen Masyarakat Terhadap Resesi Ekonomi Global 2023 Menggunakan Algoritma Naïve Bayes Classifier', 16(2), pp. 442–450



