

Analisis Sentimen Komentar Youtube Masterchef Indonesia Menggunakan Algoritma Support Vector Machine Dan Gaussian Naïve Bayes

Sentiment Analysis of YouTube Comments on Masterchef Indonesia Using Support Vector Machine and Gaussian Naïve Bayes Algorithms

Dirgahayu Marganingsih¹, Hardian Oktavianto², Ginanjar Abdurrahman³

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

*e-mail: dirgahayumrg@gmail.com¹, hardian@unmuhjember.ac.id², abdurrahmanginanjar@unmuhjember.ac.id³

Received:
09.04.2025

Revised:
21.04.2025

Accepted:
17.05.2025

Available online:
30.05.2025

Abstract: Social media has become the main platform for sharing information, including video sharing-based platforms such as YouTube, which provides educational, informational and entertainment content for the public. One content that is in great demand is a cooking show adapted from England, namely Masterchef Indonesia. Which ran for 11 seasons from 2011 to 2023 with 260 episodes. However, in the 11th season, the grand final episode, there were pros and cons regarding the participants' victory, which provoked the public by the presence of presidential candidate Ganjar Pranowo. Through the comments feature, users can express opinions and responses, which can then be analyzed to understand the public's views. This research aims to analyze positive and negative comments on the Masterchef Indonesia channel using the Support Vector Machine and Gaussian Naïve Bayes algorithms. Before applying the oversampling technique for data balancing, the Support Vector Machine algorithm produced 82% accuracy, 88% precision and 72% recall, while Gaussian Naïve Bayes produced 65% accuracy, 52% precision and 81% recall. After the oversampling technique was applied, the performance of the Support Vector Machine increased with 85% accuracy, 84% precision, and 89% recall, while Gaussian Naïve Bayes obtained 72% accuracy, 68% precision, and 72% recall.

Keywords: Sentiment Analysis, Support Vector Machine, Gaussian Naïve Bayes

Abstrak: Media sosial menjadi platform utama untuk berbagi informasi, termasuk platform berbasis video sharing seperti Youtube, yang menyediakan konten edukasi, informasi, dan hiburan bagi masyarakat. Salah satu konten yang banyak diminati adalah acara memasak yang diadaptasi dari Inggris, yaitu Masterchef Indonesia. Yang berjalan selama 11 musim dari tahun 2011 sampai 2023 dengan 260 episode. Namun, pada musim ke 11 episode grand final terdapat pro dan kontra tentang kemenangan peserta yang membuat publik terprovokasi dengan kehadiran sosok calon presiden Ganjar Pranowo. Melalui fitur komentar, pengguna dapat menyampaikan pendapat dan respon, yang kemudian dapat dianalisis untuk memahami pandangan masyarakat. Penelitian ini bertujuan untuk menganalisis komentar yang bernilai positif dan negatif pada *channel* Masterchef Indonesia dengan algoritma *Support Vector Machine* dan *Gaussian Naïve Bayes*. Sebelum diterapkan teknik *oversampling* untuk penyeimbangan data, algoritma *Support Vector Machine* menghasilkan akurasi 82%, presisi 88%, dan *recall* 72%, sedangkan *Gaussian Naïve Bayes* menghasilkan akurasi 65%, presisi 52%, dan *recall* 81%. Setelah teknik *oversampling* diterapkan, performa *Support Vector Machine* meningkat dengan akurasi 85%, presisi 84%, dan *recall* 89%, sementara *Gaussian Naïve Bayes* memperoleh akurasi 72%, presisi 68%, dan *recall* 72%.

Kata kunci: Analisis Sentimen, Support Vector Machine, Gaussian Naïve Bayes

1. PENDAHULUAN

Media sosial merupakan *platform* atau layanan daring yang digunakan oleh individu dan kelompok untuk berkomunikasi, berbagi informasi, dan berinteraksi secara *online*. Selain itu, media sosial digunakan untuk mendapatkan dan menyampaikan informasi kepada masyarakat dan berbagai aspek kehidupan modern. Terdapat berbagai jenis media sosial, salah satunya adalah Youtube, yang merupakan platform berbasis platform video-sharing. Dimana pengguna dapat mengunggah, menonton, membagikan video, kehadiran Youtube memberikan dampak yang signifikan bagi masyarakat terutama bagi mereka yang gemar membuat video blog serta film pendek (Alwie, 2020). Sebagai platform berbasis video, Youtube menyediakan fitur interaksi langsung antara kreator dan penonton melalui tombol like, share dan komentar, yang menjadi sarana bagi penonton untuk menyampaikan pujian, kritik dan saran. Dengan demikian, Youtube tidak hanya menjadi platform pasif, tetapi juga sebagai wadah untuk berbagi opini dan ide secara langsung.

Salah satu konten yang banyak diminati adalah acara memasak yang diadaptasi dari Inggris, yaitu Masterchef Indonesia, yang disiarkan di televisi indonesia dan juga dapat ditonton melalui *platform*

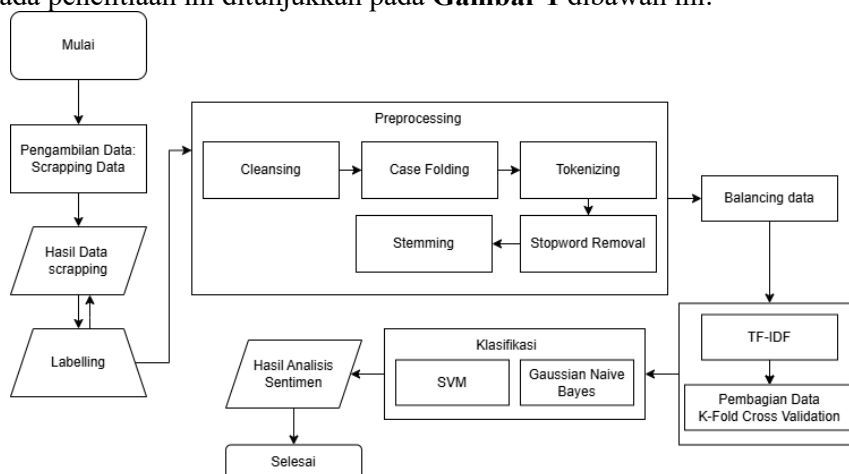
video-sharing Youtube (Khodijah dkk, 2021). Penayangan Mastechef Indonesia di Youtube bagian dari strategi untuk menjangkau lebih banyak penonton, terutama mereka yang sering mengakses platform digital daripada menonton televisi. Masterchef di *platform* Youtube menayangkan ulang video-video terkait Masterchef Indonesia, seperti cuplikan acara memasak, wawancara dengan peserta. Ulasan pada setiap episode sering kali menarik perhatian dari penggemar. Masterchef Indonesia berjalan selama 11 musim dari tahun 2011 sampai 2023 dengan 260 episode. Namun, pada *season* 11 episode *grand final* yang ditayangkan pada akhir pekan. 25-26 November 2023, muncul drama dan kontroversi tentang kemenangan peserta. kontroversi ini membuat publik dibuat terprovokasi, terutama dengan musim 11 dengan kehadiran sosok calon presiden dari PDI Perjuangan Ganjar Pranowo. Pada episode ini mempertemukan 2 *grand finalis*, Rizkisyah dengan Belinda. Belinda berhasil meraih kemenangan dengan menerima sebuah mobil secara simbolik yang diserahkan oleh Ganjar Pranowo.

Respon, opini, pendapat dan reaksi terhadap video dapat dituangkan melalui fitur yang disediakan Youtube yaitu komentar yang merupakan tempat mengekspresikan suara bagi pengguna dalam menikmati sebuah video yang ditonton dan dalam komentar tersebut bersifat positif, negatif dan netral (Mualfah dkk, 2023). Untuk mendapatkan data komentar oleh karena itu, diperlukan pengolahan data melalui analisis komentar pada video Youtube. Analisis sentimen terhadap komentar-komentar ini dapat memberikan wawasan tentang bagaimana persepsi masyarakat terhadap berbagai aspek acara, untuk mendapatkan data komentar diperlukan pengolahan data melalui analisis komentar pada video Youtube.

Analisis sentimen merupakan proses menganalisis teks digital untuk menentukan nilai positif, negatif atau netral. Analisis sentimen adalah metode untuk memahami, mengambil, dan memproses informasi teks secara otomatis untuk mengidentifikasi pernyataan emosional dalam menyampaikan pemikiran (Gifari dkk, 2022). Sentimen analisis digunakan untuk menilai pandangan atau kesepakatan opini terhadap suatu masalah atau objek, cenderung positif atau negatif. Berdasarkan dari permasalahan diatas adanya pro dan kontra terkait video Youtube Masterchef Indonesia season 11, penulis ingin melakukan analisis sentimen pada komentar youtube Masterchef Indonesia Season 11 dengan menerapkan klasifikasi analisis sentimen komentar Youtube dengan menggunakan algoritma *Support Vector Machine* dan *Gaussian Naïve Bayes*.

2. METODE

Adapun alur pada penelitian ini ditunjukkan pada **Gambar 1** dibawah ini:



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Pengumpulan data ini didapatkan dengan cara scrapping data pada kometar Youtube Masterchef Indonesia, setelah mendapatkan data hasil *scrapping*, selanjutnya akan dilakukan tahap pelabelan secara manual oleh ahli NLP (*Natural language processing*). Dan hasil validasi sentimen menunjukkan bahwa dari total 2000 data terdapat 1175 komentar dengan sentimen negatif dan 825 komentar dengan sentimen positif.

B. Preprocessing

Preprocessing adalah langkah awal yang akan dilakukan dalam pemrosesan teks. Ada beberapa tahapan *preprocessing* seperti berikut (Hudha dkk, 2022):

1. *Cleansing*
Tahapan ini merupakan tahapan menghapus karakter yang bukan alfabet seperti contoh karakter seperti angka, #, @, emoji dan link.
2. *Case Folding*
Tahapan yang melibatkan setiap huruf dalam sebuah paragraf atau kata menjadi huruf kecil semua. Karena untuk menyamakan format huruf sehingga memudahkan dalam proses analisis teks
3. *Tokenizing*
Pada tahapan ini, merupakan proses memisahkan kalimat menjadi beberapa bagian kata
4. *Stopword Removal*
Pada tahapan ini, dilakukan proses untuk menghapus kata- kata yang dianggap tidak berpengaruh dan tidak penting pada kalimat atau biasa disebut dengan *filtering*.
5. *Stemming*
Pada tahap ini, merupakan proses pengembalian kata menjadi kata dasar dengan membuang awalan, sisipan, imbuhan, atau akhiran.

C. *Balancing Data*

Balancing data adalah proses untuk menyeimbangkan jumlah sampel antar kelas mayoritas dan minoritas dalam dataset yang tidak seimbang. Ketidakseimbangan data terjadi Ketika salah satu kelas memiliki jumlah sampel lebih besar dibandingkan kelas lainnya, sehingga dapat membuat model cenderung berpihak pada kelas mayoritas (Al-Areef & Saputra S, 2023). Terdapat beberapa metode untuk mengatasi ketidakseimbangan data diantaranya *Random Undersampling*, *Random Oversampling* dan *SMOTE* (Almie dkk, 2022).

D. TF-IDF

TF-IDF merupakan singkatan dari "*Term Frequency-Inverse Document Frequency*" atau metode statistik yang digunakan dalam pemrosesan teks dalam mengevaluasi seberapa penting suatu kata pada sebuah dokumen. Berikut merupakan rumus yang digunakan untuk menghitung nilai *TF* (Derryl Qinanda dkk, 2022) :

$$tf_{t,d} = \frac{n_{t,d}}{N} \quad (1)$$

Keterangan:

$tf_{t,d}$: frekuensi kemunculan kata dalam suatu dokumen

$n_{t,d}$: nilai istilah yang muncul

N : semua *term* pada dokumen

Berikut rumus *idf* (Derryl Qinanda dkk, 2022) :

$$idf_d = \log \left(\frac{n_d}{df} \right) \quad (2)$$

Keterangan:

idf_d : jumlah kemunculan pada suatu kata pada dokumen

N_d : banyaknya total dokumen

df : banyak dokumen yang mengandung *term*

Pembobotan kata TF-IDF dengan rumus sebagai berikut (Derryl Qinanda dkk, 2022):

$$tfidf_{t,d} = tf_{t,d} \times idf_d \quad (3)$$

Keterangan:

$tfidf_{t,d}$: *term-Frequency-Inverse Document Frequency*

$tf_{t,d}$: nilai TF

idf_d : nilai IDF

E. *K-Fold Cross Validation*

K-Fold Cross Validation merupakan teknik evaluasi kinerja model yang digunakan dalam pembelajaran mesin untuk mengukur seberapa baik model tersebut akan bekerja diluar dataset yang digunakan untuk melatihnya (Wijiyanto dkk, 2024). Dengan metode ini, data dibagi menjadi beberapa bagian, dimana setiap bagian bergantian sebagai set uji, sementara bagian lainnya menjadi set latih.

Sehingga K menentukan jumlah pemisah dan pemilahan data yang digunakan untuk pembagian train dan test (Ridwansyah, 2022). Adapun rumus Kfold sebagai berikut (Indra Pratama dkk, 2023):

$$D = \frac{n}{k} \quad (4)$$

Keterangan:

D : hasil pembagian data pada dokumen
n : jumlah data
k : nilai *k-fold cross validation*

F. Support Vector Machine

Support Vector Machine (SVM) adalah suatu teknik pengklasifikasi data atau regresi yang menggunakan data-data sebelumnya dan pemodelannya disupervisi terlebih dahulu, prinsip utama metode *Support Vector Machine* adalah mengoptimalkan batas hyperplane yang memisahkan suatu set data (Rina Noviana & Isram Rasal, 2023). Berikut tahapan metode *Support Vector Machine* (Tineges dkk, 2020):

1. Menentukan kata yang sering muncul dan kemudian menentukan inisialisasi parameter $a=0.5$, $C=1$, $\lambda = 0.5$, $\gamma = 0.5$, $\epsilon = 0.001$.
2. Langkah selanjutnya dengan melakukan Perhitungan Matriks Hessian dengan persamaan menurut (Tineges dkk, 2020) sebagai berikut:

$$D_{ij} = \gamma_i \gamma_j (K(x_i, x_j))^2 + \lambda^2 \quad (5)$$

Keterangan:

y_i = kelas data ke- i
 y_j = kelas data ke- j
 λ = turunan bata teoritis
 D_{ij} = matriks data ke- ij
 $K(x_i, x_j)$ = fungsi kernel

3. Langkah selanjutnya dengan menghitung nilai dari *error*, *delta alfa*, dan *alfa* dari data i hingga data menurut (Tineges dkk, 2020) sebagai berikut:

$$E_i = \sum_{j=1}^i a_j D_{ij} \quad (6)$$

$$\delta a_i = \min(\max[\gamma(1 - E_i), a_i], C - a_i) \quad (7)$$

$$a_i = a_i + \delta a_i \quad (8)$$

keterangan:

a_j = alfa ke- j
 E_i = nilai error
 $\max_{(i)} D_{ij}$ = nilai maksimum diagonal matriks

4. untuk melakukan pencarian nilai bias bisa menggunakan rumus persamaan menurut (Tineges dkk, 2020) sebagai berikut:

$$b = -\frac{1}{2} (w \cdot x^+ \cdot w \cdot x^-) \quad (9)$$

5. Menguji pada dokumen yang akan diuji.
6. perhitungan keputusan.

$$h(x) = \begin{cases} +1, & \text{if } w \cdot x + b \geq 0 \\ -1, & \text{if } w \cdot x + b \leq 0 \end{cases} \quad (10)$$

7. jika hasil dari $\text{sign } h(x)$ dalam perhitungan keputusan lebih besar atau sama dengan 0, maka nilai tersebut adalah +1 dan termasuk dalam kelas positif. Sebaliknya, jika hasil dari $\text{sign } h(x)$ dalam perhitungan keputusan kurang dari 0, maka nilainya adalah -1 dan termasuk dalam kelas negatif. Persamaan ini digunakan untuk menentukan nilai keputusan (Tineges dkk, 2020).

$$h(x) = w \cdot x + b \quad (11)$$

atau

$$h(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b \quad (12)$$

G. Gaussian Naïve Bayes

Gaussian Naïve Bayes adalah teknik klasifikasi pembelajaran mesin berdasarkan pendekatan probablistik yang mengasumsikan setiap kelas mengikuti distribusi normal. Dalam *Gaussian Naïve*

Bayes, setiap fitur dianggap mengikuti distribusi normal, dan parameter distribusi (mean dan varians) dihitung dari data pelatihan (Naufal dkk, 2024). Berikut merupakan rumus pada *teorema bayes* menurut (Rayuwati dkk, 2022) sebagai berikut:

1. Mencari nilai mean dan standart deviasi dari masing – masing parameter. Untuk menghitung nilai rata – rata (mean) menggunakan rumus sebagai berikut (Rayuwati dkk, 2022) :

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (13)$$

atau

$$\mu = \frac{x_1 + x_2 + x_3 \dots + x_n}{n} \quad (14)$$

Keterangan:

μ = rata – rata hitung (mean)

x_i = nilai sampel ke -i

n = jumlah sampel

2. Kemudian untuk menghitung nilai standart deviasi dapat menggunakan rumus sebagai berikut (Rayuwati , 2022) :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (15)$$

Keterangan:

σ = standart deviasi

μ = nilai rata-rata hitung (mean)

x_i = nilai x ke-i

n = jumlah sampel

3. Langkah selanjutnya dengan menghitung nilai gaussian dengan rumus sebagai berikut (Rayuwati dkk, 2022) :

$$P(x|y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (16)$$

Keterangan:

P = peluang

x = nilai x

y = kelas yang dicari

σ = standart deviasi

μ = nilai rata-rata hitung (mean)

4. Langkah selanjutnya dengan perhitungan keputusan, apakah nilai tersebut masuk ke dalam kelas positif atau negatif, dengan rumus sebagai berikut (Rayuwati dkk, 2022) :

$$P(y|X) = P(y) \cdot \prod_{i=1}^n P(x_i|y) \quad (17)$$

Keterangan:

$P(y)$ = Prior Probability (kelas Positif atau Negatif)

$P(x_i|y)$ = Likelihood untuk setiap fitur x_i

n = jumlah fitur

$\prod$ = Perkalian semua likelihood fitur

3. HASIL DAN PEMBAHASAN

A. Preprocessing

1. *Cleansing*

Cleansing merupakan tahapan untuk menghapus karakter seperti i angka, #, @, emoji dan link. Adapun hasil dari *cleansing* dapat ditunjukkan pada **Tabel 1** dibawah ini:

Tabel 1. Hasil *Cleansing*

No	Komentar	<i>Cleansing</i>
1	SAYA KECEWA. PAYAHHHHHHHHHH... DARIMANYA BELINDA BISA WIN LAWAN KIKI...	SAYA KECEWA PAYAHHHHHHHHHH DARIMANYA BELINDA BISA WIN LAWAN KIKI

2	Terlalu blunder buat tayangan.... Saya sebagai penonton setia kecewa berat terliat sekali ngawurnya dari ronde pertama itu komen juri lebih baik ke kiki	Terlalu blunder buat tayangan Saya sebagai penonton setia kecewa berat terliat sekali ngawurnya dari ronde pertama itu komen juri lebih baik ke kiki
...	...	...
2000	Selamat kepada belinda pemenang masterchef	Selamat kepada belinda pemenang masterchef

2. Case Folding

Case Folding merupakan tahapan untuk merubah hirif besar menjadi huruf kecil semua. Adapun hasil dari *Case Folding* dapat ditunjukkan pada **Tabel 2** dibawah ini:

Tabel 2. Hasil *Case Folding*

No	Komentar	Case Folding
1	SAYA KECEWA. PAYAHHHHHHHHHH... DARIMANANYA BELINDA BISA WIN LAWAN KIKI...	saya kecewa payahhhhhhhhhh darimananya belinda bisa win lawan kiki
2	Terlalu blunder buat tayangan.... Saya sebagai penonton setia kecewa berat terliat sekali ngawurnya dari ronde pertama itu komen juri lebih baik ke kiki	terlalu blunder buat tayangan saya sebagai penonton setia kecewa berat terliat sekali ngawurnya dari ronde pertama itu komen juri lebih baik ke kiki
...	...	...
2000	Selamat kepada belinda pemenang masterchef	selamat kepada belinda pemenang masterchef

3. Tokenizing

Tokenizing merupakan proses memisahkan kalimat menjadi perkata. Adapun hasil dari *Tokenizing* dapat ditunjukkan pada **Tabel 3** dibawah ini:

Tabel 3. Hasil *Tokenizing*

No	Komentar	Tokenizing
1	SAYA KECEWA. PAYAHHHHHHHHHH... DARIMANANYA BELINDA BISA WIN LAWAN KIKI...	['saya', 'kecewa', 'payahhhhhhhhhh', 'darimananya', 'belinda', 'bisa', 'win', 'lawan', 'kiki']
2	Terlalu blunder buat tayangan.... Saya sebagai penonton setia kecewa berat terliat sekali ngawurnya dari ronde pertama itu komen juri lebih baik ke kiki	['terlalu', 'blunder', 'buat', 'tayangan', 'saya', 'sebagai', 'penonton', 'setia', 'kecewa', 'berat', 'terliat', 'sekali', 'ngawurnya', 'dari', 'ronde', 'pertama', 'itu', 'komen', 'juri', 'lebih', 'baik', 'ke', 'kiki']
...	...	...
2000	Selamat kepada belinda pemenang masterchef	['selamat', 'kepada', 'belinda', 'pemenang', 'masterchef']

4. Stopword Removal

Stopword Removal merupakan proses menghapus kata yang dianggap tidak penting pada kalimat. Adapun hasil dari *Stopword Removal* ditunjukkan pada **Tabel 4** dibawah ini:

Tabel 4. Hasil *Stopword Removal*

No	Komentar	Stopword Removal
1	SAYA KECEWA. PAYAHHHHHHHHHH... DARIMANANYA BELINDA BISA WIN LAWAN KIKI...	['kecewa', 'payahhhhhhhhhh', 'darimananya', 'belinda', 'win', 'lawan', 'kiki']
2	Terlalu blunder buat tayangan.... Saya sebagai penonton setia kecewa berat terliat sekali ngawurnya dari ronde pertama itu komen juri lebih baik ke kiki	['blunder', 'tayangan', 'penonton', 'setia', 'kecewa', 'berat', 'terliat', 'ngawurnya', 'ronde', 'komen', 'juri', 'kiki']
...	...	...
2000	Selamat kepada belinda pemenang masterchef	['selamat', 'belinda', 'pemenang', 'masterchef']

5. Stemming

Stemming merupakan proses mengembalikan kata menjadi kata dasar dengan membuang awalan, imbuhan dan akhiran. Adapun hasil dari *Stemming* ditunjukkan pada **Tabel 5** dibawah ini:

Tabel 5. Hasil *Stemming*

No	Komentar	Stemming
1	SAYA KECEWA. PAYAHHHHHHHHHH... DARIMANANYA BELINDA BISA WIN LAWAN KIKI...	kecewa payahhhhhhhhhh darimana belinda win lawan kiki
2	Terlalu blunder buat tayangan.... Saya sebagai penonton setia kecewa berat terliat sekali ngawurnya dari ronde pertama itu komen juri lebih baik ke kiki	blunder tayang tonton setia kecewa berat liat ngawurnya ronde komen juri kiki
3	sorry dah ga mood nonton pemenang pun disetting.. kalo drama doang yg setingan si g masalah	sorry dah ga mood nonton menang disetting kalo drama doang yg ting si g
...	...	...
2000	Selamat kepada belinda pemenang masterchef	selamat belinda menang masterchef

B. Pengujian Sebelum *Balancing*

Pada penelitian ini menggunakan *K-fold Cross Validation* nilai K yaitu 2,5,10. Dengan menggunakan 2000 dataset, dan pada tahap ini akan menggunakan 2 skenario pengujian dengan menggunakan algoritma *Gaussian Naïve Bayes* dan *Support Vector Machine* yang kemudian akan diketahui tingkat akurasi pada masing-masing pengujian. Berikut adalah hasil pengujian menggunakan algoritma *Gaussian Naïve Bayes* dan *Support Vector Machine* dengan mengukur akurasi, presisi dan *recall* yang didapatkan pada pembagian masing-masing *fold*.

1. *Support Vector Machine* sebelum *Balancing*

Adapun hasil dari pengujian *Support Vector Machine* sebelum *Balancing* ditunjukkan pada Tabel 6 dibawah ini:

Tabel 6. Hasil pengujian *Support Vector Machine* sebelum *Balancing*

K-Fold Cross	Langkah Uji	<i>Support Vector Machine</i>		
		Akurasi	Presisi	Recall
2	1	74%	65%	82%
	2	81%	82%	71%
5	1	78%	69%	81%
	2	74%	65%	78%
	3	73%	63%	81%
	4	75%	67%	71%
	5	81%	85%	72%
10	1	79%	73%	75%
	2	80%	70%	90%
	3	74%	67%	79%
	4	75%	64%	77%
	5	74%	67%	79%
	6	72%	60%	79%
	7	71%	63%	67%
	8	68%	57%	77%
	9	76%	70%	79%
	10	82%	88%	72%

2. *Gaussian Naïve Bayes* sebelum *Balancing*

Adapun hasil dari pengujian *Gaussian Naïve Bayes* sebelum *Balancing* ditunjukkan pada Tabel 7 dibawah ini:

Tabel 7. Hasil pengujian *Gaussian Naïve Bayes* sebelum *Balancing*

K-Fold Cross	Langkah Uji	<i>Gaussian Naïve Bayes</i>		
		Akurasi	Presisi	Recall
2	1	61%	52%	76%
	2	63%	54%	77%
	1	62%	52%	84%

5	2	61%	51%	83%
	3	60%	51%	78%
	4	60%	50%	78%
	5	61%	55%	76%
10	1	60%	50%	83%
	2	60%	51%	83%
	3	60%	52%	85%
	4	61%	49%	82%
	5	57%	50%	76%
	6	65%	52%	81%
	7	60%	49%	79%
	8	59%	50%	76%
	9	64%	56%	79%
	10	58%	53%	78%

Berdasarkan pengujian yang telah dilakukan dengan menggunakan *Support Vector Machine* dan *Gaussian Naïve Bayes*, diperoleh rekapitulasi hasil sebelum *Balancing data* ditunjukkan pada **Tabel 8** seperti berikut:

Tabel 8. Rangkuman hasil sebelum *Balancing*

<i>K-Fold Cross</i>	Langkah Uji	<i>Gaussian Naïve Bayes</i>			<i>Support Vector Machine</i>		
		Akurasi	Presisi	<i>Recall</i>	Akurasi	Presisi	<i>Recall</i>
2	1	61%	52%	76%	74%	65%	82%
	2	63%	54%	77%	81%	82%	71%
5	1	62%	52%	84%	78%	69%	81%
	2	61%	51%	83%	74%	65%	78%
	3	60%	51%	78%	73%	63%	81%
	4	60%	50%	78%	75%	67%	71%
	5	61%	55%	76%	81%	85%	72%
10	1	60%	50%	83%	79%	73%	75%
	2	60%	51%	83%	80%	70%	90%
	3	60%	52%	85%	74%	67%	79%
	4	61%	49%	82%	75%	64%	77%
	5	57%	50%	76%	74%	67%	79%
	6	65%	52%	81%	72%	60%	79%
	7	60%	49%	79%	71%	63%	67%
	8	59%	50%	76%	68%	57%	77%
	9	64%	56%	79%	76%	70%	79%
	10	58%	53%	78%	82%	88%	72%

Dalam klasifikasi komentar YouTube dengan data sebanyak 2000, algoritma *Gaussian Naïve Bayes* menghasilkan nilai akurasi tertinggi sebesar 65%, dengan presisi sebesar 52%, dan *recall* sebesar 81% pada uji ke-6 dari 10-*old cross-validation*. Sementara itu, algoritma *Support Vector Machine* mencapai nilai akurasi tertinggi sebesar 82%, dengan presisi sebesar 88%, dan *recall* sebesar 72% pada uji ke-10 dari 10-*fold cross-validation*.

C. Pengujian Setelah *Balancing*

Pada penelitian ini menggunakan *K-fold Cross Validation* nilai K yaitu 2,5,10. Dengan menggunakan 2000 dataset, dan menggunakan algoritma *Gaussian Naïve Bayes* dan *Support Vector Machine* yang kemudian akan diketahui tingkat akurasi pada masing-masing pengujian. Berikut adalah hasil pengujian menggunakan algoritma *Gaussian Naïve Bayes* dan *Support Vector Machine* dengan mengukur akurasi, presisi dan *recall* yang didapatkan pada pembagian masing-masing *fold*.

1. *Support Vector Machine* setelah *Balancing*

Adapun hasil dari pengujian *Support Vector Machine* sebelum *Balancing* ditunjukkan pada **Tabel 9** dibawah ini:

Tabel 9. Hasil pengujian *Support Vector machine* setelah *Balancing*

<i>K-Fold Cross</i>	Langkah Uji	<i>Support Vector Machine</i>		
		Akurasi	Presisi	<i>Recall</i>
2	1	75%	70%	85%
	2	83%	84%	81%
5	1	80%	76%	86%
	2	83%	79%	88%
	3	79%	76%	89%
	4	81%	78%	87%
	5	85%	82%	88%
10	1	77%	73%	84%
	2	81%	78%	86%
	3	79%	72%	91%
	4	84%	79%	92%
	5	79%	76%	89%
	6	85%	84%	89%
	7	80%	76%	84%
	8	83%	78%	95%
	9	79%	69%	96%
	10	84%	85%	87%

2. *Gaussian Naïve Bayes* setelah *Balancing*

Adapun hasil dari pengujian *Support Vector Machine* sebelum *Balancing* ditunjukkan pada **Tabel 10** dibawah ini:

Tabel 10. Hasil pengujian *Gaussian Naïve Bayes* setelah *Balancing*

<i>K-Fold Cross</i>	Langkah Uji	<i>Gaussian Naïve Bayes</i>		
		Akurasi	Presisi	<i>Recall</i>
2	1	65%	60%	85%
	2	66%	61%	87%
5	1	70%	64%	87%
	2	68%	61%	91%
	3	68%	65%	86%
	4	66%	62%	88%
	5	68%	62%	90%
10	1	70%	63%	91%
	2	69%	64%	86%
	3	66%	59%	90%
	4	71%	63%	95%
	5	71%	68%	87%
	6	69%	64%	89%
	7	64%	58%	88%
	8	72%	68%	93%
	9	66%	57%	92%
	10	69%	67%	90%

Berdasarkan pengujian yang telah dilakukan dengan menggunakan *Support Vector Machine* dan *Gaussian Naïve Bayes*, diperoleh rekapitulasi hasil setelah *Balancing data* ditunjukkan pada **Tabel 11** seperti berikut:

Tabel 11. Rangkuman hasil setelah *Balancing*

<i>K-Fold Cross</i>	Langkah Uji	<i>Gaussian Naïve Bayes</i>			<i>Support Vector Machine</i>		
		Akurasi	Presisi	<i>Recall</i>	Akurasi	Presisi	<i>Recall</i>
2	1	65%	60%	85%	75%	70%	85%
	2	66%	61%	87%	83%	84%	81%
5	1	70%	64%	87%	80%	76%	86%
	2	68%	61%	91%	83%	79%	88%
	3	68%	65%	86%	79%	76%	89%
	4	66%	62%	88%	81%	78%	87%
	5	68%	62%	90%	85%	82%	88%
10	1	70%	63%	91%	77%	73%	84%
	2	69%	64%	86%	81%	78%	86%
	3	66%	59%	90%	79%	72%	91%
	4	71%	63%	95%	84%	79%	92%
	5	71%	68%	87%	79%	76%	89%
	6	69%	64%	89%	85%	84%	89%
	7	64%	58%	88%	80%	76%	84%
	8	72%	68%	93%	83%	78%	95%
	9	66%	57%	92%	79%	69%	96%
	10	69%	67%	90%	84%	85%	87%

Dalam klasifikasi komentar YouTube dengan data sebanyak 2000, dengan teknik *balancing* data algoritma *Gaussian Naïve Bayes* menghasilkan nilai akurasi tertinggi sebesar 72%, dengan presisi sebesar 68%, dan *recall* sebesar 93% pada uji ke-8 dari 10-fold cross-validation. Sementara itu, algoritma *Support Vector Machine* mencapai nilai akurasi tertinggi sebesar 85%, dengan presisi sebesar 84%, dan *recall* sebesar 89% pada uji ke-6 dari 10-fold cross-validation.

Hasil yang diperoleh dari penerapan sebelum dan sesudah *balancing* data pada algoritma *Support Vector Machine* dan *Gaussian Naïve Bayes* menunjukkan bahwa penggunaan teknik oversampling untuk menangani data yang tidak seimbang dapat meningkatkan kinerja dalam memprediksi atau mengklasifikasikan sentimen kelas minoritas (positif dan negatif). Pengujian tersebut membuktikan bahwa penerapan teknik *balancing oversampling* dapat meningkatkan kinerja algoritma *Support Vector Machine* dan *Gaussian Naïve Bayes* dan lebih efektif jika dibandingkan dengan klasifikasi tanpa *balancing*.

4. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan penelitian yang sudah dilakukan terhadap komentar Youtube Masterchef Indonesia menggunakan algoritma *Gaussian Naïve Bayes* dan *Support Vector Machine*. Dapat diambil beberapa kesimpulan sebagai berikut:

1. Data penelitian yang digunakan sebanyak 2000 komentar Youtube dengan masing-masing komentar positif memperoleh 825 data yang berisi dukungan serta ucapan terhadap peserta dan komentar negatif sebanyak 1175 data yang berisi kekecewaan terhadap juri serta *channel* Masterchef Indonesia.
2. Pada penelitian ini, algoritma *Gaussian Naïve Bayes* menunjukkan performa terbaik berdasarkan data normal pada pengujian *K-fold* 10 langkah Uji 6 dengan akurasi tertinggi sebesar 65%, presisi sebesar 52% dan *recall* sebesar 81%. Kemudian dengan algoritma *Support Vector Machine* didapatkan nilai akurasi tertinggi sebesar 82% pada *K-fold* 10 Langkah Uji 10 dan akurasi terendah sebesar 71%, presisi sebesar 88% dan *recall* sebesar 72% pada *K-fold* 10 Langkah Uji 7.
3. Pada penelitian ini, algoritma *Gaussian Naïve Bayes* menunjukkan performa terbaik berdasarkan *Balancing Data Gaussian Naïve Bayes* pada pengujian *K-Fold* 10 langkah uji 8 dengan akurasi 72%, presisi 68%, *recall* 93%. Kemudian *Balancing Data Support Vector Machine* didapatkan nilai akurasi 85%, presisi 84% dan *recall* 89% pada *K-Fold* 10 langkah uji 6.

B. Saran

Berdasarkan penelitian yang telah dilakukan, penulis dapat memberikan beberapa saran yang dapat dikembangkan dipenelitian selanjutnya, sebagai berikut:

1. Apabila komentar positif dan negatif tidak seimbang, disarankan untuk mencoba teknik *Balancing Data* lain seperti metode *Random Undersampling* dan metode SMOTE.
2. Dianjurkan untuk melakukan *hyperparameter tuning* lebih lanjut pada algoritma *Support Vector Machine* dengan menggunakan metode *Grid Search* guna menemukan kombinasi parameter yang lebih optimal.

DAFTAR PUSTAKA

- Al-Areef, M. H., & Saputra S, K. (2023). Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 22(2), 270.
- Almie, A. U., Nilogiri, A., & Wardoyo, A. E. (2022). Klasifikasi Sentimen Ulasan Pelanggan Layanan Ekspedisi Menggunakan Algoritma Naïve Bayes. *JASIE: Jurnal Aplikasi Sistem Informasi Dan Elektronika*, 4(1), 28–35.
- Alwie, rahayu deny danar dan alvi furwanti, Prasetio, A. B., Andespa, R., Lhokseumawe (2020). Klasifikasi Komentar Terhadap Polisi Pada Media Sosial Youtube Menggunakan Support Vector Machine. *Jurnal Ekonomi*, 2(1), 41–49.
- Derryl Qinanda, M., Nilogiri, A., & Timur, T. (2022). Sentimen Pada Komentar Youtube Tentang Pencegahan Dan Penanganan Kekerasan Seksual Pada Permendikbud Berbasis Naïve Bayes Dan Support Vector Machine. *Jurna Sistem dan Teknologi Informasi (JUSTINDO)*, 7(2), 114–121.
- Gifari, O. I., Adha, M., Freddy, F., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40.
- Hudha, M., Supriyati, E., & Listyorini, T. (2022). Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes Classifier. *JIKO (Jurnal Informatika Dan Komputer)*, 5(1), 1–6.
- Indra Pratama, J., Nilogiri, A., & Wardoyo, A. E. (2023). Klasifikasi Sentimen Terhadap Penyelenggaraan Haji Indonesia Tahun 2023 Melalui Media Sosial Twitter Menggunakan Teknik Balancing Dan Metode Gaussian Naïve Bayes. *Jurnal Smart Teknologi*, 5(1), 117–126.
- Khodijah, S., Hayuningtyas UIN KHAS Jember, N., & Hayuningtyas, N. (2021). Komunikasi Media Sosial Perspektif Sociolinguistik (Analisis Kode pada Program Master Chef Junior Indonesia di YouTube). *Indonesian Journal of Islamic Communication*, 4(2), 1–19.
- Mualfah, D., Ramadhoni, Gunawan, R., & Mulyadipa Suratno, D. (2023). Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM. *Jurnal Fasikom*, 13(01), 72–80.
- Naufal, Y., Putro, R., Afriansyah, A., & Bagaskara, R. (2024). Penggunaan Algoritma Gaussian Naïve Bayes & Decision Tree Untuk Klasifikasi Tingkat Kemenangan Pada Game Mobile Legends. *Jurnal Komputer dan Informatika (JUKI)*, 6(Mei), 10–26.
- Rayuwati, Husna Gemasih, & Irma Nizar. (2022). Implementasi Algoritma Naïve Bayes Untuk Memprediksi Tingkat Penyebaran Covid. *Jurnal Riset Rumpun Ilmu Teknik*, 1(1), 38–46.
- Ridwansyah, T. (2022). Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 2(5), 178–185.
- Rina Noviana, & Isram Rasal. (2023). Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter. *Jurnal Teknik Dan Science*, 2(2), 51–60.
- Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *Jurnal Media Informatika Budidarma*, 4(3), 650.
- Wijiyanto, W., Pradana, A. I., Sopingi, S., & Atina, V. (2024). Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa. *Jurnal Algoritma*, 21(1), 239–248.