



Penerapan Algoritma Machine Learning SVM dan NBC pada Sentimen Analisis Komentar Youtube Program Pengaduan Masyarakat Lapor Mas Wapres

Muhammad Bais Al Hakiki^{1*}, Yulia Darmi²

¹Universitas Muhammadiyah Bengkulu, Bengkulu, Indonesia, baiskiki0@gmail.com

²Universitas Muhammadiyah Bengkulu, Bengkulu, Indonesia, yuliadarmi10juli@gmail.com

*Corresponding Author: baiskiki0@gmail.com¹

Abstract: The public complaint program “Lapor Mas Wapres” initiated by Vice President Gibran Rakabuming Raka has garnered public attention for facilitating direct communication between the government and citizens. This study aims to evaluate public responses to the program by analyzing sentiment in YouTube comments using two machine learning algorithms: Support Vector Machine (SVM) and Naïve Bayes Classifier (NBC). A total of 2,500 comments were collected from six YouTube videos through web crawling and were then processed through stages such as text cleaning, normalization, tokenization, and sentiment labeling. The performance of both algorithms was evaluated using accuracy, precision, recall, and f1-score metrics. The results showed that SVM outperformed NBC with an accuracy of 76.2%, while NBC achieved 59.8%. SVM also provided a more balanced classification across sentiment categories (positive, neutral, negative), whereas NBC tended to be biased toward positive sentiment. These findings suggest that SVM is more suitable for analyzing public opinion on social media platforms that feature informal and diverse language styles such as YouTube.

Keywords: Sentiment Analysis, YouTube, SVM, Naïve Bayes, Lapor Mas Wapres, Public Comments

Abstrak: Program pengaduan masyarakat “Lapor Mas Wapres” yang digagas oleh Wakil Presiden Gibran Rakabuming Raka menarik perhatian publik karena membuka kanal komunikasi langsung antara pemerintah dengan masyarakat. Penelitian ini bertujuan guna mengevaluasi respons masyarakat terhadap program tersebut melalui analisis sentimen pada komentar YouTube, dengan memanfaatkan dua algoritma machine learning yaitu Support Vector Machine (SVM) serta Naïve Bayes Classifier (NBC). Sebanyak 2500 komentar dikumpulkan dari enam video YouTube menggunakan metode crawling, kemudian diproses melalui tahapan pembersihan data, normalisasi, tokenisasi, dan pelabelan sentimen. Kinerja kedua algoritma diukur berdasarkan akurasi, presisi, recall, serta f1-score. Temuan penelitian menunjukkan bahwa SVM mempunyai performa klasifikasi yang lebih unggul dengan akurasi mencapai 76,2%, sementara NBC hanya mencapai akurasi 59,8%. SVM juga

menunjukkan distribusi klasifikasi yang lebih seimbang pada ketiga kategori sentimen (positif, netral, negatif), sedangkan NBC cenderung bias terhadap sentimen positif. Temuan ini menunjukkan bahwa algoritma SVM lebih sesuai untuk menganalisis opini publik pada platform media sosial yang memiliki ragam bahasa informal seperti YouTube.

Kata Kunci: Sentimen Analisis, YouTube, SVM, Naïve Bayes, Lapor Mas Wapres, Komentar Masyarakat

PENDAHULUAN

Layanan pengaduan masyarakat "Lapor Mas Wapres" yang digagas oleh Wakil Presiden Gibran Rakabuming Raka.[1] Belakangan ini, perhatian publik tertuju pada layanan pengaduan bertajuk "Lapor Mas Wapres". Layanan pengaduan tersebut diinisiasi oleh Wakil Presiden Republik Indonesia, Gibran Rakabuming Raka, yang secara resmi meluncurkannya di Istana Wakil Presiden, berlokasi di Jalan Kebon Sirih, Jakarta Pusat. Sejak tanggal 11 November 2024, tepatnya hari Senin, layanan ini telah mulai dioperasikan. Adapun waktu operasional pos pengaduan tersebut berlangsung setiap hari kerja, yaitu Senin hingga Jumat, mulai pukul 08.00 hingga pukul 14.00 WIB.[2]

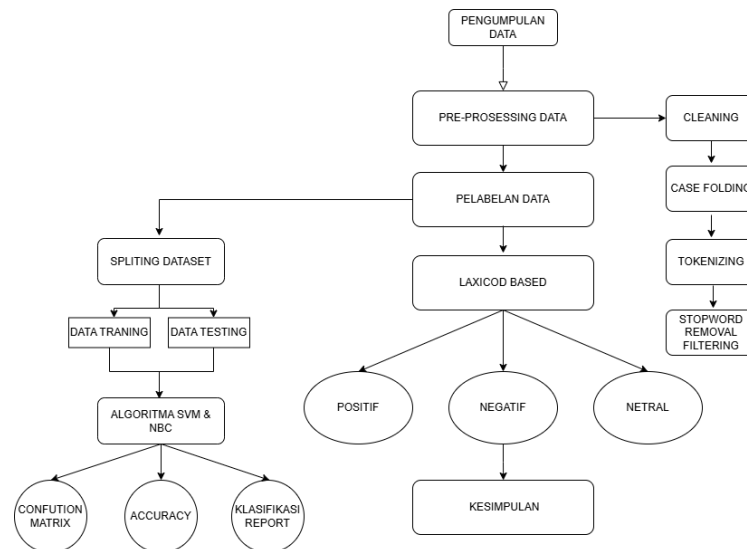
Analisis sentimen merupakan metode penting dalam mengkaji opini publik terhadap berbagai isu melalui data teks, seperti komentar di media sosial.[3] Algoritma *Support Vector Machine* (SVM) serta *Naïve Bayes Classifier* (NBC) sering digunakan dalam analisis ini karena efektivitasnya dalam klasifikasi teks. Beberapa penelitian telah membandingkan kinerja SVM dan NBC dalam analisis sentimen. Misalnya, studi oleh Muhammadin dan Sobari (2021) yang menganalisis ulasan aplikasi Kredivo menunjukkan bahwa SVM mencapai akurasi 83,3%, sementara NBC mencapai 80,8%, dengan SVM menunjukkan performa yang lebih unggul dalam klasifikasi sentimen positif dan negatif.[4] Demikian pula, penelitian oleh Azzahra et al. (2024) dalam menganalisis sentimen kebencanaan di YouTube menemukan bahwasannya SVM mempunyai akurasi 92%, lebih tinggi dibandingkan NBC yang mencapai 79%[5]

YouTube merupakan sebuah situs berbagi media audiovisual yang berfungsi sebagai salah satu bentuk hiburan virtual, di mana pengguna dapat membagikan konten berupa video maupun suara. Selain itu, YouTube juga dimanfaatkan sebagai sarana untuk menyampaikan berbagai informasi, ekspresi perasaan, serta opini mengenai beragam topik. Saat ini, YouTube telah berkembang menjadi salah satu platform paling populer dalam hal konsumsi konten video. Dengan demikian, tidaklah mudah untuk menentukan kecenderungan sikap pengguna internet terhadap situs YouTube, apakah bersifat positif, netral, atau justru negatif [6]

Dalam konteks program pengaduan masyarakat seperti "LAPOR MAS WAPRES", analisis sentimen terhadap komentar YouTube bisa memberikan wawasan berharga tentang persepsi publik terhadap layanan tersebut. Meskipun belum ditemukan penelitian spesifik mengenai "LAPOR MAS WAPRES", studi-studi sebelumnya menunjukkan bahwa penerapan algoritma SVM dan NBC efektif dalam menganalisis sentimen komentar YouTube terkait isu-isu publik. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma SVM dan NBC dalam analisis sentimen komentar YouTube terhadap program "LAPOR MAS WAPRES", guna memahami opini publik dan meningkatkan kualitas layanan pengaduan masyarakat.

METODE

Penelitian terkait analisis sentimen ini dilaksanakan melalui serangkaian tahapan yang bisa diamati pada gambar berikut ini.



Gambar 1. Tahapan penelitian.

1. Pengumpulan Data

Data dihimpun dari video *youtube* yang akan kita kumpulkan dengan *keyword* “lapor mas wapres” pada kanal *youtube channel* , langkah selanjutnya melakukan scrapping menggunakan *web APIFY*, yang dimana akan menghasilkan file dalam bentuk CSV.

2. Pre Proseessing Data

Proses pembersihan data teks dilakukan pada tahap ini dengan tujuan untuk memperoleh bentuk dasar dari setiap kata, yang nantinya akan digunakan dalam analisis lanjutan. Agar data dapat siap diolah pada tahapan selanjutnya, diperlukan beberapa proses dalam fase *preprocessing*. Di antara tahapan-tahapan dalam pengolahan teks tersebut meliputi *case folding*, *cleaning*, *tokenizing*, *slang word replacement*, *stop word removal*, dan *stemming*[7]

3. Naive Bayes Classifier

Algoritma *Naïve Bayes Classifier* adalah metode yang menerapkan konsep probabilitas bersyarat.[8]

Naïve Bayes Classifier merupakan algoritma klasifikasi yang sederhana, namun memiliki efisiensi yang tinggi dalam penerapannya. Algoritma ini berasal dari prinsip Bayesian terhadap permasalahan klasifikasi, di mana metode tersebut mengandalkan asumsi independensi antar atribut yang disederhanakan[9]

Pengukuran kinerja klasifikasi dengan menggunakan rumus akurasi dari matriks konfusi, sebagai berikut

Akurasi = $(\text{Jumlah benar} / \text{Jumlah data uji}) \times 100\%$ Klasifikasi adalah sebuah proses dalam ilmu data yang bertujuan untuk membentuk model atau fungsi yang mampu menjelaskan serta membedakan kategori atau kelas data yang ada. Model tersebut dikembangkan melalui proses analisis terhadap data pelatihan (*training data*). Sebelum pelaksanaan klasifikasi, biasanya dilakukan tahap analisis relevansi untuk mengidentifikasi atribut-atribut mana yang memiliki relevansi terhadap proses klasifikasi tersebut. Keakuratan hasil prediksi dari sebuah model klasifikasi dapat direpresentasikan dalam bentuk tabel kontingensi (*contingency table*) atau *confusion matrix* (Flach) seperti pada tabel berikut:

	Prediksi(+)	Prediksi(-)
Aktual (+)	TP (True Positives)	FN (False Negatives)
Aktual (-)	FP (False Positives)	TN (True Negatives)

Tabel 1

Keterangan:

TP: Jumlah kasus positif yang diklasifikasi sebagai positif

FP: Jumlah kasus negatif yang diklasifikasi sebagai positif

TN: Jumlah kasus negatif yang diklasifikasi sebagai negatif

FN: Jumlah kasus positif yang diklasifikasi sebagai negatif

Informasi mengenai keberhasilan pengklasifikasi diberikan oleh TP dan TN, sedangkan kesalahan pengklasifikasian diindikasikan oleh FP dan FN.

Untuk menilai performa suatu model klasifikasi, dapat digunakan sejumlah ukuran evaluatif, antara lain:

a. Akurasi

Akurasi merupakan indikator yang digunakan untuk mengukur sejauh mana tingkat ketepatan prediksi suatu model klasifikasi terhadap label kelas yang benar. Perhitungan akurasi klasifikasi dapat dilakukan dengan menggunakan rumus sebagai berikut:

$$\text{Akurasi} = \frac{\text{Jumlah klasifikasi benar}}{\text{jumlah total data}} = \frac{TP+TN}{TP+TN+FP+FN} = (1)$$

b. Precision

Precision merupakan indikator yang digunakan untuk menilai sejauh mana prediksi yang dilakukan oleh model tepat dalam mengklasifikasikan data ke dalam kelas tertentu. Rumus yang digunakan untuk menghitung precision adalah sebagai berikut:

$$\text{Precision} = \frac{TP}{TP+FP} = (2)$$

c. Sensitivitas

Sensitivitas merupakan ukuran untuk mengukur seberapa besar kemampuan suatu model klasifikasi dalam mengidentifikasi data yang benar dari suatu kelas secara akurat. Rumus yang digunakan dalam perhitungan sensitivitas adalah:

$$\text{Sensitivitas} = \frac{TP}{TP+FN} = (3)$$

d. Spesifitas

Spesifitas adalah ukuran yang digunakan untuk mengetahui tingkat keberhasilan model klasifikasi dalam mengenali data yang berasal dari kelas tertentu secara tepat. Rumus untuk menghitung

spesifitas klasifikasi sesuai persamaan berikut:

$$\text{Spesifitas} = \frac{TN}{TN+FP} = (4)$$

Untuk membandingkan performa dari berbagai algoritma klasifikasi, digunakan beberapa parameter sebagai berikut:

a. Test Mode: Mendefinisikan metode evaluasi yang digunakan, yaitu *cross-validation* dan *percentage split*, untuk mengukur akurasi pengujian.

- b. Time to build model: Istilah ini merujuk pada waktu yang dibutuhkan untuk menyusun model klasifikasi dengan masing-masing algoritma
- c. Correctly classified instances: Merupakan jumlah data yang dapat diklasifikasikan secara benar.
- d. Incorrectly classified instances: merujuk pada jumlah data yang tidak dapat diklasifikasikan dengan benar. [10]

4. Support Vector Machine (Svm)

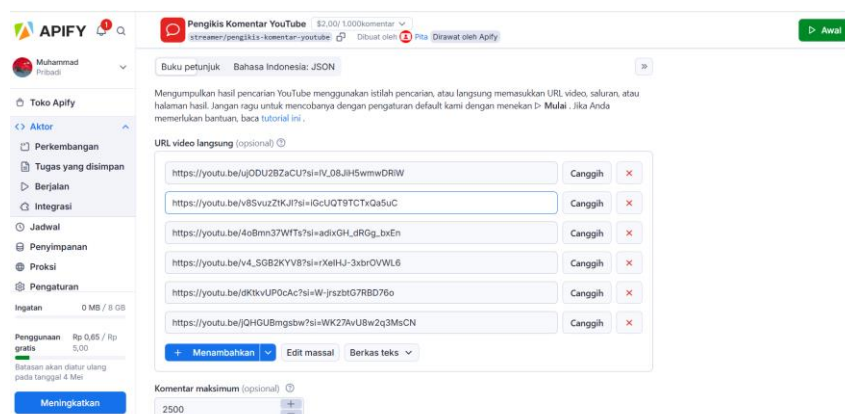
Support Vector Machine (SVM) pertama kali dikembangkan oleh Boser, Guyon, dan Vapnik untuk tujuan prediksi, baik dalam konteks klasifikasi maupun regresi, yang belakangan ini semakin populer. Pengembangan SVM dimulai pada tahun 1992 dalam acara Annual Workshop on Computational Learning Theory. SVM merupakan algoritma dalam bidang machine learning yang beroperasi berdasarkan prinsip Structural Risk Minimization (SRM), dengan tujuan utama untuk menemukan hyperplane optimal yang memisahkan dua kelas dalam ruang input. Dalam kategori supervised learning, SVM sebanding dengan ANN, di mana implementasinya mencakup tahap *training* yang diikuti dengan tahap *testing*. [11]

Proses klasifikasi diikuti dengan evaluasi terhadap model kinerja yang dihasilkan menggunakan teknik Support Vector Machine (SVM). Selama tahap evaluasi, Confusion Matrix digunakan sebagai alat untuk mengukur berbagai parameter seperti Accuracy, Precision, Recall, F1-Score, serta Error rate yang kemudian menghasilkan data yang dapat divalidasi. Confusion Matrix mencakup beberapa parameter, di antaranya TP (True Positive), yang mewakili sampel dengan nilai yang benar yang diklasifikasikan dengan benar; FN (False Negative), yang menunjukkan sampel dengan nilai yang benar namun diklasifikasikan salah; FP (False Positive), yang merujuk pada sampel dengan nilai yang salah namun diklasifikasikan sebagai benar; dan TN (True Negative), yang menggambarkan sampel dengan nilai yang salah yang diklasifikasikan dengan tidak benar. Akurasi dihitung menggunakan rumus yang tertera di bawah ini. [12]

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

HASIL DAN PEMBAHASAN

Dataset



Gambar 2. Crawling dataset dengan website apify

Dataset dikumpulkan dengan menggunakan website apify agar mendapatkan data komentar dalam bentuk CSV, yang dimana komentar youtube tersebut diambil pada 6 link

video youtube dengan keyword “lapor mas wapres” dengan mengumpulkan data komentar sebanyak 2500 *dataset*.

```
import pandas as pd

data = pd.read_csv("dataset_youtube-comments-scraper_2025-03-04_01-56-51-897.csv")
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2500 entries, 0 to 2499
Data columns (total 14 columns):
 #   Column              Non-Null Count  Dtype  
---  --
 0   author              2500 non-null   object  
 1   authorIsChannelOwner 2500 non-null   bool    
 2   cid                 2500 non-null   object  
 3   comment             2500 non-null   object  
 4   commentsCount       2500 non-null   int64   
 5   hasCreatorHeart     2500 non-null   bool    
 6   pageUrl             2500 non-null   object  
 7   publishedTimeText    2500 non-null   object  
 8   replyCount          2500 non-null   int64   
 9   replyToId           856 non-null    object  
10   title               2500 non-null   object  
11   type                2500 non-null   object  
12   videoId             2500 non-null   object  
13   voteCount           2500 non-null   int64   
dtypes: bool(2), int64(3), object(9)
memory usage: 239.4+ KB
```

Gambar 3. Hasil *crawling* dataset.

Menghapus Data Duplikat

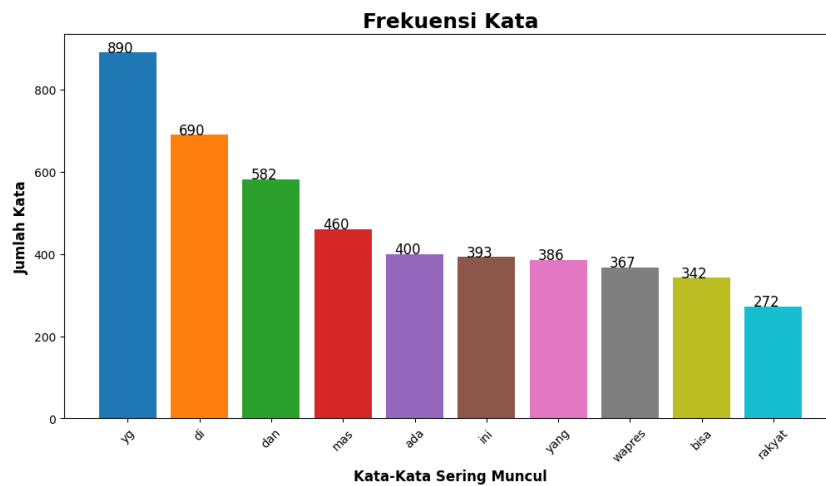
	author	comment
0	@GitoGati-gr6zz	Ide yang positif. Menuju era keterbukaan
1	@dsetia561	Ini baru yang namanya "Pejabat"ndigaji oleh r...
2	@ahmadjamil3557	Apresiasi luar biasa kepada wapres indonesia k...
3	@kotakualakapasaman6571	LUAR BIASA, sehat terus PA GIBRAN dan JAJARAN...
4	@youngshin7879	Semakin menyala mas Gibran pilihan ku ❤️❤️
5	@abassomantri5839	Mantap Ya. Perangkat dan Ketentuannya sudah di...
6	@umirudiat8231	Mantap ...jd semuanya bnar2 bekerja tdk ada yg...
7	@latifahmad	Gak salah pilih saya „paket komplit pak prabo...
8	@fernandosarindat3903	jangan mendengarkan mereka yang tidak suka 😊 tetap...
9	@darmawi-p8v	Wapres Gibran seorang pekerja lapangan yg sela...

Gambar 4. Hasil setelah menghapus Data Duplikat.

Wordcloud Sebelum di Preprocessing



Gambar 5. Wordcloud sebelum di preprocessing



Gambar 6. Frekuensi kata-kata yang sering muncul.

Cleaning Data

	author	comment	cleaning
0	@GitoGati-gr6zz	Ide yang positif. Menujui era keterbukaan	Ide yang positif Menujui era keterbukaan
1	@dsetia561	Ini baru yang namanya "Pejabat"ndigaji oleh r...	Ini baru yang namanya Pejabatndigaji oleh rak...
2	@ahmadjamil3557	Apresiasi luar biasa kepada wapres indonesia k...	Apresiasi luar biasa kepada wapres indonesia k...
3	@kotakualakapuasaman6571	LUAR BIASA, sehat terus PA GIBRAN dan JAJARANN...	LUAR BIASA sehat terus PA GIBRAN dan JAJARANNY...
4	@youngshin7879	Semakin menyala mas Gibran pilihan ku ♥♥	Semakin menyala mas Gibran pilihan ku

Tabel 2. Hasil dari *cleaning Data*.

Case Folding

	author	comment	cleaning	case_folding
0	@GitoGati-gr6zz	Ide yang positif. Menujui era keterbukaan	Ide yang positif Menujui era keterbukaan	ide yang positif menujui era keterbukaan
1	@dsetia561	Ini baru yang namanya "Pejabat"ndigaji oleh r...	Ini baru yang namanya Pejabatndigaji oleh rak...	ini baru yang namanya pejabatndigaji oleh rak...
2	@ahmadjamil3557	Apresiasi luar biasa kepada wapres indonesia k...	Apresiasi luar biasa kepada wapres indonesia k...	apresiasi luar biasa kepada wapres indonesia k..
3	@kotakualakapuasaman6571	LUAR BIASA, sehat terus PA GIBRAN dan JAJARANN...	LUAR BIASA sehat terus PA GIBRAN dan JAJARANNY...	luar biasa sehat terus pa gibran dan jajaranny...
4	@youngshin7879	Semakin menyala mas Gibran pilihan ku ♥♥	Semakin menyala mas Gibran pilihan ku	semakin menyala mas gibran pilihan ku

Tabel 3. Hasil dari *case folding*

Normalisasi Kata

	author	comment	cleaning	case_folding	hasil_normalisasi
0	@GitoGati-gr6zz	Ide yang positif. Menujui era	Ide yang positif Menujui era	ide yang positif	ide yang positif menujui era

		keterbukaan	keterbukaan	menujui era keterbukaan	keterbukaan
1	@dsetia561	Ini baru yang namanya "Pejabat"ndigaji oleh r...	Ini baru yang namanya Pejabat\ndigaji oleh rak...	ini baru yang namanya pejabat\ndigaji oleh rak...	ini baru yang namanya pejabat digaji oleh raky...
2	@ahmadjamil3557	Apresiasi luar biasa kepada wapres indonesia k...	Apresiasi luar biasa kepada wapres indonesia k...	apresiasi luar biasa kepada wapres indonesia k..	apresiasi luar biasa kepada wapres indonesia k...
3	@kotakualakapuasaman6571	LUAR BIASA, sehat terus PA GIBRAN dan JAJARANN...	LUAR BIASA sehat terus PA GIBRAN dan JAJARANNY..	luar biasa sehat terus pa gibrann dan jajaranny...	luar biasa sehat terus apa gibrann dan jajarann...
4	@youngshin7879	Semakin menyala mas Gibran pilihan ku ♥♥	Semakin menyala mas Gibran pilihan ku	semakin menyala mas gibrann pilihan ku	semakin menyala mas gibrann pilihan ku

Tabel 4. Hasil dari Normalisasi Kata.

Tokenization

	author	comment	cleaning	case_folding	hasil_normalisasi	tokenize
0	@GitoGati-gr6zz	Ide yang positif. Menujui era keterbukaan	Ide yang positif Menujui era keterbukaan	ide yang positif menujui era keterbukaan	ide yang positif menujui era keterbukaan	[ide, yang, positif, menujui, era, keterbukaan]
1	@dsetia561	Ini baru yang namanya "Pejabat"ndigaji oleh r...	Ini baru yang namanya Pejabat\ndigaji oleh rak...	ini baru yang namanya pejabat\ndigaji oleh rak...	ini baru yang namanya pejabat digaji oleh raky...	[ini, baru, yang, namanya, pejabat, digaji, ol...
2	@ahmadjamil3557	Apresiasi luar biasa kepada wapres indonesia k...	Apresiasi luar biasa kepada wapres indonesia k...	apresiasi luar biasa kepada wapres indonesia k..	apresiasi luar biasa kepada wapres indonesia k...	[apresiasi, luar, biasa, kepada, wapres, indon...
3	@kotakualakapuasaman6571	LUAR BIASA, sehat terus PA GIBRAN dan JAJARANN..	LUAR BIASA sehat terus PA GIBRAN dan JAJARANNY..	luar biasa sehat terus pa gibrann dan jajaranny...	luar biasa sehat terus apa gibrann dan jajarann...	[luar, biasa, sehat, terus, apa, gibrann, dan, ...
4	@youngshin7879	Semakin menyala mas Gibran pilihan ku ♥♥	Semakin menyala mas Gibran pilihan ku	semakin menyala mas gibrann pilihan ku	semakin menyala mas gibrann pilihan ku	[semakin, menyala, mas, gibrann, pilihan, ku]

Tabel 5. Hasil dari Tokenization.

Proses Stopword Removal

	author	comment	cleaning	case_folding	hasil_normalisasi	tokenize	stopword removal
0	@GitoGati-gr6zz	Ide yang positif. Menujui era keterbukaan	Ide yang positif Menujui era keterbukaan	ide yang positif menujui era keterbukaan	ide yang positif menujui era keterbukaan	[ide, yang, positif, menujui, era, keterbukaan]	[ide, positif, menujui, era, keterbukaan]
1	@dsetia561	Ini baru yang namanya "Pejabat"n digaji oleh r...	Ini baru yang namanya Pejabat\ndi gaji oleh rak...	ini baru yang namanya pejabat\ndi gaji oleh rak...	ini baru yang namanya pejabat digaji oleh raky...	[ini, baru, yang, namanya, pejabat, digaji, ol...	[namanya, pejabat, digaji, rakyat, melayani, r...
2	@ahmadjamil3557	Apresiasi luar biasa kepada wapres indonesia k...	Apresiasi luar biasa kepada wapres indonesia k...	apresiasi luar biasa kepada wapres indonesia k..	apresiasi luar biasa kepada wapres indonesia k...	[apresiasi, luar, biasa, kepada, wapres, indon...	[apresiasi, wapres, indonesia, mas, gibran, rr...
3	@kotakualakapuasaman6571	LUAR BIASA, sehat terus PA GIBRAN dan JAJARAN N...	LUAR BIASA sehat terus PA GIBRAN dan JAJARAN NY...	luar biasa sehat terus pa gibran dan jajaranny..	luar biasa sehat terus apa gibran dan jajarann...	[luar, biasa, sehat, terus, apa, gibran, dan, ...	[sehat, gibran, jajarannya, trimakasih, bu, ms...
4	@youngshin7879	Semakin menyala mas Gibran pilihan ku ❤️❤️	Semakin menyala mas Gibran pilihan ku	semakin menyala mas gibran pilihan ku	semakin menyala mas gibran pilihan ku	[semakin, menyala, mas, gibran, pilihan, ku]	[menyala, mas, gibran, pilihan, ku]

Tabel 6. Hasil dari proses *Stopword removal*.

Proses Stemming Data

	author	comment	cleaning	case_folding	hasil_normalisasi	tokenize	stopword removal	stemming_data
0	@GitoGati-gr6zz	Ide yang positif. Menujui era keterbukaan	Ide yang positif Menujui era keterbukaan	ide yang positif menujui era keterbukaan	ide yang positif menujui era keterbukaan	[ide, yang, positif, menujui, era, keterbukaan]	[ide, positif, menujui, era, keterbukaan]	ide positif tuju era buka
1	@dsetia561	Ini baru yang namanya	Ini baru yang namanya	ini baru yang namanya	ini baru yang namanya	[ini, baru, yang,	[namanya, pejabat,	nama jabat gaji

		"Pejabat"\ ndigaji oleh r...	Pejabat\n digaji oleh rak...	pejabat\n digaji oleh rak...	pejabat digaji oleh raky...	namany a, pejabat, digaji, ol...	digaji, rakyat, melaya ni, r...	rakyat layan rakyat syaa allah...
2	@ahmadjamil3557	Apresiasi luar biasa kepada wapres indonesia k...	Apresiasi luar biasa kepada wapres indonesia k...	apresiasi luar biasa kepada wapres indonesi a k..	apresiasi luar biasa kepada wapres indonesia k...	[apresia si, luar, biasa, kepada, wapres, indon...	[apresia si, wapres, indones ia, mas, gibran, rr...	apresiasi wapres indonesi a mas gibran rr kerja
3	@kotakualakapua saman6571	LUAR BIASA, sehat terus PA GIBRAN dan JAJARAN N...	LUAR BIASA sehat terus PA GIBRAN dan JAJARA NNY...	luar biasa sehat terus pa gibran dan jajaranny ...	luar biasa sehat terus apa gibran dan jajarann...	[luar, biasa, sehat, terus, apa, gibran, dan, ...	[sehat, gibran, jajarann ya, trimaka sih, bu, ms...	sehat gibran jajar trimakas ih bu msyarak at seh...
4	@youngshin7879	Semakin menyala mas Gibran pilihan ku ❤️❤️	Semakin menyala mas Gibran pilihan ku	semakin menyala mas gibran pilihan ku	semakin menyala mas gibran pilihan ku	[semaki n, menyal a, mas, gibran, pilihan, ku]	[menyal a, mas, gibran, pilihan, ku]	nyala mas gibran pilih ku

Tabel 7. proses stemming data

Proses Hapus Data Bernilai Kosong (NaN)

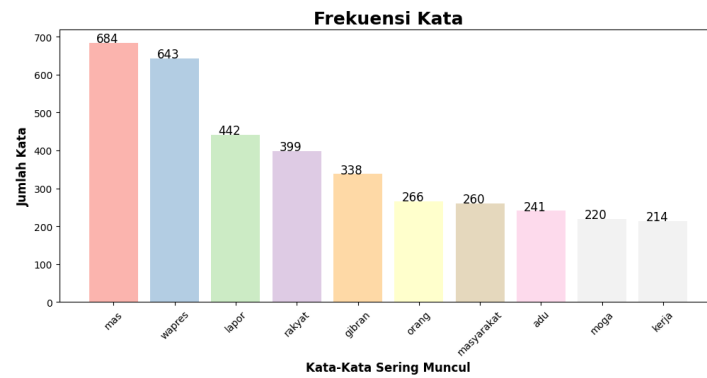
```
<class 'pandas.core.frame.DataFrame'>
Index: 2463 entries, 0 to 2499
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   author                                2463 non-null   object
1   comment                              2463 non-null   object
2   cleaning                             2463 non-null   object
3   case_folding                         2463 non-null   object
4   hasil_normalisasi                   2463 non-null   object
5   tokenize                             2463 non-null   object
6   stopword removal                     2463 non-null   object
7   stemming_data                       2463 non-null   object
dtypes: object(8)
memory usage: 173.2+ KB
```

Gambar 7. Proses hapus data bernilai kosong (NaN)

Wordcloud Setelah Preprocessing



Gambar 8. wordcloud setelah preprocessing

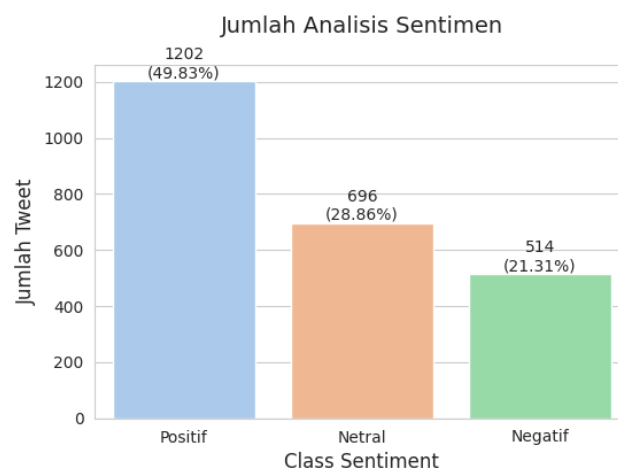


Gambar 9. Frekuensi kata-kata yang sering muncul

Pelabelan Data

	steming_data	Score	Sentiment
0	ide positif tuju era buka	1	Positif
1	nama jabat gaji rakyat layan rakyat syaa allah...	2	Positif
2	apresiasi wapres indonesia mas gibran rr kerja	1	Positif
3	sehat gibran jajar trimakasih bu msyarakat seh...	3	Positif
4	nyala mas gibran pilih ku	-1	Negatif

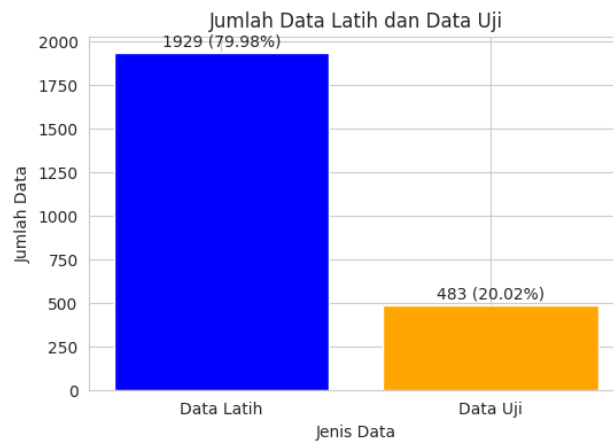
Gambar 10. Pelabelan data



Gambar 11. Laxicon Based

Terlihat dari grafik diatas menampilkan jumlah sentimen positif mendapatkan nilai (49,83%) serta sentimen netral mendaptkan nilai (28,86%) dan sentimen negatif mendapatkan nilai (21,31%)

Spliting Dataset

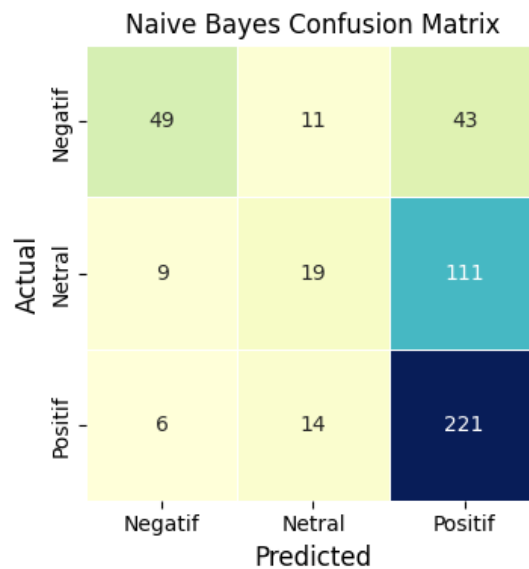


Gambar 12. Splitting dataset

Jumlah Data Latih: 1929

Jumlah Data Uji: 483

Penerapan Algoritma Naive Bayes Classifier (NBC)



Gambar 13. Confusion Matrix Naive Bayes:

precision	recall	f1-score	support	
Negatif	0.766	0.476	0.587	103.000
Netral	0.432	0.137	0.208	139.000
Positif	0.589	0.917	0.718	241.000
accuracy	0.598	0.598	0.598	0.598
macro avg	0.596	0.510	0.504	483.000
weighted avg	0.582	0.598	0.543	483.000

Tabel 8. Classification Report for Naive Bayes:

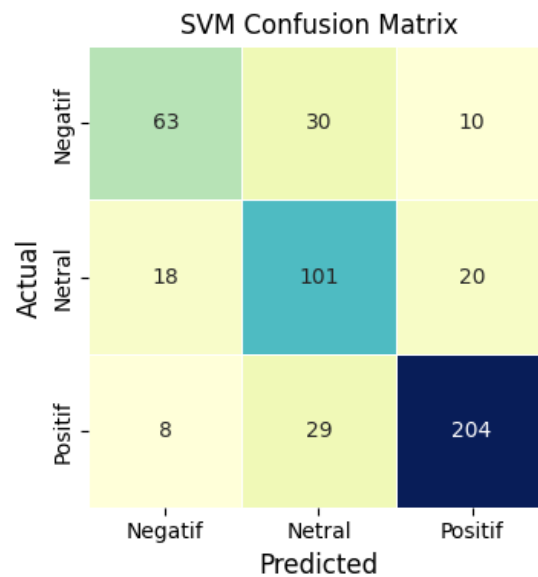
$$\text{Akurasi} = \frac{\text{Jumlah klasifikasi benar}}{\text{jumlah total data}} = \frac{49+19+221}{483} = \frac{289}{483} = 0.598$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{221}{221+43+111} = \frac{221}{375} = 0.589$$

$$\text{Sensitivitas} = \frac{TP}{TP+FN} = \frac{221}{221+20} = \frac{221}{241} = 0.917$$

$$\text{Spesifitas} = \frac{TN}{TN+FP} = \frac{88}{88+154} = \frac{88}{243} = 0.364$$

Penerapan Algoritma Support Vector Machine (SVM)



Gambar 14. Confusion Matrix SVM

precision	recall	f1-score	support	
Negatif	0.708	0.612	0.656	103.000
Netral	0.631	0.727	0.676	139.000
Positif	0.872	0.846	0.859	241.000
accuracy	0.762	0.762	0.762	0.762
macro avg	0.737	0.728	0.730	483.000
weighted avg	0.768	0.762	0.763	483.000

Tabel 9. Classification Report SVM:

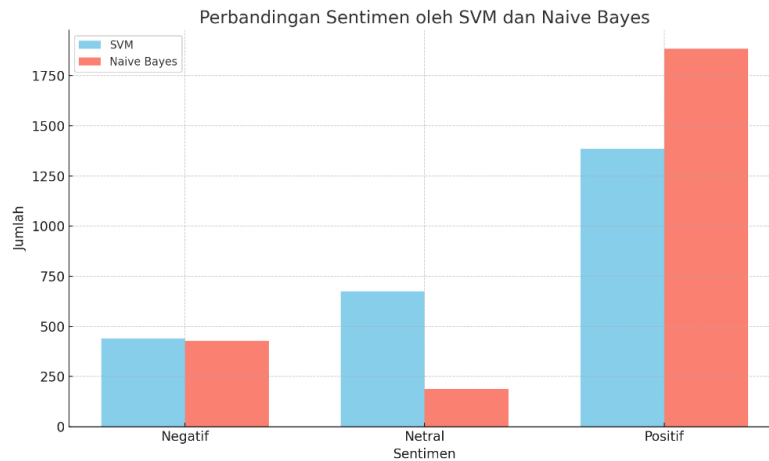
$$\text{Akurasi} = \frac{\text{Jumlah klasifikasi benar}}{\text{jumlah total data}} = \frac{\text{Total TP semua kelas}}{\text{Total data}} = \frac{63+101+204}{483} = \frac{368}{483} = 0.762$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{204}{204+30} = \frac{204}{234} = 0.8718$$

$$\text{Sensitivitas} = \frac{TP}{TP+FN} = \frac{204}{204+37} = \frac{204}{241} = 0.8465$$

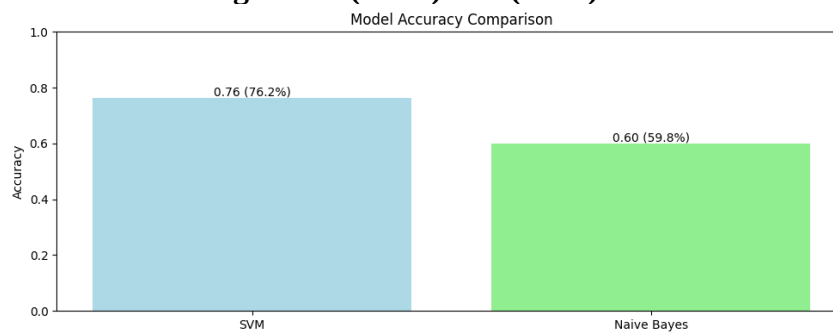
$$\text{Spesifitas} = \frac{TN}{TN+FP} = \frac{212}{212+30} = \frac{212}{242} = 0.8750$$

Perbandingan Sentimen Penggunaan Algoritma (SVM) dan (NBC)



Gambar 15. Grafik perbandingan sentimen

Model Komparasi Akurasi Algoritma (SVM) dan (NBC)



Gambar 16. Grafik komparasi

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan terhadap analisis sentimen komentar YouTube pada program *Lapor Mas Wapres* Menyatakan bahwa sentimen positif mendominasi urutan ke 1 dan sentimen netral di urutan ke 2 serta sentimen negatif di urutan ke 3, dengan membandingkan dua algoritma klasifikasi yakni Support Vector Machine (SVM) dan Naïve Bayes, maka dapat disimpulkan sebagai berikut:

1. Performa Model

- Model SVM menghasilkan akurasi sebesar 76,2%, sementara model Naïve Bayes menghasilkan akurasi 59,8%. Dengan demikian, SVM terbukti lebih akurat dalam mengklasifikasikan sentimen komentar.
- SVM menghasilkan klasifikasi sentimen yang lebih merata,
- Sebaliknya, Naïve Bayes menghasilkan distribusi yang berat . Hal ini menunjukkan bahwa Naïve Bayes cenderung overfitting terhadap fitur-fitur sentimen positif dalam data.

2. Evaluasi Berdasarkan Confusion Matrix

- Dari confusion matrix dapat dilihat bahwa SVM memiliki jumlah klasifikasi benar (True Positive dan True Negative) yang lebih tinggi dibandingkan Naïve Bayes. Ini memperkuat hasil bahwa SVM lebih unggul secara keseluruhan dalam menangani tiga kategori sentimen.

3. Kesesuaian dengan Data

- Mengingat karakteristik data komentar YouTube yang memiliki banyak variasi kata, bentuk informal, dan konteks tidak selalu eksplisit, maka pendekatan SVM yang mampu memisahkan data secara maksimal dalam ruang vektor lebih efektif dibandingkan pendekatan probabilistik yang digunakan oleh Naïve Bayes.

REFERENSI

- [1] ObserverID, "Lapor Mas Wapres Command Post Receives 296 Complaints," <https://observerid.com/lapor-mas-wapres-command-post-receives-296-complaints>.
- [2] mohammad hatta muarabagja, "https://www.tempo.co/politik/menilik-deretan-respons-kantor-komunikasi-kepresidenan-soal-lapor-mas-wapres-1169152."
- [3] H. Hidayat, F. Santoso, and L. F. Lidimillah, "Analisis Sentimen Pengguna YouTube Tentang Rohingya Menggunakan Algoritma SVM (Support Vector Machine)," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 1729–1738, Jul. 2024, doi: 10.33379/gtech.v8i3.4497.
- [4] A. Muhammadin and I. A. Sobari, "Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma SVM Dan NBC," *Jurnal Rekayasa Perangkat Lunak*, vol. 2, no. 2, 2021, [Online]. Available: <http://jurnal.bsi.ac.id/index.php/reputasi>
- [5] T. A. Azzahra *et al.*, "Perbandingan Efektivitas Naïve Bayes dan SVM dalam Menganalisis Sentimen Kebencanaan di Youtube," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 1, p. 312, Jan. 2024, doi: 10.30865/mib.v8i1.7186.
- [6] A. Karimah and G. Dwilestari, "ANALISIS SENTIMEN KOMENTAR VIDEO MOBIL LISTRIK DI PLATFORM YOUTUBE DENGAN METODE NAIVE BAYES," 2024. [Online]. Available: <https://www.kaggle.com/datasets/billycemerson/anali>
- [7] R. A. Fauzan, "(SENAFTI) 30 Agustus 2023-Jakarta," 2023.
- [8] N. Bayes, O. : Sulistia, M. Harahap, and R. Kurniawan, "Analisis Sentimen Komentar Youtube terhadap Food Vlogger dengan Menggunakan Metode Analisis Sentimen Komentar Youtube terhadap Food Vlogger dengan Menggunakan Metode Naïve Bayes," vol. 9, no. 1, [Online]. Available: http://ejournal.ust.ac.id/index.php/Jurnal_Means/
- [9] M. Rizki *et al.*, "PERBAIKAN ALGORITMA NAIVE BAYES CLASSIFIER MENGGUNAKAN TEKNIK LAPLACIAN CORRECTION," 2021.
- [10] A. H. Yunial, "Prosiding Seminar Nasional Informatika dan Sistem Informasi ANALISA PERBANDINGAN ALGORITMA KLASIFIKASI SUPPORT VECTOR MACHINE, DECESSION TREE DAN NAIVE BAYES".
- [11] M. A. Subarkah *et al.*, "Analisis Sentimen Terhadap Video Ulasan Produk Menggunakan Metode Support Vector Machine Dengan Sequential Minimal Optimization Article History ABSTRAK," *Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika (KERNEL)*, vol. 3, no. 2, 2022.
- [12] R. Oktaria Mardiyanto and dan Ferry Wahyu Wibowo, "Mardiyanto, Kusrini, dan Wibowo-Analisis Sentimen Pengguna Aplikasi Bank Syariah Indonesia Dengan Menggunakan Algoritma Support Vector Machine (SVM) ANALISIS SENTIMEN PENGGUNA APLIKASI BANK SYARIAH INDONESIA DENGAN0MENGGUNAKAN0ALGORITMA SUPPORT VECTOR MACHINE (SVM) (ANALYSIS OF USER SENTIMENT OF BANK INDONESIA APPLICATIONS USING SUPPORT VECTOR MACHINE (SVM) ALGORITHM)."