



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak cipta milik UIN Suska Riau



State Islamic University of Sultan Syarif Kasim Riau

ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP ANIES BASWEDAN SEBAGAI BAKAL CALON PRESIDEN 2024 MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

CHELY AULIA MISRUN
NIM. 11950121677



UIN SUSKA RIAU

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2023

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP ANIES BASWEDAN SEBAGAI BAKAL CALON PRESIDEN 2024 MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

TUGAS AKHIR

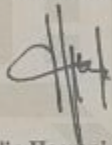
Oleh

CHELY AULIA MISRUN

NIM. 11950121677

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 22 Juni 2023

Pembimbing I,



DR. Elin Haerani, S.T., M.Kom

NIP. 19810523 200710 2 003

UIN SUSKA RIAU

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP ANIES BASWEDAN SEBAGAI BAKAL CALON PRESIDEN 2024 MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Oleh

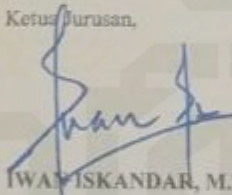
CHELY AULIA MISRUN

NIM. 11950121677

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 22 Juni 2023

Mengesahkan,
Ketua Jurusan,



IWAN ISKANDAR, M.T.

NIP. 19821216 201503 1 003

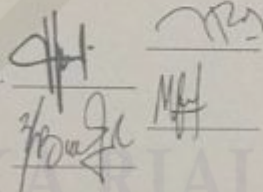


DR. HARTONO, M.PD.

NIP. 19640301 199203 1 003

DEWAN PENGUJI

Ketua : Novriyanto, S.T., M.Sc.
Pembimbing I : DR. Elin Haerani, S.T., M.Kom.
Penguji I : Muhammad Fikry, S.T., M.Sc.
Penguji II : Elvia Budianita, S.T., M.Cs.



SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Chely Aulia Misrun

NIM : 11950121677

Tempat/Tgl.Lahir : Pekanbaru, 12 Juli 2001

Fakultas : Sains dan Teknologi

Prodi : Teknik Informatika

Judul Skripsi : Analisis Sentimen Komentar Youtube Terhadap Anies Baswedan Sebagai Bakal Calon Presiden 2024 Menggunakan Metode *Naive Bayes Classifier*

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya
3. Oleh karena itu Skripsi saya ini, saya nyatakan bebas dari plagiat
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah surat pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga

Pekanbaru, 4 Juli 2023

Yang membuat pernyataan



CHELY AULIA MISRUN
NIM. 11950121677

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier

Chely Aulia Misrun¹, Elin Haerani², Muhammad Fikry³, Elvia Budianita⁴

Email: ¹11950121677@students.uin-suska.ac.id, ²elin.haerani@uin-suska.ac.id, ³muhammad.fikry@uin-suska.ac.id, ⁴elvia.budianita@uin-suska.ac.id,

^{1,2,3,4} Teknik Informatika, UIN Sultan Syarif Kasim Riau

Diterima: 31 Maret 2023 | Direvisi: - | Disetujui: 30 April 2023

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Suatu tokoh sebagai bakal calon presiden adalah Anies Baswedan mantan gubernur DKI Jakarta yang menerima banyak penghargaan dan memiliki kebijakan program kerja yang efektif dalam permasalahan di wilayah DKI Jakarta. Komentar mengenai Anies Baswedan sebagai bakal calon presiden 2024 banyak dijumpai pada media sosial youtube. Youtube memfasilitasi pengguna untuk memberikan komentar dalam menanggapi video yang dapat dijadikan sebuah informasi analisis sentimen untuk mengetahui komentar positif serta komentar negatif. Algoritma yang dipakai pada penelitian ini ialah *naïve bayes classifier*. Terdapat lima proses utama pada penelitian ini, yaitu pengumpulan data, pembobotan kata (TF-IDF), *text preprocessing*, klasifikasi (*naïve bayes classifier*) dan pengujian. Dari 1009 data komentar di youtube berbahasa Indonesia terkait video Anies Baswedan sebagai bakal calon presiden 2024. Berdasarkan hasil analisis, terdapat 610 komentar positif serta 399 negatif. Hasil akurasi menggunakan algoritma *naïve bayes classifier* sebesar 78% yang di dapat dengan menggunakan perbandingan 10% data uji serta 90% data latih.

Kata kunci: *anies baswedan, naïve bayes classifier, analisis sentimen, youtube.*

Sentiment analysis of youtube comments on Anies Baswedan as a 2024 presidential candidate using the naïve bayes classifier method

Abstract

One of the figures as a presidential candidate is Anies Baswedan, the former governor of DKI Jakarta who received many awards and has an effective work program policy for problems in the DKI Jakarta area. Many comments about Anies Baswedan as a 2024 presidential candidate are found on YouTube social media. Youtube facilitates users to provide comments in response to videos which can be used as sentiment analysis information to find out positive comments and negative comments. The algorithm used in this research is the *naïve bayes classifier*. There are five main processes in this research, namely data collection, *text preprocessing*, word weighting (TF-IDF), classification (*Naïve Bayes Classifier*) and testing. From 1009 comment data on Indonesian-language youtube related to the Anies Baswedan video as a 2024 presidential candidate. Based on the analysis results, there are 610 positive comments and 399 negative comments. The accuracy result using the *naïve bayes classifier* algorithm is 78% which is obtained by using a comparison of 90% training data and 10% test data.

Keywords: *anies baswedan, naïve bayes classifier, analisis sentimen, youtube.*

1. PENDAHULUAN

Indonesia merupakan sebuah negara yang memiliki total masyarakat terbanyak keempat di dunia dengan total warga 274.790.244 dan menganut sistem demokrasi. Di Indonesia, pemilu dilaksanakan perlima tahun [1]. Sejak tahun 2024 ini dilaksanakan hal

tersebut. Presiden adalah pemimpin suatu negara yang memiliki peran menjadi kepala pemerintahan serta negara. Komentar dari rakyat menentukan tingkat popularitas seorang tokoh politik yang hendak mencalonkan menjadi presiden. Salah satu tokoh politik bakal calon presiden 2024 adalah Anies Baswedan mantan gubernur DKI Jakarta periode menjabat 16 Oktober 2017-16 Oktober 2022 yang telah menerima banyak penghargaan dan memiliki kebijakan program kerja yang cukup efektif dalam menangani permasalahan di wilayah DKI Jakarta.

Youtube merupakan perolehan laporan berbentuk video dengan pemakai aktif terbanyak, pemakai bisa berkomunikasi secara berbagai video, membagikan *dislike* atau *like*, menambahkan *viewer* pada sebuah video yang berlangganan (*subscribe*) terhadap sebuah *channel* [2]. Fasilitas yang diberikan oleh youtube kepada pengguna untuk menanggapi video dengan cara memberikan pendapat di kolom komentar. 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 sentimen.

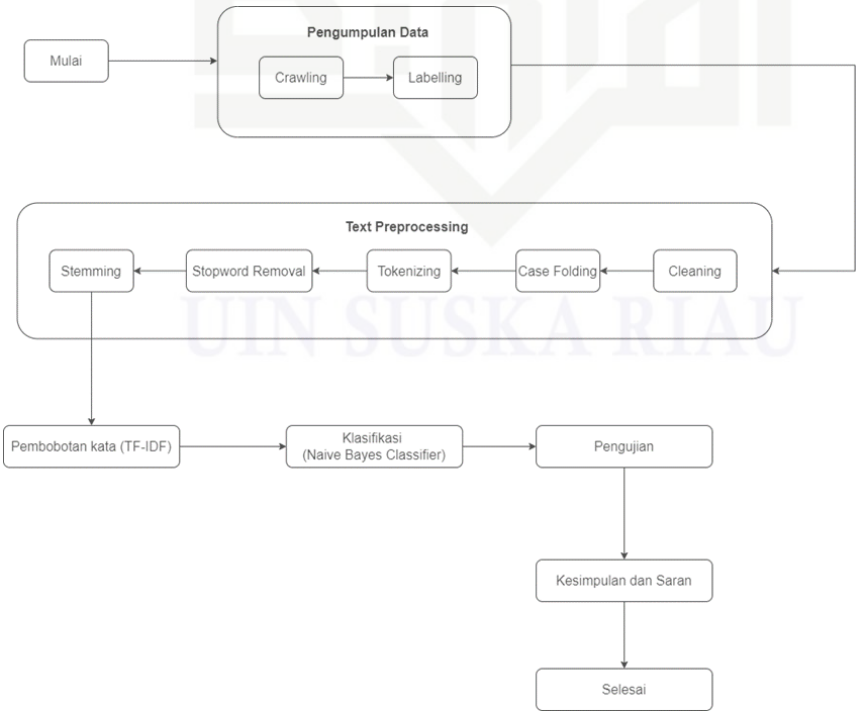
Salah satu bagian dari bidang ilmu text mining adalah analisis sentimen yang bertarget guna mengenali opini sebuah seseorang serta organisasi pada sebuah objek apakah bersifat positif dan negatif [3]. Beberapa algoritma yang dapat dimanfaatkan dalam mengerjakan analisis sentimen yaitu *support vector machine*, *naïve bayes classifie*, KNN atau *K-Nearest Neighbors* serta lainnya.

Terdapat beberapa penelitian yang berhubungan pada analisis sentimen memakai algoritma *naive bayes classifier*, seperti penelitian yang dilaksanakan Dhaifa Farah Zhafira, dkk dengan judul Analisis sentimen kebijakan kampus merdeka menggunakan Naive Bayes Classifier serta pembobotan TF-IDF melalui komentar terhadap youtube menghasilkan akurasi terbaik sebesar 97%, nilai *precision* 90.35%, *recall* 93.6%, dan *f-measure* 91.95% [4]. Selanjutnya pada penelitian berjudul Analisis sentimen pengguna youtube terhadap tayangan #MATANAJWAMENANTITERAWAN dengan metode Naïve Bayes Classifier mendapatkan akurasi 90.36% dengan data netral berjumlah 1232, data negatif 90 dan 78 data positif [5].

Dalam penelitian ini memakai algoritma *Naïve Bayes Classifier*. Hal tersebut memiliki taraf dugaan yang maksimal dan tepat daripada model *classifier* yang lain [6]. Penelitian ini bertujuan untuk mengklasifikasikan kalimat atau sentimen terhadap Anies Baswedan sebagai bakal calon presiden 2024 kedalam dua golongan yaitu, positif serta negatif secara menerapkan metode Naïve Bayes Classifier dan mengamati taraf akurasi dari metode Naïve Bayes Classifier guna mengimplementasikan sentimen pada topik Anies Baswedan sebagai bakal calon presiden 2024 terhadap media sosial youtube.

2. METODE PENELITIAN

Metode penelitian merupakan langkah untuk menggabungkan informasi atau data yang di dapatkan sebagai tujuan dari penelitian. Berikut tahapan yang hendak diakhiri, diamati terhadap gambar 1.



Gambar 1. Tahapan Flowchart Metodologi Penelitian

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

State Islamic University of Sultan Syarif Kasim Riau

2.1. Pengumpulan Data

Pada tahap pertama dilakukan pengumpulan data yang di dapat dari sosial media youtube dengan keyword “Anies Baswedan” sebagai bakal calon presiden Republik Indonesia 2024. Selanjutnya proses labelling bertujuan untuk memberikan label pada dataset yang telah didapatkan [7], pelabelan dilakukan secara Crowdsourcing. Crowdsourcing pada penelitian ini terdiri dari tiga dosen bahasa. Data yang berhasil diambil sebanyak 1009 komentar yang mencakup dua kelas atau label ialah kelas positif serta kelas negatif. Berdasarkan hasil crawling serta pelabelan yang dilakukan oleh tiga dosen bahasa diperoleh sentimen positif sejumlah 610 serta sentimen negatif sejumlah 399 data.

2.2. Text Preprocessing

Setelah memberikan label pada setiap data yang diperoleh, selanjutnya dilakukan text preprocessing. Tahapan ini dilakukan seleksi data supaya data yang tak tersusun sebagai data yang tersusun [8].

Dalam text preprocessing ada sebagian langkah yang dilaksanakan, mencakup :

- 1) Cleaning : bertujuan guna menghapus karakter atau simbol yang tidak dibutuhkan pada data yang diperoleh terhadap Anies Baswedan sebagai bakal calon presiden 2024 yang berjumlah 1009 data.
- 2) Case Folding : bertujuan guna menyamakan seluruh huruf terhadap komentar sebagai huruf kecil.
- 3) Tokenizing : tahap ini bertujuan memisah suatu kalimat menjadi potongan kata per kata setelah melakukan case folding.
- 4) Stopword Removal : tahap ini bermaksud guna menghapus kata yang tak mempunyai arti pada data yang diperoleh terhadap Anies Baswedan sebagai bakal calon presiden 2024.
- 5) Stemming : bertujuan untuk menghilangkan imbuhan awalan, akhiran dan sisipan untuk kembali dalam bentuk kata dasar sesuai KBBI.

2.3. Pembobotan Kata (TF-IDF)

Pembobotan kata TF-IDF (*Term frequency-Inverse Document Frequency*) merupakan sebuah tahap dalam melakukan perubahan data tekstual pada data numerik perkata [9]. TF-IDF ialah suatu algoritma untuk menghitung bobot setiap kata pada data yang telah di dapatkan. TF ialah frekuensi kehadiran kata terhadap setiap dokumen. IDF ialah nilai DF. IDF melibatkan fungsi algoritma karena jika tidak penilaian yang efektif dalam dua dokumen terlalu ekstrim [10]. DF adalah frekuensi dokumen yang mengandung kata. Berikut ini rumus menghitung TF-IDF :

$$W_{t,d} = TF_{t,d} \cdot IDF_t \quad (1)$$

$$IDF_t = \log \left(\frac{N}{n_t} \right) \quad (2)$$

$$TF - IDF_{t,d} = TF_{t,d} \cdot IDF_t \quad (3)$$

Keterangan :

- $W_{t,d}$ = bobot melalui t (*term*) pada suatu dokumen
- $TF_{t,d}$ = frekuensi kehadiran t (*term*) pada dokumen d
- IDF_t = inverse document frequency
- $TF - IDF_{t,d}$ = kalimat bobot (d) terhadap kata (t)
- N = total dokumen
- n_t = total dokumen yang terkandung kata t

2.4. Klasifikasi (Naïve Bayes Classifier)

Metode pembelajaran probabilitas pada Machine Learning untuk mengklasifikasikan data uji secara independen [11]. Teorema bayes merupakan teorema untuk memprediksi peristiwa masa depan berdasarkan pengalaman dimasa sebelumnya [12]. Kelebihan algoritma naïve bayes classifier yaitu mudah di implementasikan dan mempunyai tingkat akurasi yang tinggi [13]. Pada tahap ini menentukan apakah suatu data termasuk kelas positif dan kelas negatif. Berikut rumus pengkalkulasian Naïve Bayes :

$$P(C|X) = \frac{P(X|C) P(C)}{P(X)} \quad (4)$$

Rumus untuk menghitung Naïve Bayes ditunjukkan pada persamaan (4). Dimana X menguji dara dari kelas yang tidak diketahui. Nilai C berarti kumpulan data X adalah kelas yang terdefinisi dengan baik. P(C) ialah probabilitas C, P(X) adalah probabilitas X, P(C|X) ialah probabilitas hipotesis C akibat keadaan X dan terakhir P(X|C) berarti probabilitas X sehubungan dengan situasi hipotesis C [14].

2.5. Pengujian

Pada tahap ini pengujian menggunakan metode confusion matrix. Metode yang dipakai guna mengukur kinerja atau tingkat kebenaran sebuah klasifikasi yaitu confusion matrix [15]. Pada tahap pengujian confusion matrix, terdiri dari empat istilah yaitu, TN atau True Negative yaitu total data negatif yang terklasifikasikan tepat, TP atau True Positive ialah data positif yang terklasifikasikan tepat, serta FN atau False Negative ialah sebuah data positif tetapi terduga menjadi data negative juga FP atau False Positive yaitu data negatif tetapi terdeteksi menjadi data positif.

Tabel 1. Confusion Matrix

	Prediksi Positif	Prediksi Negatif
Realita Positif	TP	FN
Realita Negatif	FP	TN

Melalui metode confusion matrix bisa didapat nilai presisi, akurasi serta recall. Nilai akurasi dapat melihat tingkat akurat sistem dalam mengategorikan data dengan tepat dengan keseluruhan data. Nilai akurasi di dapatkan secara melakukan persamaan (5). Taraf akurasi laporan yang dimintakan pemakai serta anggapan yang dibagikan sistem disebut nilai presisi (6). Recall merupakan taraf kesuksesan sistem guna memilih laporan (7) [16]. F-Measure, menghitung rata-rata harmonik dari recall dan precision yang memberikan masing-masing bobot yang sama (8) [17].

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \tag{5}$$

$$\text{Presisi} = \frac{TP}{TP+FP} * 100\% \tag{6}$$

$$\text{Recall} = \frac{TP}{TP+FN} * 100\% \tag{7}$$

$$F - \text{Measure} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

3. HASIL DAN PEMBAHASAN

3.1. Dataset

Dataset yang dipakai dalam penelitian ini merupakan data yang tersedia secara publik pada media sosial youtube. Teknik pengumpulan data yang dipakai merupakan teknik *scraping*, menggunakan bahasa pemograman python. Proses *scraping* data yang bersumber dari Youtube dapat memanfaatkan fasilitas Youtube Data API yang sudah tersedia pada menu *resources* [4]. Pengambilan dan pengumpulan data dilaksanakan sejak 12 Desember 2022, terhimpun sejumlah 1009 data. Berdasarkan hasil crawling dan pelabelan yang dilakukan oleh tiga dosen bahasa diperoleh sentimen positif sejumlah 610 serta sentimen negatif sejumlah 399 data. Total dataset serta perolehan melalui klasifikasi dalam tabel 2.

Tabel 2. Jumlah Dataset dan Hasil

Sentimen	Jumlah	Persen
Positif	610	60,4559%
Negatif	399	39,5441%
Total	1009	100,00%

3.2. Preprocessing Data

Sesudah sukses mengumpulkan dataset dan memberikan label pada semua data yang di dapat, langkah selanjutnya melakukan preprocessing data. Dalam langkah preprocessing data dilaksanakan secara lima langkah ialah stopword removal, cleaning, case folding, tokenizing serta stemming. Proses pembobotan dan pengklasifikasian merupakan tahapan awal yang sangat penting [18]. Hasil dari preprocessing data dalam tabel 3.

Tabel 3. Contoh Hasil Preprocessing Data

Preprocessing	Input	Output
Cleaning	Anies pemimpin yang cerdas dan berintegritas, sangat cocok menjadi calon presiden 2024	Anies pemimpin yang cerdas dan berintegritas sangat cocok menjadi calon presiden 2024
Case Folding	Anies pemimpin yang cerdas dan berintegritas sangat cocok menjadi calon presiden 2024	anies pemimpin yang cerdas dan berintegritas sangat cocok menjadi calon presiden 2024
Tokenizing	anies pemimpin yang cerdas dan berintegritas sangat cocok menjadi calon presiden 2024	[anies, pemimpin, yang, cerdas, dan, berintegritas, sangat, cocok, menjadi, calon, presiden, 2024]
Stopword Removal	[anies, pemimpin, yang, cerdas, dan, berintegritas, sangat, cocok, menjadi, calon, presiden, 2024]	[anies, pemimpin, cerdas, berintegritas,, cocok, calon, presiden, 2024]
Stemming	[anies, pemimpin, cerdas, berintegritas,, cocok, calon, presiden, 2024]	[anies, pimpin, cerdas, integritas, cocok, calon, presiden, 2024]

Selanjutnya, dilakukan pengujian teknik *preprocessing* yang berbeda-beda. Pada pengujian ini dilakukan 2 (dua) kali, yaitu melakukan teknik *preprocessing* tanpa *stopword removal* dan melakukan seluruh tahapan *preprocessing*. Tabel 4 pengujian teknik *preprocessing*.

Tabel 4. Pengujian Teknik Preprocessing

Preprocessing	Akurasi
Full Preprocessing (Data Latih 70% dan 30% Data Uji)	77%
Full Preprocessing (Data Latih 80% dan 20% Data Uji)	76%
Full Preprocessing (Data Latih 90% dan 10% Data Uji)	78%
Tanpa Stopword Removal (Data Latih 70% dan 30% Data Uji)	77%
Tanpa Stopword Removal (Data Latih 80% dan 20% Data Uji)	77%
Tanpa Stopword Removal (Data Latih 90% dan 10% Data Uji)	79%

Melalui tabel 4, kinerja sistem terbaik diperoleh ketika menggunakan teknik *preprocessing* tanpa *stopword Removal* dengan perbandingan 10% data uji serta data latih 90% menghasilkan akurasi 79%, dibandingkan dengan hasil akurasi full *preprocessing* pada 10 % data uji serta data latih 90% yang memperoleh akurasi yaitu 78%. Hal tersebut bisa disebabkan karena tidak melakukan *stopward removal* jadi informasi dari kalimat tidak hilang dan mendapatkan akurasi lebih tinggi dibandingkan melakukan *stopward removal* [19].

3.3. Pembobotan Kata (TF-IDF)

Tahap ini memakai *TfidfVectorizer* dan *CountVectorizer*. Tabel 5 adalah hasil pembobotan kata TF-IDF.

Tabel 5. Contoh Penggalan Data Hasil Pembobotan Kata TF-IDF

Kata	IDF
pimpin	0,097
anies	0,222
siapa	0,398
bangsa	0,699

3.4 Klasifikasi Naïve Bayes Classifier

Data yang sudah melewati proses *preprocessing*, labelling dan term weighting (TF-IDF) menjadi input pada proses testing dan training. Tahapan pertama yang dilakukan adalah memisahkan data training dan data testing. Lalu dilaksanakan tahap klasifikasi secara memakai algoritma Naïve Bayes Classifier guna membuat model yang bisa digunakan untuk pengolahan klasifikasi data tanpa pemberian label yang dilakukan secara manual. Implementasi melalui algoritma naïve bayes classifier dilaksanakan secara

memakai naive_bayes, python sklearn secara memanggikan Pustaka MultinomialNB [20]. Pengujian ini dilakukan menggunakan metode confusion matrix yang akan dimasukkan ke dalam perhitungan accuracy, precision, dan recall.

Pada penelitian ini dilakukan tiga model eksperimen melalui total data latih. Total data latih yang pertama yaitu sejumlah 70% serta 30% data uji, data latih kedua ialah sejumlah 80% serta 20% data uji, dan yang terakhir data latih sejumlah 90% serta 10% data uji. Data latih pertama sejumlah 70%.

	precision	recall	f1-score	support
Negatif	0.97	0.47	0.63	128
Positif	0.72	0.99	0.83	175
accuracy			0.77	303
macro avg	0.84	0.73	0.73	303
weighted avg	0.82	0.77	0.75	303

Gambar 2. Hasil Klasifikasi Data Latih 70% & 30% Data Uji

Selanjutnya hasil klasifikasi data latih kedua yaitu sejumlah 80% serta 20% data uji.

	precision	recall	f1-score	support
Negatif	0.95	0.47	0.62	86
Positif	0.71	0.98	0.83	116
accuracy			0.76	202
macro avg	0.83	0.72	0.73	202
weighted avg	0.81	0.76	0.74	202

Gambar 3. Hasil Klasifikasi Data Latih 80% & 20% Data Uji

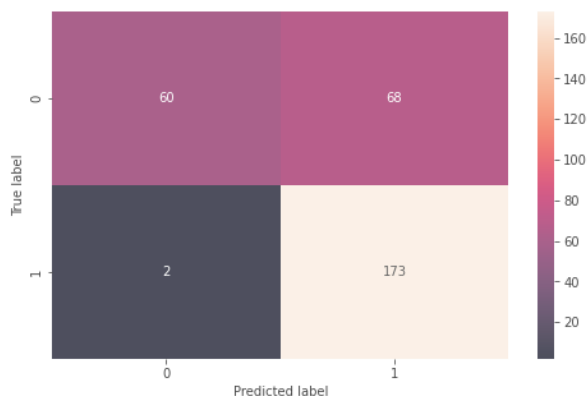
Selanjutnya klasifikasi yang terakhir data latih yaitu sejumlah 90% serta 10% data uji.

	precision	recall	f1-score	support
Negatif	0.87	0.51	0.65	39
Positif	0.76	0.95	0.84	62
accuracy			0.78	101
macro avg	0.81	0.73	0.74	101
weighted avg	0.80	0.78	0.77	101

Gambar 4. Hasil Klasifikasi Data Latih 90% & 10% Data Uji

3.5 Pengujian

Pada penelitian ini dilaksanakan pengujian melalui data uji serta data latih yang berjumlah 1009 data. Pembagiannya memakai 3 perbandingan ialah data latih 80% : 20% data uji, selanjutnya perbandingan data latih 70% : 30% data uji, serta perbandingan data latih 90% : 10% data uji.



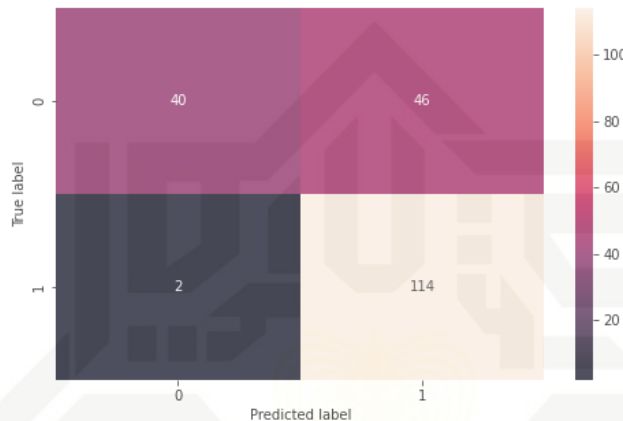
Gambar 5.. Hasil Confusion Matrix Perbandingan 70% : 30%

Dengan hasil dari confusion matrix perbandingan data latih 70% : 30% data uji, maka didapatkan nilai akurasi sejumlah :

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} * 100\% = \frac{60 + 173}{60 + 173 + 68 + 2} * 100\%$$

$$= \frac{233}{303} * 100\% = 0,77 * 100\% = 77\%$$

Hasil akurasi algoritma naïve bayes classifier pada perbandingan data latih 70% serta 30% data uji sejumlah 0,77 atau 77% dalam klasifikasi komentar youtube.



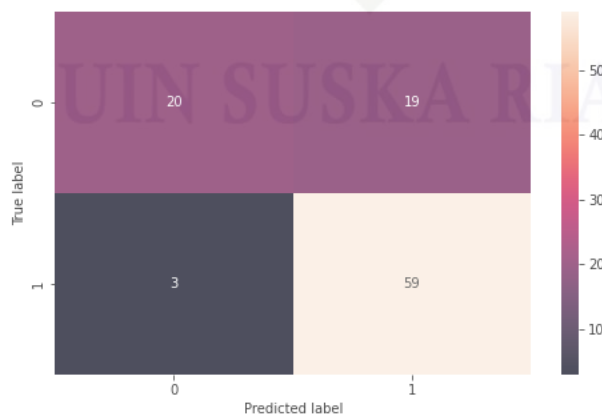
Gambar 6.. Hasil Confusion Matrix Perbandingan 80% : 20%

Dengan hasil dari confusion matrix perbandingan data latih 80% : 20% data uji, maka didapatkan nilai akurasi sejumlah :

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} * 100\% = \frac{40 + 114}{40 + 114 + 46 + 2} * 100\%$$

$$= \frac{154}{202} * 100\% = 0,76 * 100\% = 76\%$$

Hasil akurasi algoritma naïve bayes classifier pada perbandingan data latih 80% serta 20% data uji sejumlah 0,76 atau 76% dalam klasifikasi komentar youtube.



Gambar 7.. Hasil Confusion Matrix Perbandingan 90% : 10%

Dengan hasil dari confusion matrix perbandingan data latih 90% : 10% data uji, maka didapatkan nilai akurasi sejumlah :

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} * 100\% = \frac{20 + 59}{20 + 59 + 19 + 3} * 100\%$$



$$= \frac{79}{101} * 100\% = 0,78 * 100\% = 78\%$$

Hasil akurasi algoritma naïve bayes classifier pada perbandingan data latih 90% serta 10% data uji sejumlah 0,78 atau 78% dalam klasifikasi komentar youtube.

4. KESIMPULAN

Berdasarkan hasil penelitian analisis sentimen komentar di youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode *naïve bayes classifier* dengan total 1009 data, menunjukkan lebih banyak komentar positif yang berjumlah 610 dari pada komentar negatif yang berjumlah 399. Hasil pengujian pada tahapan *preprocessing* dengan dua kali pengujian di dapatkan tingkat *accuracy* terbaik ketika melakukan teknik *preprocessing* tanpa *stopword removal* pada perbandingan data latih 90% serta 10% data uji menghasilkan *accuracy* ialah 79%. Hal ini disebabkan karena tidak melakukan *stopward removal* jadi informasi dari kalimat tidak hilang dan mendapatkan akurasi lebih tinggi dibandingkan melakukan *stopward removal*. Selain itu, dari tahap uji memakai algoritma *naïve bayes classifier* pada tiga model perbandingan data uji serta data latih. Jumlah data latih yang pertama yaitu sejumlah 70% serta 30% data uji, menghasilkan nilai *accuracy* sejumlah 77%, *recall* 73% serta *precision* 84%. Jumlah data latih kedua sejumlah 80% serta 20% data uji, menghasilkan nilai *accuracy* sejumlah 76%, *precision* 83% serta *recall* 72%. Pembagian total data latih serta data uji pada *accuracy* tertinggi di dapat pada perbandingan jumlah data latih 90% serta 10% data uji, sehingga memperoleh nilai *accuracy* sejumlah 78%, *recall* 73% serta *precision* 81% yang mengindikasikan jika algoritma *naïve bayes classifier* cukup tepat guna menyelesaikan klasifikasi sentimen terhadap komentar youtube terhadap topik Anies Baswedan sebagai bakal calon presiden 2024. Saran untuk penelitian selanjutnya, menggunakan algoritma lain untuk menentukan seberapa bagus kinerja algoritma tersebut dibandingkan algoritma naïve bayes classifier guna menyelesaikan klasifikasi sentimen komentar youtube pada topik Anies Baswedan sebagai bakal calon presiden 2024.

DAFTAR PUSTAKA

- [1] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook," *Eksplora Informatika*, vol. 9, no. 1, pp. 60–69, Sep. 2019, doi: 10.30864/eksplora.v9i1.254.
- [2] K. A. B. Permana, M. Sudarma, and W. G. Ariastina, "Analisis Rating Sentimen pada Video di Media Sosial Youtube Menggunakan STRUCT-SVM," *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 1, p. 113, May 2019, doi: 10.24843/mite.2019.v18i01.p17.
- [3] D. Angraina and A. Putri, "Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma Support Vector Machine," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 472–478, Dec. 2022, doi: 10.37859/coscitech.v3i3.4260.
- [4] D. F. Zhafira, B. Rahayudi, and Indriati, "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF," 2021.
- [5] M. Hudha, E. Supriyati, and T. Listyorini, "Analisis Sentimen Pengguna Youtube Terhadap Tayangan #MATANAJWAMENANTITERAWAN Dengan Metode Naive Bayes Classifier," *Jurnal Informatika dan Komputer Akreditasi KEMENRISTEKDIKTI*, vol. 5, no. 1, pp. 2614–8897, 2022, doi: 10.33387/jiko.
- [6] T. Sanubari, C. Prianto, and N. Riza, *Odol (One desa one product unggulan online) penerapan metode Naive Bayes pada pengembangan aplikasi e-commerce menggunakan Codeigniter*. Bandung, Jawa Barat: Kreatif, 2020.
- [7] S. Ash and S. Agustian, "Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 pada Media Sosial Twitter menggunakan Metode Logistic Regression," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 99–106, Aug. 2022, doi: 10.37859/coscitech.v3i2.3836.
- [8] M. I. Aditama, I. P. Rizqeya, K. H. U. Wiwaha, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19," 2020.
- [9] A. S. Jeremy, M. F. Tresna, and N. Aryo, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," Aug. 2019.
- [10] I. D. Dinov, *Data Science and Predictive Analytics: Biomedical and Health Applications using R*. in The Springer Series in Applied Machine Learning. Springer International Publishing, 2023. [Online]. Available: <https://books.google.co.id/books?id=hrquEAAAQBAJ>
- [11] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 3, no. 3, pp. 394–401, Oct. 2018, doi: 10.30591/jpit.v3i3.977.
- [12] P. D. Duei, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.
- [13] A. A. Permana et al., *Machine Learning*. Padang Sumatera Barat: Global Eksekutif Teknologi, 2023.
- [14] J. Suntoro, *Data mining: algoritma dan implementasi dengan pemrograman PHP*, Pertama. Jakarta: Elex Media Komputindo, 2019.

- [15] S. Smys, R. Bestak, R. Palanisamy, and I. Kotuliak, *Computer Networks and Inventive Communication Technologies: Proceedings of Fourth ICCNCT 2021*. in Lecture Notes on Data Engineering and Communications Technologies. Springer Nature Singapore, 2021. [Online]. Available: <https://books.google.co.id/books?id=sdFCEAAAQBAJ>
- [16] M. M. Farhan and U. R. Febrian, "Analisis Sentimen Mahasiswa Terhadap Fasilitas Universitas Telkom Menggunakan Metode Jaringan Saraf Tiruan dan TF-IDF," 2019.
- [17] N. I. Aliffiyanti, I. Ernawati, and Y. Widiastiwi, "Klasifikasi Diagnosis Penyakit Stroke Dengan Menggunakan Metode Random Forest," 2022.
- [18] R. F. Nur and S. D. Harini, "Klasifikasi Berita Hoax Pilpres Menggunakan Metode Modified K-Nearest Neighbor dan Pembobotan Menggunakan TF-IDF," 2019.
- [19] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.
- [20] E. B. Susanto, P. A. Christianto, M. R. Maulana, and S. W. Binabar, "Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 234–241, Dec. 2022, doi: 10.37859/coscitech.v3i3.4343.

