

Implementasi Algoritme Multinomial Naïve Bayes Pada Analisis Sentimen Terhadap Isu Presiden 3 Periode

Andre Kautsar^{1*}, Mohammad Syafrullah²

^{1,2}Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia

Email: ^{1*}andre.ktsr9@gmail.com, ²mohammad.syafrullah@budiluhur.ac.id
(* : corresponding author)

Abstrak-Negara Indonesia dalam waktu dekat akan melaksanakan pemilihan umum (pemilu) presiden pada tahun 2024 mendatang. Pemilihan umum ini diselenggarakan setiap 5 tahun sekali menurut UUD 1945. Namun, dikala persiapan pemilu presiden ini muncul narasi untuk memperpanjang masa jabatan presiden Indonesia atau dikenal presiden 3 periode. Narasi tersebut tentunya berseberangan dengan aturan hukum yang berlaku dan mendapat beragam komentar dari masyarakat, terutama pada media sosial. Penentuan sentimen dari media sosial akan tidak efektif jika dilakukan secara manual. Atas dasar itu, penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan pandangan masyarakat tentang isu perpanjangan masa jabatan presiden Indonesia, serta berapa akurasi dari algoritme yang diimplementasikan dalam aplikasi berbasis *website*. Analisis dilakukan melalui proses klasifikasi data yang mengandung sentimen masyarakat, ke dalam kelas positif dan negatif. Dataset yang dianalisis bersumber dari media sosial Twitter berupa *tweets* berbahasa Indonesia. Algoritme yang digunakan untuk klasifikasi *tweets* adalah Multinomial Naïve Bayes karena kecepatan dan kesederhanaannya dalam klasifikasi data. Hasil uji coba pada 21 data uji, didapat 13 sentimen positif dan 8 sentimen negatif pada periode akhir bulan Maret hingga April 2022. Perhitungan menggunakan *confusion matrix* menghasilkan 86% akurasi, 86% presisi, dan 88% *recall* dengan pemecahan data terbaik yaitu 90% data latih dan 10% data uji.

Kata Kunci: analisis sentimen, multinomial naïve bayes, akurasi

Implementation Of Multinomial Nave Bayes Algorithm In Sentiment Analysis Of Presidential Issues 3 Period

Abstract-Soon, Indonesia will hold a presidential general election (election) in 2024. This general election is held every 5 years according to the 1945 Constitution. However, during the preparation for the presidential election, a narrative emerged to extend the term of office of the Indonesian president or known as the 3rd term of president. The narrative is of course contrary to the applicable legal rules and has received various comments from the public, especially on social media. Determination of sentiment from social media will be ineffective if it is done manually. On that basis, this study aims to analyze and classify public views on the issue of extending the term of office of the Indonesian president, as well as the accuracy of the algorithms implemented in web-based applications. The analysis is carried out through a process of classifying data containing public sentiment, into positive and negative classes. The dataset analyzed was sourced from Twitter social media in the form of Indonesian-language tweets. The algorithm used for tweets classification is Multinomial Naïve Bayes because of its speed and simplicity in data classification. The test results on 21 test data, obtained 13 positive sentiments and 8 negative sentiments in the period end of March to April 2022. Calculations using the confusion matrix resulted in 86% accuracy, 86% precision, and 88% recall with the best data solving of 90% data training and 10% test data.

Keywords: Sentiment Analysis, Multinomial Nave Bayes, Accuracy

1. PENDAHULUAN

Indonesia merupakan negara yang menerapkan sistem demokrasi, di mana penerapan sistem pemerintahannya dilaksanakan atas suara rakyat sebagai kedaulatan tertinggi [1]. Hal ini tercermin dari sistem pemilihan umum (pemilu) yang mengambil suara tertinggi, mulai dari kepala daerah, bupati, gubernur dan termasuk memilih presiden dan wakil presiden. Pemilihan umum presiden Indonesia ini biasa diselenggarakan secara periodik.

Dalam waktu dekat, Indonesia akan memasuki babak baru pelaksanaan pemilu presiden pada tahun 2024 mendatang. Namun, ditengah persiapan pesta demokrasi ini muncul isu perpanjangan masa jabatan presiden Indonesia atau lebih dikenal presiden 3 periode. Pembicaraan hangat ini sudah lama mencuat dan sekarang telah didengungkan kembali oleh para tokoh *elite* politik Indonesia, dengan beragam argumen dan mengklaim pribadi mempunyai data akan dukungan penundaan tersebut [2]. Rencana penundaan pemilu presiden secara tidak langsung akan memaksa untuk mengamandemen konstitusi lama dan tentunya akan menuai pro kontra dari masyarakat, terutama di media sosial.

Twitter menjadi satu dari sekian *platform* media sosial yang banyak dimanfaatkan masyarakat untuk beropini. Terdapat 18,45 juta pengguna Twitter per januari 2022 lalu, capaian ini menempatkan Indonesia sebagai negara pengguna Twitter terbanyak ke-5 di dunia [3]. Layanan jejaring sosial berbasis mikroblog daring ini juga kerap

kali dimanfaatkan untuk sumber data penelitian karena kemudahan akses, penyajian data yang *realtime*, dan jumlah karakter sebatas 280 karakter sehingga di dapat kalimat singkat, padat dan jelas yang memudahkan penganalisaan sentimen.

Pencarian sentimen masyarakat dari suatu *tweet* dapat dilakukan secara manual, tetapi akan jauh lebih tidak efektif seiring dengan jumlah opini yang bertambah banyak. Oleh karena itu, penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan pandangan masyarakat tentang isu perpanjangan masa jabatan presiden Indonesia, serta berapa akurasi dari algoritme yang diimplementasikan dalam aplikasi berbasis *website*.

Analisis sentimen adalah sebuah metode yang digunakan untuk mengekstrak, memahami serta mengolah tekstual data secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini [4]. Analisis sentimen akan melakukan klasifikasi sifat dari kalimat dengan melihat kecenderungan pendapat terhadap suatu masalah, apakah kalimat tersebut bermakna positif atau negatif.

Penggunaan algoritme dalam melakukan klasifikasi sentimen pada penelitian ini adalah algoritme Naïve Bayes. Satu dari model Naïve Bayes yang terkenal adalah Multinomial Naïve Bayes. Model ini memperhitungkan frekuensi setiap kata yang muncul pada dokumen [5]. Multinomial Naïve Bayes juga digunakan karena kecepatan dan kesederhanaannya dalam klasifikasi kalimat.

Ada banyak penelitian sebelumnya yang terkait penelitian ini. Pertama, penelitian yang melakukan perbandingan klasifikasi menggunakan Naïve Bayes dan Multinomial Naïve Bayes pada berita *hoax* kategori kesehatan berbahasa Indonesia secara otomatis yang dilakukan oleh [6]. Lalu ada penelitian oleh [7] tentang klasifikasi sentimen sara, hoaks dan radikal pada media sosial Twitter dengan mengimplementasikan Naïve Bayes Multinomial Text. Kemudian, pada penelitian [8] tentang opini masyarakat Indonesia terhadap isu *new normal* skenario dan penelitian yang dilakukan oleh [9] mengenai isu radikalisme menggunakan Naïve Bayes. Adapula penelitian yang terkait penanganan Covid-19 oleh [10] dan vaksinasi Covid-19 dari penelitian [11] yang keduanya menggunakan algoritme Naïve Bayes. Selanjutnya, analisis sentimen mengenai isu RUU KPK menggunakan *Support Vector Machine* (SVM) yang diteliti oleh [12]. Terakhir, membahas seputar polemik persepakbolaan Indonesia oleh [13] yang menerapkan algoritme *K-Nearest Neighbor*.

1.1 Pembobotan TF-IDF

Pembobotan kata dengan TF-IDF terdapat dua bagian proses yaitu TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*). TF ialah frekuensi kemunculan tiap kata pada sebuah dokumen yang semakin banyak kata muncul pada tiap dokumen maka semakin besar nilai TF, sedangkan IDF merupakan frekuensi nilai dokumen yang berbanding terbalik terhadap TF.

Pada penelitian ini, menerapkan *class TfidfVectorizer()* pada pustaka Sklearn, dimana pada perhitungan IDF ditambahkan 1 pada numerator dan denominator untuk mencegah pembagian dengan angka nol. Adapun rumus perhitungan TF-IDF seperti pada persamaan rumus (1) berikut :

$$W_{t,d} = tf_{t,d} \times idf_t = tf_{t,d} \times \text{LN} \frac{(N + 1)}{(df_t + 1)} + 1 \quad (1)$$

Keterangan :

$W_{t,d}$: Nilai bobot TF-IDF
$tf_{t,d}$: Frekuensi kemunculan kata
idf_t	: Nilai <i>inverse</i> frekuensi dokumen tiap kata
df_t	: Frekuensi dokumen tiap kata
N	: Nilai total dokumen

Kemudian, hasil nilai bobot tiap kata ini akan dilakukan normalisasi kata dengan *Euclidean Norm*. Adapun perhitungannya dapat dilihat pada persamaan rumus (2) berikut :

$$v_{norm} = \frac{v}{\|v\|_2} = \frac{v}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \quad (2)$$

Keterangan :

v_{norm}	: Nilai normalisasi tiap kata
v	: Bobot tiap kata

1.2 Multinomial Naïve Bayes

Multinomial Naïve Bayes adalah sebuah metode klasifikasi yang memperhitungkan kemunculan setiap kata pada dokumen, kemudian mencari probabilitas hipotesis menggunakan nilai probabilitas suatu kelas berdasarkan dari atribut yang dimiliki dengan penentuan nilai kelas tertinggi. Rumus umum pencarian probabilitas dapat diamati dengan persamaan (3) berikut :

$$P(c_j|w_i) = \frac{P(c_j) P(w_i|c_j)}{P(w_i)} \quad (3)$$

Keterangan :

- $P(c_j|w_i)$: Probabilitas kategori c_j jika terdapat kemunculan kata w_i
- $P(c_j)$: Probabilitas kemunculan kelas *prior*
- $P(w_i|c_j)$: Probabilitas sebuah kata w_i yang ada pada kelas c_j
- $P(w_i)$: Probabilitas kemunculan kata w_i

Pada proses perhitungan klasifikasi, Probabilitas kemunculan kata w_i dapat dihapus sebab tidak memiliki pengaruh terhadap perbandingan hasil klasifikasi setiap kelas. Kemudian, pencarian nilai *prior* untuk setiap kelas dapat dihitung dengan persamaan rumus (4) berikut :

$$P(c_j) = \frac{\text{count}(c_j)}{N} \quad (4)$$

Keterangan :

- $\text{count}(c_j)$: Banyaknya kelas c_j yang ada pada dokumen latihan
- N : Nilai seluruh kelas dalam kata

Perhitungan *posterior* merupakan hasil kali nilai *prior* dengan *likelihood* atau probabilitas kelas c_j terhadap kata w_i dengan persamaan (5) berikut :

$$P(c_j|w_i) = P(c_j) \times P(w_i|c_j) \times \dots \times P(w_n|c_n) \quad (5)$$

Pada persamaan (5) akan muncul permasalahan dimana data berisi fitur dengan model belum diketahui sebelumnya, sehingga dalam proses *training* akan mengarah ke-0 probabilitas apriori dan itu tidak dapat memprediksi suatu kelas. Maka, digunakan pendekatan parameter *Laplace Smoothing* untuk menghindari probabilitas nol dengan menambahkan nilai 1 ke dalam setiap perhitungan, sehingga persamaan *Laplace Smoothing* dapat diamati pada persamaan (6) berikut :

$$P(w_i|c_j) = \frac{\text{count}(w_i|c_j) + 1}{(\sum \text{count}(w|c_j) + |V|)} \quad (6)$$

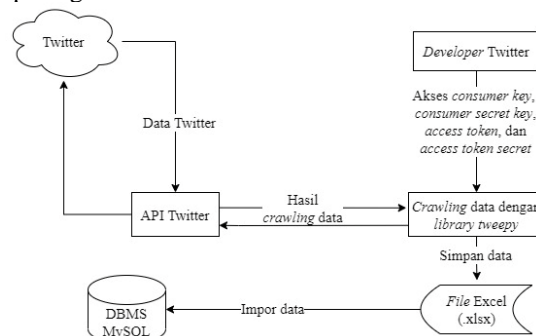
Keterangan :

- $P(w_i|c_j)$: *Likehood* atau probabilitas kata ke- i yang muncul pada kelas- j
- $\text{count}(w_i|c_j)$: Nilai kemunculan kata w pada kelas c_j
- $\sum \text{count}(w|c_j)$: Nilai seluruh kata dalam kelas c_j
- $|V|$: Nilai kata unik yang ada pada seluruh kelas

2. METODE PENELITIAN

2.1 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan teknik *crawling* data dan dibantu oleh pustaka Tweepy yang memanfaatkan API Twitter. API Twitter ini dapat diakses dengan cara mendaftar dan membuat proyek pada akun *Developer* Twitter untuk mendapatkan *consumer key*, *consumer secret key*, *access token*, dan *access token secret* sebagai syarat pengambilan data Twitter. Kemudian, dilakukan *crawling* data pada periode 27 Maret hingga 25 April 2022 dengan kata kunci 'jokowi', '3 periode', 'presiden' dan 'jabatan'. Setelah proses *crawling*, data siap disimpan dalam bentuk *file* Excel untuk masuk pada proses lebih lanjut. Ilustrasi metode pengumpulan data dapat dilihat pada gambar 1.



Gambar 1. Metode Pengumpulan Data

Data yang sudah terkumpul akan diterapkan teknik *stratified sampling* ketika proses pembagian data pada saat memasuki tahapan klasifikasi. *Stratified sampling* didefinisikan sebagai suatu teknik pengambilan sampel dengan melihat suatu tingkatan (*strata*) pada data, yang dalam hal ini adalah data Twitter. Pada penelitian ini, data yang dilakukan pembagian menjadi data latih dan data uji digunakan beberapa presentase rasio pembagian data untuk mendapat akurasi tertinggi. Rasio pembagian data dapat dilihat pada tabel 1.

Tabel 1. Rasio Pembagian Data

Data Uji	Data Latih
90%	10%
80%	20%
70%	30%
60%	40%
50%	50%

2.2 Pelabelan

Pada tahap ini, mulai masuk proses pelabelan kelas sentimen pada setiap *tweet* sebelum pembobotan kata dan pembagian data. Pelabelan ini dilakukan secara manual berdasarkan subjektivitas terhadap suatu kalimat secara satu per satu pada *tweet*. Kalimat *tweet* yang maksud dan tujuan berupa ajakan untuk mendukung penolakan isu perpanjangan masa jabatan presiden Indonesia dilabeli sentimen positif. Sedangkan, kalimat yang mengandung hasutan atau ajakan menuju pada arah dukungan isu akan dilabeli sentimen negatif.

2.3 Text Preprocessing

Data yang sudah terkumpul bersumber dari Twitter dalam proses *crawling* masih memiliki struktur teks yang tidak beraturan atau lebih banyak teks yang mengandung *noise*. Struktur teks yang seperti ini mempengaruhi proses *text mining* pada tahapan berikutnya. Berdasarkan hal tersebut, proses *preprocessing* ini bertujuan agar data tidak terstruktur diubah menjadi data yang terstruktur. Proses ini bertujuan untuk menghasilkan *tweet* bersih yang nantinya dapat diolah untuk memberikan informasi yang bermakna. Langkah-langkah *preprocessing* untuk penelitian ini adalah :

a. Case folding

Pada tahap ini dilakukan penyeragaman seluruh teks *tweets* dari huruf berkapital menjadi huruf kecil. Seperti contoh kata "Fakta" diubah menjadi "fakta" dan seterusnya.

b. Cleansing

Tahapan *cleansing* dilakukan dengan tahapan penghapusan *noise* pada teks *tweets* yang berisi seperti angka, url, dan simbol lainnya. Proses *cleansing* pada penelitian ini terdiri dari beberapa tahapan sebagai berikut :

1. Menghapus URL

URL atau tautan kerap kali muncul pada data Twitter yang biasanya tersisipkan pada teks *tweets*. URL atau tautan ini terkadang mengarah pada halaman promosi suatu produk, berita atau bahkan lebih buruknya halaman yang disediakan untuk pencurian data pribadi (*phising*). Keberadaan URL dianggap sebagai data yang tidak berarti dan tidak valid serta harus dihapus. Contoh URL <https://t.co/V24ffF167h> dihapus saat mulai memasuki tahapan *cleansing*.

2. Menghapus Karakter Selain Huruf

Setelah proses penghapusan URL selesai, dilanjut dengan tahapan menghapus semua karakter non alfabet seperti angka, *username* atau *mention* (@), tanda pagar atau *hashtag* (#), *delimiter* seperti koma, titik atau tanda baca yang terkait dan beragam simbol lainnya. Semua karakter non alfabet ini jika dibiarkan akan menjadi *noise* pada sebuah dokumen.

3. Menghapus 1 Karakter

Tahapan terakhir dan tidak kalah pentingnya dari proses *cleansing* adalah menghapus 1 karakter pada sebuah kalimat, teks atau dokumen. Hal ini dianggap tidak memiliki arti jelas dan mempengaruhi pada tahapan pembobotan kata. Seperti contoh pada kalimat "y kali tidak maju" mungkin dimaksudkan penulis adalah "iya kali tidak maju", maka huruf "y" harus dihilangkan pada tahapan ini.

c. Normalization

Tahapan ini untuk menormalisasikan kata dari setiap kata yang mengandung kata tidak baku, kata gaul atau singkatan tertentu menjadi kata baku. Proses ini melibatkan kamus *slangwords* yang terdapat dalam basis data, ditambahkan secara manual dengan kata - kata yang berkaitan dengan isu politik yang diteliti. Misalnya, kata "tdk" diubah menjadi kata "tidak", kata "yg" menjadi kata "yang" dan seterusnya.

d. Stopwords Removal

Proses berikutnya, dilanjut dengan proses *stopwords removal* atau penghapusan kata - kata yang dianggap tidak penting atau tidak memberi pengaruh yang besar, biasanya kata umum yang kemunculannya sering ditemui pada dokumen. Pada penelitian ini, memanfaatkan daftar *stopword* berbahasa Indonesia bersumber dari penelitian

Fadillah Z. Tala. Daftar ini juga dilakukan penambahan secara manual dengan kata - kata yang berkaitan dengan isu politik yang sedang diteliti saat ini. Misalnya, kata “ada” yang ternyata ada dalam kamus *stopwords* akan dihilangkan dalam suatu dokumen

e. *Stemming*

Tahapan berikutnya yaitu proses *stemming* pada setiap kata berimbuhan pada suatu dokumen menjadi kata dasar menggunakan bantuan pustaka Sastrawi yang menerapkan Algoritme Nazief dan Adriani. Contohnya kata “perpanjangan” akan diubah menjadi kata “panjang” dengan menghapus kata “per-” dan kata “-an” dan seterusnya pada kata yang mengandung imbuhan.

2.4 Pembobotan Kata

Proses ini merupakan proses pemberian bobot pada setiap kata pada dokumen atau teks *tweets* untuk mengetahui hubungannya dalam sebuah sekumpulan data pada dokumen. Proses pembobotan ini harus melalui tahap pencarian kata unik dari *tweets* yang sudah melalui tahap *preprocessing*. Daftar kata ini akan memasuki tahap perhitungan TF (*term frequency*) per dokumen sebuah *tweet* dari kumpulan kata, maka didapat nilai kemunculan kata. Selanjutnya, dilakukan perhitungan DF(*document frequency*) yaitu frekuensi kalimat yang mengandung kata tertentu pada dataset. Kemudian, menghitung IDF (*inverse document frequency*) dengan cara membagi jumlah data *tweets* dengan DF. Terakhir, mengalikan nilai TF dengan IDF dan menormalisasi kata dengan *Euclidean Norm*.

2.5 Klasifikasi

Tahap klasifikasi pada penelitian ini menggunakan Naïve Bayes dengan model Multinomial. Tahapan ini melakukan pelatihan terhadap dokumen yang sudah melewati tahap pelabelan dan *preprocessing* serta proses perhitungan bobot kata dengan metode pembobotan TF-IDF untuk membuat model dari data latih. Pembuatan model ini juga termasuk tahap pembelajaran dengan cara menganalisis data latih.

2.6 Pengujian

Pada pengujian ini, dilakukan untuk mengetahui tingkat akurasi, presisi dan *recall* dari klasifikasi menggunakan Multinomial Naïve Bayes. Pengujian ini menggunakan perbandingan data hasil prediksi terhadap sekumpulan data aktual dengan bantuan *Confusion Matrix*.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dataset penelitian ini bersumber dari media sosial Twitter berupa teks. Dataset tersebut diperoleh melalui proses *crawling* menggunakan pustaka Tweepy pada periode 27 Maret hingga 25 April 2022 menggunakan kata kunci ‘jokowi’ ‘3 periode’, ‘presiden’ dan ‘jabatan’. Hasil pengumpulan data Twitter ini berjumlah 206 *tweets* dengan informasi yang terdiri dari *created_at*, *username* dan *tweets*. Dari *tweets* yang terkumpul dilakukan pelabelan secara manual dengan memperhatikan subjektivitas keseluruhan kalimat. Hasil dari pengumpulan dataset dapat dilihat pada tabel 2.

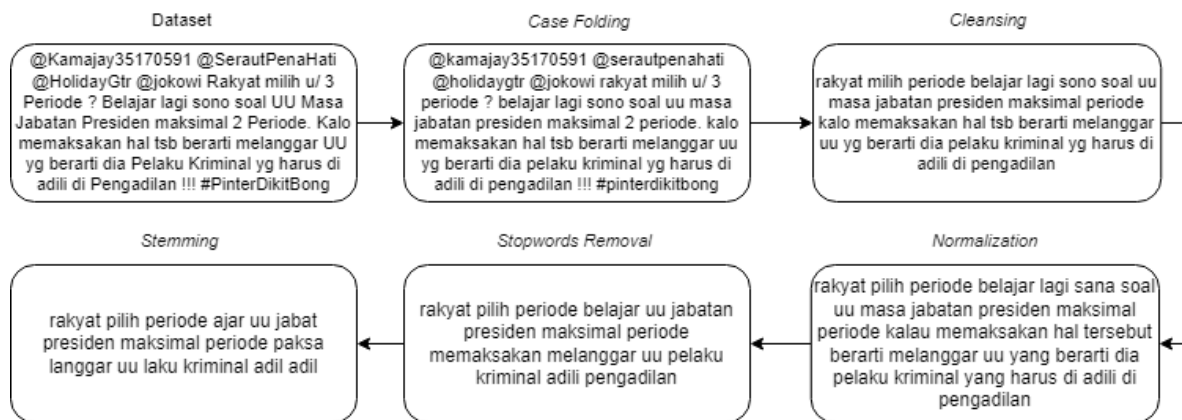
Tabel 2. Sampel dataset

No.	Created_at	Username	Tweets	Label
1.	2022-04-04 19:47:31	pemudaindohebat	"Tolak...penundaan pemilu 2024 atau perpanjang masa jabatan presiden 2 tahun. #sukseskanpemilu2024 #regenerasinasional2024	Positif
2.	2022-04-10 17:40:58	IDCorner	https://t.co/bospkgnDQ3" "Jokowi tak Mau Ada Spekulasi 3 Periode Jabatan Presiden - https://t.co/sBEn3qBE7P	Positif
3.	2022-04-11 09:31:29	Sherly0ctaviani	#IDCORNER" @rezkydoank @jokowi @airlangga_hrt @TitoKarnavian @bahlillahadalia Presiden menolak dengan tegas perpanjangan masa jabatan 3 Periode.	Positif

3.2 Preprocessing Data

Data yang sudah terkumpul bersumber dari Twitter dalam proses *crawling* masih memiliki struktur teks yang tidak beraturan atau lebih banyak teks yang mengandung *noise*. Struktur teks yang seperti ini mempengaruhi proses *text mining* pada tahapan berikutnya. Maka, tahap *preprocessing* harus mengubah data tidak terstruktur menjadi

data terstruktur. Proses ini bertujuan untuk menghasilkan *tweets* bersih agar dapat diproses pada tahapan selanjutnya dan menghasilkan informasi yang bermakna. Tahapan *preprocessing* dapat dilihat pada gambar 2.



Gambar 2. Tahapan *Preprocessing*

Setelah semua tahap *preprocessing* telah dilalui. Data *tweets* ini sudah terbebas dari beragam *noise* dari hasil *crawling* data dan lebih terstruktur. Data sampel yang sudah bersih untuk selanjutnya dilakukan proses pembobotan kata dapat dilihat pada tabel 3 berikut :

Tabel 3. Hasil *preprocessing*

No.	Created at	Username	Tweets	Label
1.	2022-04-04 19:47:31	pemudaindohebat	tolak tunda milu panjang jabat presiden	Positif
2.	2022-04-10 17:40:58	IDCorner	jokowi spekulasi periode jabat presiden	Positif
3.	2022-04-11 09:31:29	Sherly0ctaviani	presiden tolak panjang jabat periode	Positif

Selanjutnya, tahapan pembobotan kata untuk mengubah kata menjadi bilangan numerik yang dapat dihitung menggunakan TF-IDF. Konsep ini dengan mencari seberapa banyak kemunculan kata dalam suatu dokumen. Hasil pembobotan dapat dilihat pada tabel 4.

Tabel 4. Hasil pembobotan TF-IDF

Dokumen	TF - IDF (Normalized)								
	jabat	jokowi	milu	panjang	periode	presiden	spekulasi	tolak	tunda
D1	0,341	0	0,559	0,424	0	0,341	0	0,424	0,341
D2	0,327	0,553	0	0	0,419	0,327	0,338	0	0
D3	0,380	0	0	0,487	0,487	0,380	0	0,487	0

3.3 Pengujian Model

Pengujian ialah satu dari hal yang perlu dilakukan di setiap proses pengembangan sebuah sistem untuk dilakukan evaluasi mengenai tingkat keberhasilan aplikasi yang dibuat berdasarkan nilai tertentu. Penelitian ini melakukan pengujian dengan perhitungan termasuk dari nilai akurasi, presisi dan *recall* pada penerapan algoritme Naïve Bayes dengan model Multinomial dalam memprediksi label data uji. Pada tabel 5, menunjukkan hasil pengujian menggunakan metode *confusion matrix* dengan rasio pembagian data yang telah ditentukan.

Tabel 5. Hasil pengujian pembagian data

Rasio Pembagian Data	Data Latih
90% : 10%	86%
80% : 20%	79%
70% : 30%	76%
60% : 40%	77%
50% : 50%	73%

Berdasarkan tabel 5, diketahui bahwa pembagian data dengan 90% data latih dan 10% data uji memiliki tingkat akurasi tertinggi sebesar 86%. Data latih yang digunakan sejumlah 185 data, dimana 93 data berlabel positif dan 92 data berlabel negatif. Hasil nilai prediksi dapat ditunjukkan pada tabel 6 :

Tabel 6. Sampel hasil prediksi

No.	Tweets	Label (Aktual)	Label (Prediksi)
1.	NEWS - Sebut Akan Taat Konstitusi, Presiden Jokowi Lagi-lagi Dinilai “Bersayap” Terkait perpanjangan masa jabatan presiden 3 periode, Managing Director PPPI, Ahmad Khoiril Umam, menilai, pernyataan Presiden @jokowi kembali “bersayap”. https://t.co/KyeFdOFPXV	Negatif	Positif
2.	Ternyata, Menko Marves Luhut Binsar Pandjaitan merupakan Ketua Dewan Pembina APDESI yang mendukung Jokowi Presiden 3 periode #ApdesiDukungJokowi https://t.co/nzKdMgpnGY	Negatif	Negatif
3.	@manghari500 @dapitnih Kelihatan nya Jokowi masih blusukan untuk mencari dukungan rakyat yg gak ngerti apa" untuk dijadikan dalih tap mpr jabatan presiden 3 periode.		
...
21.	Faldo mengingatkan Presiden Jokowi telah menegaskan sikapnya tentang perpanjangan masa jabatan presiden. Jokowi, kata Faldo, akan taat kepada aturan di konstitusi. https://t.co/pXQQ6OnI5A lewat @SPOSTid #Jokowi @FaldoMaldini #SPOSTid	Positif	Positif

Pada rasio pembagian data dengan akurasi tertinggi, digunakan data uji sebesar 21 data. *Output* dari nilai kalkulasi pengujian tersebut ditampilkan dalam tabel *confusion matrix* 2x2, perhatikan tabel 7 berikut :

Tabel 7. Confusion Matrix 2x2

Actual Values	Predicted Values		Total
	Positive	Negative	
Positive	10	3	13
Negative	0	8	8
Total	18	11	21

Setelah mengetahui hubungan nilai aktual dan nilai prediksi, maka dapat diperoleh nilai akurasi, presisi dan *recall* dengan *detail* perhitungan evaluasi pada tabel 8 berikut ini :

Tabel 8. Perhitungan evaluasi

Perhitungan Evaluasi		
Akurasi	$((10 + 8) / (10 + 8 + 0 + 3)) \times 100\%$	86%
Presisi	$((10 / (10 + 0)) + (8 / (8 + 3))) / 2 \times 100\%$	86%
Recall	$((10 / (10 + 3)) + (8 / (8 + 0))) / 2 \times 100\%$	88%

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa Multinomial Naïve Bayes bekerja dengan baik dalam mengklasifikasikan data berbentuk dokumen. Sebanyak 206 *tweets*, didapat pembagian data terbaik yaitu 90% data *training* dan 10% data *testing* dengan perolehan akurasi sebesar 86%, distribusi data latih sejumlah 185 data (93 bersentimen positif dan 92 bersentimen negatif) dan data uji sebesar 21 data yang menghasilkan nilai prediksi 13 sentimen positif dan 8 sentimen negatif. Maka, sentimen positif lebih dominan dari sentimen negatif tentang isu presiden 3 periode. Ini berarti sebagian besar menolak isu tersebut pada periode 27 Maret hingga 25 April 2022. Dan juga, hasil pengujian dengan akurasi tertinggi menghasilkan rata - rata presisi 86% dan rata - rata *recall* 88%.

Adapun saran yang didapat sebagai pengembangan lebih lanjut ialah diperlukan pencarian data Twitter dengan kata kunci yang lebih bervariasi sehingga menghasilkan pandangan lebih beragam. Penambahan kamus kata pada daftar *slangwords* dan *stopwords* juga harus diimbangi dengan keberagaman bahasa yang digunakan pada *tweets*. Terakhir, melakukan proses pelabelan manual direkomendasikan dengan bantuan pakar atau ahli dibidang politik dan bahasa Indonesia.

DAFTAR PUSTAKA

- [1] E. Purnamawati, “Perjalanan Demokrasi Di Indonesia”, Solusi, vol. 18, no. 2, pp. 251–264, 2020.
- [2] (2022), Kompas.com. [Online]. Available: <https://nasional.kompas.com/read/2022/03/16/07030041/isu-perpanjangan-masa-jabatan-presiden-dan-wacana-rasa-orde-baru?page=all>
- [3] (2022), Katadata.co.id, [Online]. Available: <https://databoks.katadata.co.id/datapublish/2022/03/23/pengguna-twitter-indonesia-masuk-daftar-terbanyak-di-dunia-urutan-berapa>

- [4] F. V. Sari dan A. Wibowo, “Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi”, *SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [5] A. Rahman dan A. Doewes, “Online News Classification Using Multinomial Naive Bayes”, *ITSMART*, vol. 6, no. 1, pp. 32–38, 2017.
- [6] C. N. Harahap, G. I. Marthasari, and N. Hayatin, “Perbandingan Klasifikasi Berita Hoax Kategori Kesehatan Menggunakan Naïve Bayes dan Multinomial Naïve Bayes”, *Repositor*, vol. 3, no. 4, pp. 419–424, 2021.
- [7] F. E. Purwiantono dan A. Aditya, “Klasifikasi Sentimen Sara, Hoaks Dan Radikal Pada Postingan Media Sosial Menggunakan Algoritma Naive Bayes Multinomial Text”, *Tekno Kompak*, vol. 14, no. 2, p. 68, 2020.
- [8] M. N. Rusardi, B. Rahayudi, dan P. P. Adikara, “Analisis Sentimen Masyarakat terhadap Isu New Normal Scenario berdasarkan Opini dari Twitter menggunakan Algoritma Naive Bayes Classifier”, *Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 3, pp. 1434–1440, 2022.
- [9] T. A. Sundara dan S. Ekaputri Arnas, “Naïve Bayes Classifier untuk Analisis Sentimen Isu Radikalisme”, *Prosiding Seminar Nasional Sisfotek (Sistem Inf. dan Teknol. Informasi)*, pp. 93–98, 2020.
- [10] Yuyun, N. Hidayah, dan S. Sahibu, “Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter”, *RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, 2021.
- [11] Z. Firmansyah dan N. F. Puspitasari, “Analisis Sentimen Masyarakat Terhadap Vaksinasi COVID-19 Berdasarkan Opini Pada Twitter Menggunakan Algoritma Naive Bayes”, *Teknik Informatika*, vol. 14, no. 2, pp. 171–178, 2021.
- [12] R. Nooraeni, H. D. Sariyanti, A. F. F. Iskandar, S. F. Munawwaroh, S. Pertiwi, dan Y. Ronaldias, “Analisis Sentimen Data Twitter Mengenai Isu RUU KPK Dengan Metode Support Vector Machine (SVM)”, *Paradigma - J. Komputer dan Informatika*, vol. 22, no. 1, pp. 55–60, 2020.
- [13] J. A. Septian, T. M. Fahrudin, dan A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor”, *Journal of Intelligent Systems and Computer*, vol. 1, no. 1, pp. 43–49, 2019.