

Analisis Sentimen terhadap Perpanjangan Masa Jabatan Presiden Indonesia Menggunakan Algoritma Naïve Bayes

Titan Nugraha¹, Purwantoro², Yuyun Umaidah³

^{1,2,3} Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang

Email: titan.nugraha18124@student.unsika.ac.id¹, purwantoro.masbro@staff.unsika.ac.id², yuyun.umaidah@staff.unsika.ac.id³

Abstrak

Media sosial twitter sering digunakan sebagai media untuk menyampaikan pendapat terhadap pemerintahan di Indonesia. Banyak juga hal – hal yang kontroversial di media sosial twitter terhadap pemerintah, seperti halnya pada hari ini sedang terjadi kontrovesi dimana ada sebuah usulan untuk menjadikan masa jabatan Presiden di Indonesia yaitu tiga periode, yang dimana tadinya hanya bisa menjabat sampai dua periode atau sepuluh tahun saja untuk satu orang Presiden. Opini atau sentimen masyarakat yang dalam istilah Twitter biasa disebut sebagai “ciutan” ini dapat berupa opini negatif maupun positif. Akan tetapi jumlah data tersebut cukup banyak sehingga dibutuhkan suatu metode untuk mewujudkannya yaitu dengan menggunakan analisis sentimen. Analisis sentimen dapat dijadikan solusi untuk mengolah opini tersebut dengan menggunakan algoritma *Naïve Bayes Classifier*. Pada hasil dari teknik *naïve bayes* memerlukan evaluasi untuk menentukan model terbaik. Pengujian dilakukan dengan tiga model yaitu 70:30, 80:20, dan 90:10 yang kemudian dievaluasi menggunakan *confusion matrix*. Berdasarkan perbandingan hasil pengujian evaluasi, skenario terbaik yang untuk model klasifikasi Naïve Bayes yaitu pada skenario atau model pertama (90% data training dan 10% data testing) dengan nilai *accuracy* sebesar 95% , nilai *precision* sebesar 97%, nilai *recall* sebesar 96%, dan nilai *f-measure* sebesar 96%.

Kata kunci: *Analisis Sentimen, Naïve Bayes, Presiden Indonesia, Sentimen skor, Twitter*

Abstract

Twitter social media is often used as a medium to express opinions on the government in Indonesia. There are also many controversial things on Twitter social media against the government, as is the case today there is a controversy where there is a proposal to make the term of office of the President in Indonesia three periods, which previously could only serve up to two terms or ten years. for one President. Public opinion or sentiment which in Twitter's terms is commonly referred to as "shrink" can be in the form of negative or positive opinions. However, the amount of data is quite large so it takes a method that can be used to make it happen, namely by using sentiment analysis. Sentiment analysis can be used as a solution to process these opinions using the Naïve Bayes Classifier algorithm. The results of the nave Bayes technique require evaluation to determine the best model. The test is carried out with three models, namely 70:30, 80:20, and 90:10 which are then evaluated using a confusion matrix. Based on the comparison of the evaluation test results, the best scenario for the Naïve Bayes classification model is in the first scenario or model (90% training data and 10% testing data) with an accuracy value of 95%, a precision value of 97%, a recall value of 96%, and the f-measure value is 96%.

Keywords : *Sentiment Analysis, Nave Bayes, President of Indonesia, Sentiment score, Twitter*

PENDAHULUAN

Media sosial twitter sering digunakan sebagai media untuk menyampaikan pendapat terhadap pemerintahan di Indonesia. Dan juga sering dipakai pemerintah untuk mensosialisasikan segala informasi tentang pemerintahan Indonesia. Sehingga masyarakat Indonesia dapat mengetahui informasi yang berkaitan dengan pemerintah. Banyak juga hal – hal yang kontroversial di media sosial twitter terhadap pemerintah, seperti halnya pada hari ini sedang terjadi kontrovesi dimana ada sebuah usulan untuk menjadikan masa jabatan Presiden di Indonesia yaitu tiga periode, yang dimana tadinya hanya bisa menjabat sampai dua periode atau sepuluh tahun saja untuk satu orang Presiden.

Wacana tersebut sedang menjadi kontroversi sekarang , dikarenakan banyak survei yang menyatakan masyarakat banyak yang tidak setuju terhadap wacana tersebut, salah satunya lembaga survei Indonesia Political Opinion (IPO) telah melakukan survei yang dimana menghasilkan 61% responden tidak setuju dan 39% persen setuju (Medistiara, 2022).

Tetapi lain halnya dengan Menteri Koordinator Kemaritiman dan Investasi Luhut B. Pandjaitan menyatakan bahwa dia mempunyai big data yang menghasilkan bahwa suara rakyat Indonesia setuju untuk penundaan pemilu 2024 yang dimana setuju terhadap perpanjangan masa jabatan Presiden. Hal tersebut ia ucapkan didalam *podcast Close the Door* bersama Deddy Corbuzier (Santoso & Fadil Djailaini, 2022).

Opini atau sentimen masyarakat yang dalam istilah Twitter biasa disebut sebagai “ciutan” ini dapat berupa opini negatif maupun positif. Akan tetapi jumlah data tersebut cukup banyak sehingga dibutuhkan suatu metode yang dapat digunakan untuk mewujudkannya yaitu dengan menggunakan analisis sentimen. Analisis sentimen adalah teknik pembersihan data dari kata dan simbol yang tidak relevan dan merubah data yang bernilai kualitatif menjadi data kuantitatif, kemudian data ulasan pengguna akan diklasifikasikan untuk mendapatkan ulasan yang bersifat positif dan negatif (Manik, Ernawati, & Nurlaili, 2021). Saat ini analisis sentimen merupakan salah satu topik yang banyak digunakan para peneliti, yang memiliki tujuan untuk menyediakan informasi dari sebuah dataset yang tidak terstruktur. Menurut (Arsi & Waluyo, 2021) analisis sentimen ini juga bisa diterapkan pada opini semua bidang seperti ekonomi, politik, sosial dan hukum. Analisis sentimen ini juga dapat mengelompokkan polaritas dari teks untuk mengetahui apakah opini pada dokumen termasuk positif atau negatif (Paulina, Bachtiar, & Rusydi, 2020).

Algoritma Naïve Bayes adalah algoritma klasifikasi sederhana yang mana menghitung sekumpulan probabilitas dengan cara menjumlahkan dan mengkombinasikan nilai dari dataset yang diberikan. Metode Naïve Bayes akan digunakan pada penelitian ini dalam proses analisis sentimen masyarakat Indonesia terhadap usulan masa perpanjangan Presiden Indonesia. Dari berbagai referensi penelitian, klasifikasi Algoritma Naïve Bayes lebih banyak disukai dikarenakan kesederhanaannya dan juga kecepatannya (Basit, 2020). Walaupun klasifikasi Algoritma Naïve Bayes bisa dikatakan klasifikasi yang sederhana, tetapi hasil yang didapat dari pada klasifikasi Algoritma Naïve Bayes ini sering sekali mencapai performa yang serupa dengan algoritma klasifikasi lainnya seperti contoh *Neural Network classifier & Decision tree*. Klasifikasi Algoritma Naïve Bayes ini bisa memberikan akurasi yang tinggi juga cepat di dalam memproses data yang ada di dalam jumlah yang sangat banyak (Nugroho & Cholissodin, 2021).

Pada penelitian sebelumnya tentang sentimen masyarakat Indonesia terhadap kebijakan pemerintah tentang PPKM menggunakan metode naïve bayes memperoleh hasil yaitu berupa sentimen positif sebanyak 99% dan sentimen negatif sebanyak 1% atau terlihat dari hasil klasifikasi emosi memiliki mayoritas “*unknown*” atau tidak diketahui yang sebagian besar kata tersebut

memiliki polaritas positif (Taofik Krisdiyanto, 2021). Di penelitian tersebut penulis tersebut mengatakan bahwa kekurangan yang didapatkan oleh kurangnya kata yang ada pada *library* sehingga keakuratan dalam klasifikasi berkurang. Oleh karena itu penulis ingin melakukan penelitian terhadap sentimen masyarakat Indonesia di media sosial twitter menggunakan algoritma naïve bayes.

Media Sosial

Media sosial adalah tempat dunia internet untuk berinteraksi antar pengguna yang ada di beberapa belahan dunia, berfungsi untuk menghubungkan yang jauh untuk menjadi dekat. Media sosial saat ini, yang telah memperoleh basis pengguna yang besar dan digunakan oleh orang-orang di seluruh dunia, memiliki potensi untuk menjadi salah satu teknologi yang menentukan di era kita (Appel, Grewal, Hadi, & Stephen, 2020).

Twitter

Twitter adalah platform media sosial yang menjadi kegemaran para pengguna internet, terutama para pemuda. Dikarenakan di twitter bisa mengetik suatu pendapat atau opini yang biasa disebut "cuitan". Dengan menulis sebuah opini di twitter seseorang dapat berbagi informasi dan menyatakan pendapat secara bebas tanpa terkecuali. Dengan lebih dari sepuluh tahun masa pembangunan, Twitter telah memiliki 330 juta pengguna aktif bulanan yang mengirim sekitar 500 juta tweet tiap harinya (Twitter, 2019).

Presiden

Presiden adalah suatu orang yang memimpin suatu negara, dan juga di temani oleh salah satu orang wakil presiden. Presiden berkewajiban untuk memimpin negara dan mengatur pemerintahan. Selain di temani oleh seorang wakil presiden, presiden juga ditemani oleh beberapa Menteri untuk mengatur pemerintahan.

Di Indonesia sejak merdeka di tahun 1945, terhitung sudah tujuh kali melakukan pergantian Presiden, yang dimana ke 7 Presiden itu ialah ; Soekarno, Soeharto, B.J Habibie, Gusdur, Megawati, Susilo Bambang Yudhoyono, dan Jokowi. Yang sedang menjambat hari ini ialah Jokowi sejak tahun 2014.

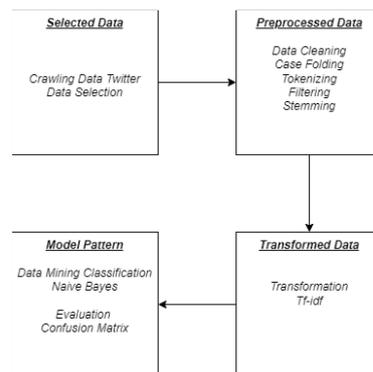
Analisis Sentimen

Analisis sentimen yaitu salah satu pendekatan untuk menganalisis dan mengkategorikan ekspresi tekstual emosi sebagai positif atau negatif disebut "analisis sentimen". Interpretasi pembaca terhadap materi dipengaruhi oleh emosi dan pandangan dunianya sendiridibutuhkan teknik analisis sentimen, dan opini tersebut akan diklasifikasikan menjadi opini positif dan negatif (Badjrie, Pratiwi, & Anggana, 2021)

Naïve Bayes Classifier

Menggunakan probabilitas bersyarat untuk setiap kelas yang mungkin diturunkan dari matriks data, pengklasifikasi *Naïve Bayes* didasarkan pada teorema Bayes. Sebaliknya, asumsi NBC mengandaikan bahwa variabel sistem independen satu sama lain. Ketika diterapkan pada masalah klasifikasi dunia nyata, pengklasifikasi berbasis NBC tetap menghasilkan hasil yang dapat diterima meskipun bergantung pada asumsi yang salah (Marcos de Moraes, Soares, & Machado, 2020).

METODE



Gambar 1 Alur Penelitian

Data Selection

Tahap pertama adalah prosedur seleksi data, dimana teknik data *crawling* akan digunakan untuk memperoleh data yang relevan dengan isu penelitian, khususnya soal perpanjangan tangan presiden Indonesia, dari platform media sosial Twitter. Dengan menggunakan *keyword* "jokowi tiga periode" pada Twitter API, kami dapat mengumpulkan tweet yang relevan untuk dianalisis. Kumpulan data diproses terlebih dahulu ketika diperoleh dari prosedur, yang memerlukan pemilihan data untuk memastikan bahwa data memenuhi persyaratan penelitian ini.

Preprocessing

Tahapan *preprocessing* ini merupakan suatu proses agar membentuk dan membersihkan data sebelum dilakukannya klasifikasi. Berikut adalah tahapan dari *preprocessing*:

1. *Cleaning*

Cleaning merupakan proses pembersihan data diantaranya untuk menghilangkan emoticon, ataupun simbol yang tidak diperlukan.

2. *Case Folding*

Tahapan awal pada proses ini adalah *case folding* yang bertujuan untuk merubah teks menjadi huruf kecil.

3. *Tokenizing*

Selanjutnya adalah tahapan *tokenizing*. Proses ini digunakan untuk melakukan pemecahan kalimat menjadi kata per kata.

4. *Stopword Removal*

Stopword removal memiliki 32 stopwords list yang akan memangkas kata pada data yang mengandung kata pada list tersebut. Stopwords list tersebut berisi kata yang tidak memiliki peran atau nilai pada proses klasifikasi

5. *Stemming*

Proses stemming merupakan proses untuk menghilangkan imbuhan. Setelah melakukan tokenizing maka akan dilakukan penghapusan imbuhan pada hasil token.

Transformation Data

Setelah dokumen diproses, selanjutnya dokumen diubah, yang berarti teks diubah menjadi data tipe numerik yang lebih menggambarkan dokumen. *Term frequency inverse document frequency* (TF-IDF) adalah sistem pembobotan yang diterapkan, dan melakukan tugasnya dengan

menetapkan nilai untuk setiap kata yang mewakili frekuensi kemunculan dokumen untuk kata kunci atau kata yang sering digunakan.

Data Mining

Data Mining adalah langkah selanjutnya. Pada langkah ini, data mentah diubah menjadi intelijen yang dapat ditindaklanjuti. Tujuan dari data mining adalah untuk menemukan dan menganalisis pola tersembunyi dalam data dalam jumlah besar untuk wawasan yang praktis dan dapat ditindaklanjuti.

Urutkan kumpulan sentimen ke dalam kelas positif dan negatif menggunakan metode klasifikasi. Naive Bayes, algoritma klasifikasi yang terkait dengan teorema Bayes. Dalam melakukan klasifikasi data akan di pisah menjadi data training dan testing.

Evaluation

Langkah terakhir adalah pemeriksaan model yang dibangun untuk memeriksa konsistensi dengan asumsi yang sudah ada sebelumnya. Menerapkan *confusion matrix* memungkinkan kami untuk mengukur akurasi algoritme. Hal ini disebabkan oleh kemampuan *confusion matrix* untuk menggambarkan prediksi model dan implementasi aktual dari algoritma. Untuk memberikan gambaran kinerja yang lebih jelas, matriks konfusi memungkinkan analisis di empat dimensi:

1. *Accuracy* : Untuk mengetahui tingkat kemiripan antara nilai prediksi dengan nilai aktual.
2. *Precision* : Untuk mengetahui seberapa tepat atau akurat model dari yang diprediksi positif. *Precision* juga merupakan ukuran yang baik untuk menentukan ketika nilai *False Positive* pada suatu model tinggi.
3. *Recall* : Untuk menghitung berapa banyak dari nilai *Actual Positive* yang berhasil diidentifikasi oleh model melalui pelabelan sebagai *True Positive*. *Recall* juga akan menjadi metrik model yang digunakan untuk memilih model terbaik ketika ada nilai yang tinggi terkait dengan *False Negative*.
4. *F-Measure* : Perbandingan rata-rata nilai *precision* dan *recall* yang dibobotkan

HASIL DAN PEMBAHASAN

Data selection

Hasil pada tahap ini yaitu berupa *dataset* yang digunakan untuk melakukan penelitian. Teknik yang digunakan yaitu *crawling* data dari Twitter API yang dilakukan pada tanggal 30 Juni dan data yang diambil merupakan data cuitan dengan rentang waktu pada tanggal 1 Juni 2022 sampai 30 Juni 2022 dengan mendapatkan data sebesar 1.000 tweet dapat dilihat pada Tabel 1 dibawah ini.

Tabel 1. Sample Hasil Data Selection

No	Hasil Tweet
1	@mojokdotco Jokowi tiga periode? Wkwkw
2	RT @temponewsroom: Kini nyaris semua partai berada di bawah kendalinya.\n\nUntuk apa? Mengembalikan ide gila tiga periode? Atau konsolidasi m...
3	Presiden Jokowi tolak jabatan tiga periode\n\n(' 🍀 `)\n #BersatuDukungRIMaju https://t.co/140WPGbdVu
4	RT @ChusnulCh__: Jokowi berkali2 menolak, yg berwacana jg sdh tdk ada lg, bahkan pemilu 2024

Preprocessing

Dilakukan tahapan pre processing untuk menghilangkan kata – kata atau karakter – karakter yang tidak dibutuhkan

1. Cleaning

Proses cleaning merupakan penghapusan URL, angka, atau simbol-simbol yang tidak diperlukan dalam klasifikasi. Berikut merupakan contoh hasil penerapan tahap cleaning, bisa dilihat pada Tabel 2 dibawah ini.

Tabel 2. Sample Hasil Cleaning

Twit sebelum dilakukan <i>cleaning</i>	Tweet setelah dilakukan <i>cleaning</i>
Partai Gelora: Parpol Pro Jokowi Munculkan Kembali Presiden Tiga Periode Dalam Bentuk Baru https://t.co/7RnjRvrqFw	Partai Gelora Parpol Pro Jokowi Munculkan Kembali Presiden Tiga Periode Dalam Bentuk Baru
Presiden Jokowi tolak jabatan tiga periode \n\n\ngl(' ➤ \`l)\n #BersatuDukungRIMaju https://t.co/140WPGbdVu	Presiden Jokowi tolak jabatan tiga periode

2. Case Folding

Di dalam tahap ini dilakukan mengubah semua huruf kapital dalam dataset menjadi huruf kecil. Pada proses ini dilakukan supaya semua data yang akan diproses memiliki penyeragaman karakter sehingga ketika proses preprocessing selanjutnya bisa lebih baik lagi dan menghasilkan nilai preprocessing yang bagus. Berikut contoh hasil tahap case folding, bisa dilihat pada tabel 3 dibawah ini.

Tabel 3. Sample Hasil Case Folding

Twit sebelum dilakukan <i>Case Folding</i>	Tweet setelah dilakukan <i>Case Folding</i>
Baru di era Jokowi ini ada Polisi kalah sama lonte yang model begini minta tiga periode	baru di era jokowi ini ada polisi kalah sama lonte yang model begini minta tiga periode
Partai Demokrat Curigai Reshuffle Menteri Hanya Strategi Jokowi Muluskan Jabatan Tiga Periode	partai demokrat curigai reshuffle menteri hanya strategi jokowi muluskan jabatan tiga periode

3. Tokenizing

Setelah data dibersihkan dan diubah ke huruf kecil, selanjutnya akan melalui tahap tokenizing. Tokenizing merupakan proses pemecahan kalimat pada data komentar menjadi kata perkata, proses ini akan mempermudah saat masuk ke tahap transformasi dengan tidak memproses kalimatnya tapi memproses kata demi kata dari kalimat tersebut. Hasil dari proses tokenizing dapat di lihat pada tabel 4 dibawah ini.

Tabel 4 Sample Hasil Tokenizing

Twit sebelum dilakukan <i>Tokenizing</i>	Tweet setelah dilakukan <i>Tokenizing</i>
Ajak Masyarakat Lawan Wacana Jokowi Tiga Periode Masinton Pasaribu Turun Penuhi Jalanan lewat Partai Demokrat Curigai Reshuffle Menteri Hanya Strategi Jokowi Muluskan Jabatan Tiga Periode	'ajak','masyarakat','lawan','wacana', 'jokowi','tiga','periode','masinton', 'pasaribu','turun','penuhi', 'jalanan','lewat' 'partai', 'demokrat', 'curigai', 'reshuffle', 'menteri', 'hanya', 'strategi', 'jokowi', 'muluskan', 'jabatan', 'tiga', 'periode'

4. Filtering

Selanjutnya akan dilakukan tahap stopword removal atau filtering, yakni menyaring kata-kata yang relevan guna proses klasifikasi selanjutnya. Contoh kata yang tidak diperlukan yaitu kata “yang”, “dan”, “di”, “dari”, “nya”. Pada proses ini akan menggunakan dokumen berisikan kata – kata stopword, yang digunakan untuk membantu menyaring kata – kata yang tidak digunakan pada proses klasifikasi. Hasil dari proses stopword removal yang dibantu menggunakan dokumen yang berisi kata – kata stopword, dapat dilihat pada tabel 5 dibawah ini.

Tabel 5. Sample Hasil Filtering

Twit sebelum dilakukan <i>Filtering</i>	Tweet setelah dilakukan <i>Filtering</i>
'elu', 'terus', 'yang', 'demo', 'banyak', 'amat', 'orderan', 'keknya', 'mmg', 'pak', 'harus', 'tiga', 'periode', 'biar', 'lu', 'dpat', 'orderan', 'terus'	'elu', 'demo', 'banyak', 'orderan', 'keknya', 'mmg', 'periode', 'biar', 'lu', 'dpat', 'orderan'
'omongan', 'bahlil', 'cuma', 'test', 'the', 'water', 'kalau', 'hasilnya', 'lanjut', 'periode'	'oke', 'omongan', 'bahlil', 'test', 'the', 'water', 'hasilnya', 'oke', 'periode'

5. *Stemming*

Proses atau tahap terakhir dari preprocessing ialah proses stemming, stemming bertujuan untuk mengubah kata-kata yang terdapat dalam dokumen ke dalam kata-kata dasar. Proses stemming dilakukan menggunakan package katadasaR. Berikut adalah hasil proses stemming di dalam dataset di tabel 6 dibawah ini .

Tabel 6. Sample Hasil Stemming

Twit sebelum dilakukan <i>Stemming</i>	Tweet setelah dilakukan <i>Stemming</i>
'elu', 'demo', 'banyak', 'orderan', 'keknya', 'mmg', 'periode', 'biar', 'lu', 'dpat', 'orderan'	'elu', 'demo', 'banyak', 'order', 'kek', 'mmg', 'periode', 'biar', 'lu', 'dpat', 'order'
'omongan', 'bahlil', 'test', 'the', 'water', 'hasilnya', 'oke', 'periode'	'omongan', 'bahlil', 'test', 'the', 'water', 'hasil', 'oke', 'periode'

Transformation Data

Tahap *transformation* atau transformasi merupakan tahap untuk memberikan bobot pada kata menggunakan pembobotan TF (*Term Frequency*) dari data komentar yang telah melalui proses pelabelan data dan pembobotan skor sentimen, di bawah ini merupakan hasil dari proses tranformastion data, dapat dilihat pada gambar 2.

Docs	dukung	jokowi	kendali	lantik	menteri	partai	periode	presiden	resmi	wakil
135	0	1	0	1	2	0	0	1	1	1
159	0	0	0	2	2	0	0	2	1	1
187	0	0	0	2	2	0	0	2	1	1
235	0	1	0	1	2	0	0	1	1	1
266	0	0	0	1	2	0	0	1	1	1
4	0	0	0	1	2	0	0	1	1	1
40	0	1	0	1	2	0	1	1	1	1
63	0	0	0	2	2	0	0	2	1	1
71	0	0	0	2	2	0	0	2	1	1
75	0	1	0	1	2	0	0	2	1	1

Gambar 2. Sample Hasil TF-IDF

Data Mining

Dalam penelitian ini digunakan proses pengolahan data menggunakan algoritma *naïve bayes* dengan 3 skenario yang terdiri dari 90:10, 80:20 dan 70:30. Hasil sentiment tentang opini perpanjangan masa jabatan Presiden Indonesia ini menggunakan *naïve bayes classifier* berupa *Confusion Matrix* dimana *Confusion Matrix* ini menghasilkan nilai akurasi tentang prediksi mesin terhadap data yang sudah diproses. Berikut adalah pembagian data training dan data testing pada tabel 7 berikut.

Tabel 7. Pembagian Data testing dan data Training

Presentase Data		Jumlah Data	
Training	Testing	Training	Testing
90%	10%	577	64
80%	20%	513	128
70%	30%	449	192
Total		641	

Evaluation

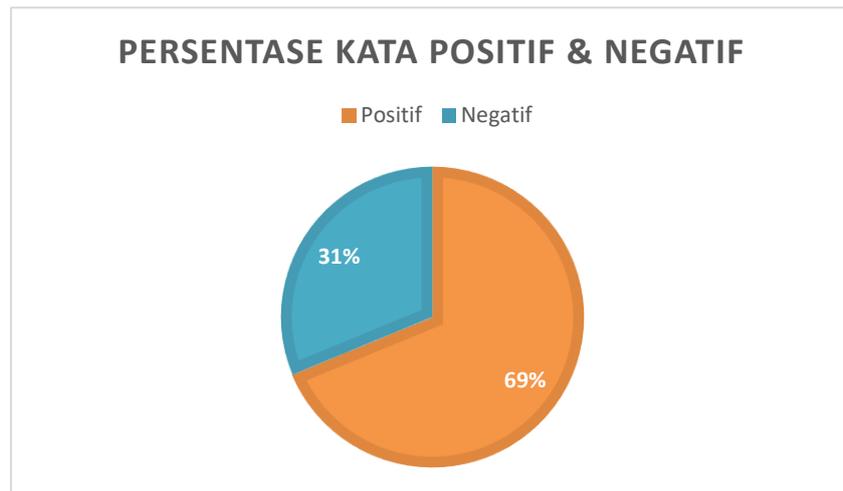
Tahap terakhir setelah klasifikasi yaitu evaluasi. Tahap ini merupakan tahap pengukuran performa terhadap model yang telah di implementasikan. Dengan menggunakan *confussion matrix*, nilai yang akan dijadikan parameter dalam penelitian ini yaitu *accuracy*, *precision*, *recall*, dan *f-measure*. Maka akan dilakukan evaluasi untuk mengetahui hasil akurasi yang didapatkan dari algoritma *naïve bayes* menggunakan skenario yang telah ditentukan. Berikut hasil perhitungan evaluasi dari setiap skenario, dapat dilihat pada tabel 8 berikut.

Tabel 8. Hasil Perhitungan Evaluasi Setiap Skenario

Skenario	Accuracy	Precision	Recall	F- Measure
90:10	95%	97%	95%	96%
80:20	93%	94%	96%	94%
70:30	94 %	94 %	98 %	96 %

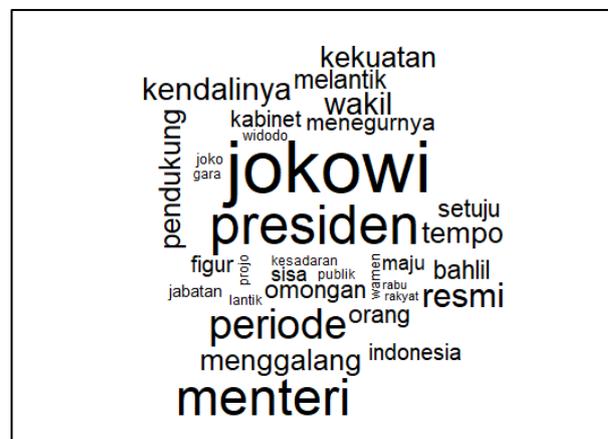
PEMBAHASAN

Penelitian ini menggunakan dataset dari twit dengan keyword "jokowi tiga periode" dengan menggunakan teknik *crawling* data pada twitter api. Tweet yang berhasil dikumpulkan sebanyak 1.000 tweet yang merupakan data tweet pada tanggal 1 Juni 2022 sampai 30 Juni 2022. Setelah melalui tahap pembersihan, data yang digunakan menjadi 641 data tweet menghasilkan 422 data *tweet* positif dan 192 data *tweet* negatif. Berikut persentase dari sentimen positif dan negatif pada gambar 3.



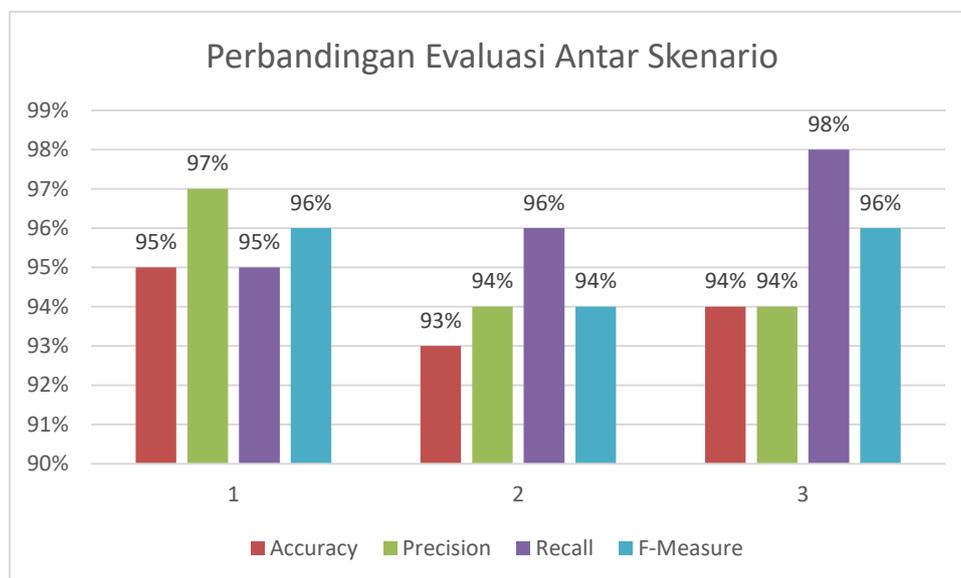
Gambar 3. Persentase Sentimen Posif dan Negatif

Word cloud yaitu kata yang sering muncul pada dataset. Kata yang sering muncul yaitu kata "jokowi", "presiden" dan "menteri". Semakin besar ukuran kata pada word cloud maka semakin tinggi frekuensi kata tersebut, artinya kata tersebut sering digunakan oleh masyarakat untuk mengirim cuitan pada twitter. Wordcloud dari penelitian ini dapat dilihat pada gambar 4 berikut.



Gambar 4. Wordcloud

Proses data mining menggunakan algoritma *naive bayes* dengan menerapkan 3 skenario yang dilakukan yaitu 90% data training dan 10% data testing, 80% data training dan 20% data testing, kemudian 70% data training dan 30% data testing. Perbandingan *performance* dari antar skenario, dapat dilihat pada gambar 5 berikut.



Gambar 5. Perbandingan Skenario

Berdasarkan perbandingan hasil pengujian evaluasi, skenario terbaik yang untuk model klasifikasi Naïve Bayes yaitu pada skenario atau model pertama (90% data training dan 10% data testing) dengan nilai *accuracy* sebesar 95% , nilai *precision* sebesar 97%, nilai *recall* sebesar 96%, dan nilai *f-measure* sebesar 96%. Model ini dinilai baik karena mempunyai persentase nilai terbesar pada 3 pengujian yaitu pada *accuracy*, *precision* dan *f-measure* dan juga pada nilai *recall* walau tidak mempunyai nilai terbesar pada skenario yang diujikan, tetapi persentasenya sudah sangat baik.

SIMPULAN

Penelitian ini melakukan analisis sentimen masyarakat di twitter terhadap usulan perpanjangan masa jabatan Presiden Indonesia dengan menggunakan algoritma *Naive Bayes*. Metodologi yang digunakan dalam penelitian ini yakni KDD (*Knowledge Discovery in Database*) yang terdiri dari 5 tahapan yakni tahap pertama yaitu *data selection* yang terdiri dari proses *crawling data*, lalu tahap *pre-processing* yang terdiri dari 5 proses didalamnya yakni *case folding*, *tokenizing*, *spelling normalization*, *filtering*, dan *stemming*, selanjutnya tahap *transformation* yaitu pembobotan kata/*term* dengan menggunakan TF-IDF, selanjutnya tahap *data mining* yaitu proses klasifikasi data dengan mengimplementasikan salah satu algoritma *naive bayes* yaitu *multinomial naive bayes*, dan tahap terakhir yaitu *evaluation* di mana pada proses ini dilakukan pengujian dari model dengan melihat nilai parameter pada *confussion matrix* yakni nilai akurasi, *precision*, *recall*, dan juga *f-measure/f1-score* nya.

Data yang dikumpulkan berjumlah 1.000 data, diproses melalui tahap *pre-processing* menjadi 641 data dengan 422 tweet positif dan 192 tweet negatif. Pengolahan untuk pelabelan data sentimen menggunakan kamus kata positif dan kata negatif, persentase skor sentimen yaitu

69% sentimen positif dan 31 % sentimen negatif, yang berarti sentimen masyarakat terhadap usulan masa perpanjangan masa jabatan Presiden Indonesia yang berjumlah 1.000 data tweet lebih banyak yang mempunyai sentimen positif dibanding sentimen negatif.

DAFTAR PUSTAKA

- Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, 48(1), 79–95.
- Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147.
- Badjrie, S. H., Pratiwi, O. N., & Anggana, H. D. (2021). Analisis Sentimen Review Customer Terhadap Produk Indihome Dan First Media Menggunakan Algoritma Convolutional Neural Network. *EProceedings of Engineering*, 8(5).
- Basit, A. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Hasil Panen Padi. *Jurnal Teknik Informatika Kaputama (JTik) 2020*, 4(2), 208–213.
- Manik, G., Ernawati, I., & Nurlaili, I. (2021). Analisis Sentimen Pada Review Pengguna E-Commerce Bidang Pangan Menggunakan Metode Support Vector Machine (Studi Kasus: Review Sayurbox dan Tanihub pada Google Play). *Senamika*, 2(2), 64–74.
- Marcos de Moraes, R., Soares, E. A. de M. G., & Machado, L. dos S. (2020). A double weighted fuzzy gamma naive bayes classifier. *Journal Of Intelligent & Fuzzy Systems*, 38(1), 577–588.
- Medistiara, Y. (2022). *Survei IPO: 61 Persen Responden Tak Setuju Jabatan Presiden 3*. Retrieved from https://news.detik.com/berita/d-5960343/survei-ipo-61-persen-responden-tak-setuju-jabatan-presiden-3-periode?_ga=2.70094928.1118145021.1649317763-1131121490.1628563534
- Nugroho, R. A., & Cholissodin, I. (2021). Implementasi Naive Bayes Classifier untuk Klasifikasi Emosi Tweet Berbahasa Indonesia pada Spark. 5(1), 301–310.
- Paulina, W., Bachtiar, F. A., & Rusydi, A. N. (2020). Analisis Sentimen Berbasis Aspek Ulasan Pelanggan Terhadap Kertanegara Premium Guest House Menggunakan Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(4), 1141–1149.
- Santoso, B., & Fadil Djailaini, M. (2022). Luhut Soal Jokowi Tiga Periode: Presiden Taat Konstitusi, Tapi... Retrieved from [suara.com website: https://www.suara.com/news/2022/03/13/104416/luhut-soal-jokowi-tiga-periode-presiden-taat-konstitusi-tapi](https://www.suara.com/news/2022/03/13/104416/luhut-soal-jokowi-tiga-periode-presiden-taat-konstitusi-tapi)
- Taofik Krisdiyanto, E. M. O. N. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naive Bayes Clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(1), 32–37. Retrieved from <http://ejournal.uin-suska.ac.id/index.php/coreit/article/view/12945>
- Twitter. (2019). *Twitter, Inc. Annual Report - Our logo? Made up of 3 sizes of circles, each representing local, topical, and global conversation!*