

Analisis Sentimen Pencalonan Jokowi dalam Tiga Periode menggunakan Metode Naïve Bayes

Sentiment Analysis of Jokowi's Candidate Discourse in Three Periods using the Naïve Bayes Method

¹Dita Nurmadewi*, ²Mia Amaliah, ³Hajar Hanifah, ⁴Ulfa Bunnaya Purwanti, ⁵Maruti Sekar Arum, ⁶Nanang Wijaya Kusuma, ⁷Sri Yuniati

¹²³⁴⁵⁶Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie

⁷Hubungan Internasional, Universitas Jember

Jalan H. R. Rasuna Said Kav. C22, Kecamatan Kuningan, DKI Jakarta 12920

Jalan Kalimantan 37, Jember 68121

*e-mail: dita.nurmadewi@bakrie.ac.id

(received: 8 November 2022, revised: 16 November 2022, accepted: 26 November 2022)

Abstrak

Menjelang pemilihan presiden 2024, banyak wacana berkembang di kalangan masyarakat terkait periode pemilihan presiden yang berlangsung setiap lima tahun. Salah satu wacana politik yang ramai diperbincangkan adalah pencalonan Presiden Jokowi untuk periode ketiga kepemimpinannya. Penelitian ini bertujuan untuk melakukan analisis sentimen masyarakat terhadap wacana tersebut. Data sampel untuk analisis sentimen diambil dari media sosial *Twitter* berupa status *tweet*. Tahapan yang dilakukan oleh peneliti dimulai dengan menentukan metode penelitian melalui *text mining* dan analisis sentimen dengan menggunakan algoritma *Naïve Bayes*. Hasil dari penelitian yang dilakukan mengungkapkan bahwa masih banyaknya sentimen positif dibandingkan sentimen negatif terkait wacana pencalonan Jokowi dalam tiga periode.

Kata kunci: Analisis sentimen, *Naive Bayes*, *Twitter*, Jokowi

Abstract

Approaching the 2024 presidential election many discourses develop among the public regarding the presidential election period which takes place every five years. One of the most widely discussed political discourses is the candidacy of President Jokowi for the third term of his leadership. This study aims to carry out a sentiment analysis of the public on the discourse. The sample data for sentiment analysis is taken from Twitter social media in the form of tweet status. The stages carried out by the researchers began with determining the research method through text mining and sentiment analysis using the Naive Bayes algorithm. The results of the research conducted revealed that there are still many positive sentiments compared to negative sentiments related to the discourse of Jokowi's candidacy in the three periods.

Keywords: *Sentiment Analysis, Naive Bayes, Twitter, Jokowi*

1 Pendahuluan

Teknologi informasi dan komunikasi saat ini berkembang kian pesat, karena adanya dukungan pengguna internet di dunia. Dalam melakukan aktifitas sehari-hari, masyarakat cenderung memanfaatkan internet untuk saling berkomunikasi melalui media sosial, mencari informasi terbaru atau *up-to-date* dan bahkan penggunaan internet juga dijadikan sebagai media pendukung dalam melakukan transaksi jual beli [1]. Penggunaan internet untuk mencari informasi menjadi rutinitas sehari-hari di kalangan masyarakat, hal ini di dukung oleh survey yang dilakukan pada tahun 2016 oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) yang menyatakan bahwa jaringan internet sebesar 97,4% digunakan untuk mengakses media sosial [2].

Media sosial memiliki peran sebagai alat komunikasi dan sudah menjadi bagian dari teknologi informasi dan komunikasi yang juga memberikan informasi terbaru mengenai kondisi yang sedang terjadi [3]. *Twitter* adalah salah satu media sosial yang mendapat perhatian tinggi dari masyarakat yang dapat dijadikan sebagai media alternatif untuk menyebarkan informasi hingga ke berbagai negara [4]. Penggunaan *twitter* sebagai salah satu media sosial juga di dukung oleh dataindonesia.id dalam laporan *we are social*, yang menyatakan bahwa Indonesia merupakan salah satu negara pengguna media sosial utamanya yaitu penggunaan Twitter terbesar di dunia, dimana pengguna *Twitter* sebanyak 18,45 juta pada tahun 2022 [5]. Dalam penggunaan media sosial seperti *twitter* tiap masyarakat bebas untuk saling berpendapat dan bertukar informasi. Kebebasan berpendapat juga sudah diatur dalam Undang-Undang Dasar 1945 (UUD 45) dalam pasal 28 ayat 3 yang menyatakan jika setiap orang berhak atas kebebasan berserikat, berkumpul, dan mengeluarkan pendapat [6]. Adanya kebebasan berpendapat ini memiliki dampak positif dan negatif bagi pengguna *twitter*. Dampak positif dari adanya media sosial sebagai media informasi yaitu informasi penting yang ingin disampaikan oleh pemerintah kepada masyarakat dapat disebarluaskan secara cepat dengan *twitter*. Akan tetapi *twitter* juga memiliki dampak negatif, yang mana terkadang beberapa masyarakat kurang memahami norma-norma yang harus dilakukan dalam memberikan pendapatnya dalam media sosial, postingan di *twitter* atau yang biasa disebut *tweet* juga dapat menimbulkan perselisihan antar akun pengguna *twitter*. Pada umumnya *tweet* digunakan untuk memberikan informasi atau melakukan postingan tentang diri pengguna *twitter* dan berbagi informasi pribadi. Isi dari *tweet* juga dapat menggambarkan perasaan dari pengguna *twitter* tersebut [7].

Saat ini sedang ramai diperbincangkan mengenai pemilihan presiden baru yang akan digelar pada tahun 2024 di *twitter*. Opini yang berkembang utamanya menyangkut pencalonan kembali Joko Widodo (Jokowi) yang menjabat sebagai Presiden Republik Indonesia untuk kembali mencalonkan diri sebagai presiden di periode selanjutnya. Pada media sosial *twitter* sedang ramai *tweet* yang berbicara tentang pencalonan Jokowi dalam tiga periode. *Tweet* yang berisi opini masyarakat ini yang nantinya akan menjadi sumber data untuk melihat analisis sentimen yang muncul di kalangan masyarakat. Analisis sentimen ini nantinya dapat menjadi informasi bagaimana masyarakat memberikan pendapatnya mengenai Jokowi tiga periode. Metode yang digunakan untuk analisis sentimen masyarakat yaitu *naïve bayes* yang merupakan salah satu metode pengklasifikasian statistik dalam suatu algoritma *machine learning*.

2 Tinjauan Penelitian

Penelitian terdahulu yang telah dilakukan tentang analisis sentimen masyarakat di media sosial *twitter* dapat dilihat pada Tabel 1. Penelitian terdahulu ini dijadikan sebagai literatur pendukung atau referensi dalam penelitian ini.

Tabel 1. Penelitian Terdahulu

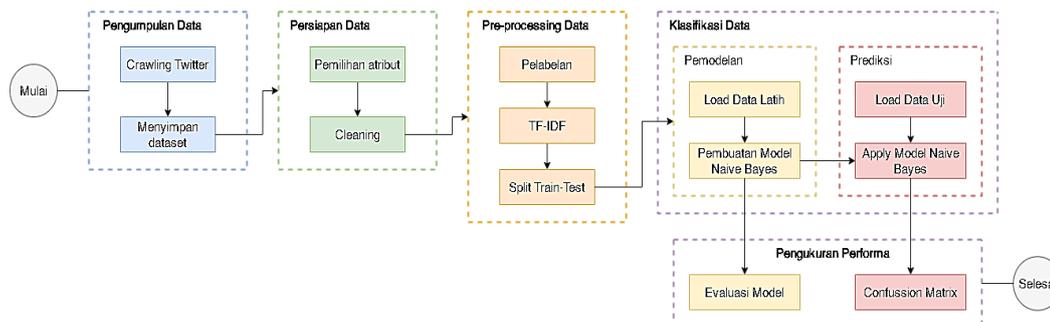
Nama Peneliti	Tahun	Hasil Penelitian Terdahulu
Yuni Nurtikasari, Syariful Alam, Teguh Iman Hermanto	2022	Penelitian yuni, dkk [8] membahas mengenai analisis sentimen film menggunakan tools orange. Perbedaan : Penelitian ini juga menganalisis sentimen masyarakat tetapi menggunakan tools rapidminer dan topik yang dibahas yaitu pencalonan kembali tokoh publik.
Muhammad Fahreza Alfa Sina Mustofa, Ahmad R. Pratama	2022	Penelitian mustofa dan pratama [9] melakukan analisis sentimen wabah covid-19 di postingan facebook yang dibuat oleh akun pemerintah dan portal berita dengan CrowdTangle. Perbedaan : Penelitian ini menggunakan data <i>twitter</i> yaitu berupa <i>tweet</i> yang digunakan sebagai sumber untuk menganalisis sentimen masyarakat. <i>Twitter</i> dipilih karena API <i>twitter</i> yang dapat diakses dengan mudah.

Alvionita Anjani	2022	Paper dari alvionita [10] menganalisis pro dan kontra di Internet mengenai kaum LGBT. Perbedaan: Penelitian ini juga melihat sentimen masyarakat terkait pro dan kontra, tetapi topik yang diangkat mengenai pencalonan presiden Jokowi tiga periode.
Cahyo Prianto, Nisa Hanum Harani, Indra Firmansyah	2019	Penelitian cahyo, dkk [11] melakukan analisis sentimen dari media sosial twitter terhadap kandidat Presiden RI pada pemilu 2019, tujuannya untuk melihat respon masyarakat terhadap calon presiden yaitu Joko Widodo dan Prabowo Subianto mana yang lebih unggul. Perbedaan : Penelitian ini hanya fokus pada analisis sentimen kepada tokoh publik Jokowi untuk pencalonan presiden tiga periode pada pemilu 2024.

Berdasarkan penelitian terdahulu diatas, banyak penelitian mengenai analisis sentimen masyarakat yang di ambil dari media sosial seperti, *twitter* dan *facebook*. *Tools* yang digunakan juga beraneka macam. Penelitian ini mengangkat tema yang saat ini sedang trend dibicarakan di masyarakat, mengingat akan adanya pemilu di tahun 2024. Opini masyarakat yang akan di analisis mengenai Jokowi tiga periode yang mana ini didasarkan dari *tweet*. Data ini sebelum diolah akan dilakukan *pre processing* yang dilanjutkan dengan melakukan pengklasifikasian menggunakan metode *naive bayes* hingga menghasilkan nilai *performance*.

3 Metode Penelitian

Metode penelitian yang dilakukan, digambarkan menggunakan *flowchart* yang ditampilkan pada gambar 1 sebagai berikut ini :



Gambar 1. Flowchart Tahapan Penelitian

Proses yang ditampilkan pada gambar 1 meliputi beberapa kegiatan, yang pertama yaitu pengambilan data atau *crawling* data di *twitter* dengan aplikasi *rapidminer*. *Crawling* merupakan salah satu teknik dalam pengumpulan data. Dilanjutkan dengan kegiatan persiapan data. Kegiatan ini bertujuan agar data yang dihasilkan bersih dan valid sehingga mudah dibaca. Setelah itu dilakukan analisis menggunakan metode klasifikasi *Naive Bayes* untuk melihat akurasi dari data yang dihasilkan dan selanjutnya mengukur *performance* yang menghasilkan hasil analisis berupa informasi yang baru.

3.1 Pengumpulan Data

Teknik pengumpulan data dimulai dengan mengkoneksikan *rapidminder* dengan *Twitter API* (*Application Programming Interface*). Dengan adanya koneksi ini, akan memudahkan untuk dapat mengambil data dari *twitter* yang selanjutnya data dapat diolah. Tahapan pengumpulan data *tweet* ditampilkan pada Gambar 2, seperti di bawah ini:



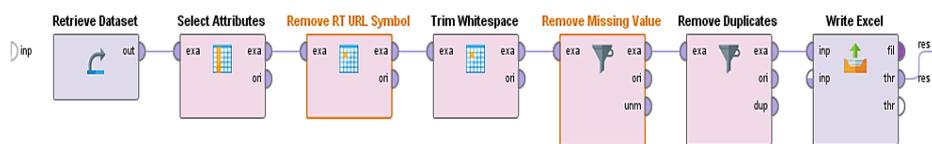
Gambar 2. Operator Proses Pengumpulan Data Twitter

Pada tahapan pengumpulan data yang ditunjukkan oleh gambar 2 menggunakan beberapa operator yang ada pada aplikasi *twitter*, yaitu :

1. *Retrieve twitter.com* : operator yang digunakan untuk mengakses informasi yang tersimpan di repositori *twitter*.
2. *Search Twitter* : operator yang digunakan untuk mengakses atau mengambil data di *twitter*.
3. *Write Excel* : operator yang digunakan untuk menyimpan data dalam bentuk *excel*.

3.2 Persiapan Data

Kegiatan yang dilakukan setelah pengumpulan data yaitu persiapan data .Kegiatan ini dilakukan untuk mempersiapkan agar data menjadi data yang bersih sehingga dapat dilakukan analisis data menggunakan metode klasifikasi data. Tahapan yang dilakukan pada persiapan data yang ditunjukkan sebagaimana ditampilkan Gambar 3, yaitu :



Gambar 3. Persiapan Data

3.2.1 Pemilihan Atribut

Dalam melakukan analisis sentimen, peneliti membutuhkan data status atau *tweet* dari pengguna Twitter untuk mengetahui bagaimana opini masyarakat mengenai topik Jokowi tiga periode. Oleh karena itu, langkah pertama yang akan dilakukan dalam mempersiapkan data adalah pemilihan atribut. Dalam hal ini, atribut *text* yang berisi data tersebut lah yang akan diambil menggunakan operator *select attributes* untuk selanjutnya diproses di tahapan selanjutnya.

Tabel 2. Dataset setelah proses pemilihan atribut

No.	Text
1	RT @NegeriKolam: A1 Jokowi akan melakukan show of force dengan mengundang massa di akhir bulan November nanti. Peserta akan didatangkan dar...
2	RT @Ari3Pras: Isu 3 periode GAGAL, Isu ijazah palsu KANDAS. Gorenganpun diganti dg mengadu domba Pak @jokowi @ganjarpranowo dengan Bu Mega...
3	@DokterTifa anda n kelompok anda slalu Dengungkan 3 periode, omongmu adalah doa. Dijabanin jokowi ngamuk
4	RT @ocongBoyy: pak Jokowi itu gak minat 3 periode krn visinya Internasional, bukan lagi nasional...ambisinya sudah tinggi, jadi Sekjen PBB...
...	
2465	@OposisiCerdas Jokowi marah ke SP / Nasdem bukan krn faktor anies nya...tp faktor bahwa itu menutup kemungkinan 3 periode.. Tq @NasDem pokoknya !!

Terlihat dari Tabel 2 di atas bahwa data yang didapatkan pada atribut *text*, merupakan kalimat dalam bentuk mentah yang artinya masih terdapat beberapa *noise* atau kesalahan acak atau kesalahan

varian dalam variabel terukur yang terdapat dalam sebuah kalimat. Proses ini menghasilkan banyak data *text* yang selanjutnya dapat diolah ke proses selanjutnya, yaitu *cleaning*.

3.2.2 Pembersihan Data (*Data Cleaning*)

Pembersihan data atau *data cleaning* merupakan proses menghapus simbol atau karakter yang tidak dibutuhkan dalam data yang akan diolah sehingga tidak akan mengganggu dalam proses klasifikasi data. Dataset yang andal, didapatkan ketika pertama kali menggunakan proses *cleansing* atau pembersihan data [12]. Oleh karena itu, dalam proses data *cleaning* beberapa kata atau karakter serta simbol yang bukan merupakan bagian dari dalam kalimat akan dihapuskan, termasuk : *RT*, *@mention*, *#hashtag*, *URL*, spasi, simbol, dan lainnya. Adapun kegiatan dalam proses pembersihan data meliputi:

- a. Penghapusan simbol atau karakter
Kata atau simbol yang tidak termasuk dalam bahasa alami (*natural language*) seperti *RT*(*Re-Tweet*), *@mention*, *#hashtag*, *URL*, maupun simbol atau karakter lainnya harus dihilangkan karena bukan bagian dari kalimat yang mana hendak dianalisis sentimanya.
- b. Menghilangkan spasi
Setelah dari penghapusan simbol diteruskan ke menghilangkan spasi yang berlebihan. Spasi atau daerah kosong yang dimaksud adalah spasi di awal dan di akhir kalimat. Hal ini bertujuan agar memudahkan untuk pembacaan kalimat.
- c. Menghapus data kosong
Diperlukan juga penghapusan data kosong atau *missing values*. Baris yang berisi data kosong harus dihilangkan karena tidak akan digunakan dalam proses analisa ini. Hal ini dikarenakan data kosong dapat bertindak sebagai *outlier* yang akan mempengaruhi keakurasian dari analisis.
- d. Menghilangkan duplikasi data
Proses akhir dari tahap pembersihan data adalah menghilangkan kemungkinan adanya duplikasi data pada dataset yang dapat mempengaruhi tingkat akurasi hasil analisis sentimen.
- e. Simpan hasil
Data akhir dari hasil pembersihan data, kemudian akan disimpan dalam bentuk excel.

Berikut ini adalah hasil perbandingan atribut *text* setelah mengamalami proses pembersihan data, yang ditunjukkan dengan Tabel 3 sebagai berikut :

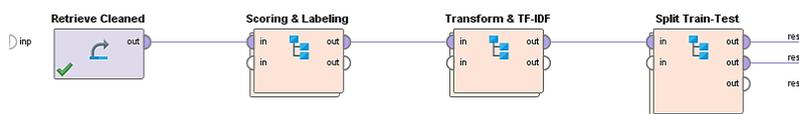
Tabel 3. Perbandingan atribut *text* sebelum dan sesudah proses pembersihan data

Dataset sebelum <i>Cleaning</i>	Dataset setelah <i>Cleaning</i>
RT @NegeriKolam: A1 Jokowi akan melakukan show of force dengan mengundang massa di akhir bulan November nanti. Peserta akan didatangkan dar...	A1 Jokowi akan melakukan show of force dengan mengundang massa di akhir bulan November nanti Peserta akan didatangkan dar...

Berdasarkan tabel 3 diatas, dapat diketahui perbandingan dataset sebelum *cleaning* dan setelah *cleaning* bahwa *text* lebih mudah dibaca ketika sudah mengalami proses *data cleaning*. Hal ini juga memudahkan atribu *text* untuk diolah ke tahap *pre-processing*.

3.3 Pre-processing Data

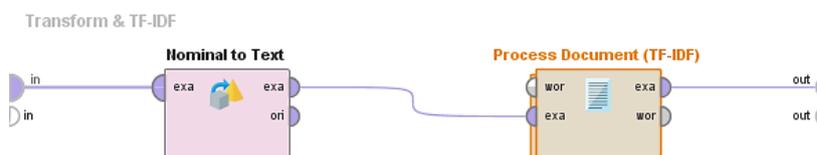
Tahapan selanjutnya adalah *pre-procesing* data, yang bertujuan untuk memproses data untuk memastikan data memiliki struktur maupun tipe-data yang sesuai dan siap untuk dianalisa. Tahapan ini ditampilkan pada gambar 4 yang meliputi proses pelabelan (*labeling*), TF-IDF (yang mencakup *tokenize*, *transform*, *stopword*, dan *steming*), dan pembagian data latih (*train*) dan data uji (*test*). Berikut adalah operator yang digunakan pada tahap *pre processing*, yang ditampilkan pada Gambar 4 berikut ini:



Gambar 4. Operator Tahapan *Pre-processing* Data

3.3.1 Pelabelan Data

Langkah pertama dalam tahapan *pre-processing* adalah proses *data labeling* atau pelabelan data. Pada proses ini, 100 *record* data pertama akan diberikan skor menggunakan operator *Dictionary-Based Sentiment*; dan berdasarkan skor tersebut, dengan menggunakan operator *Generate Attributes* untuk membuat atribut *sentiment* yang menyimpan nilai positif atau negatif. Selanjutnya penetapan atribut *sentiment* tersebut sebagai label atau target prediksi. Berikut adalah operator yang digunakan pada proses *transform* data yang ditampilkan pada Gambar 5, yaitu :

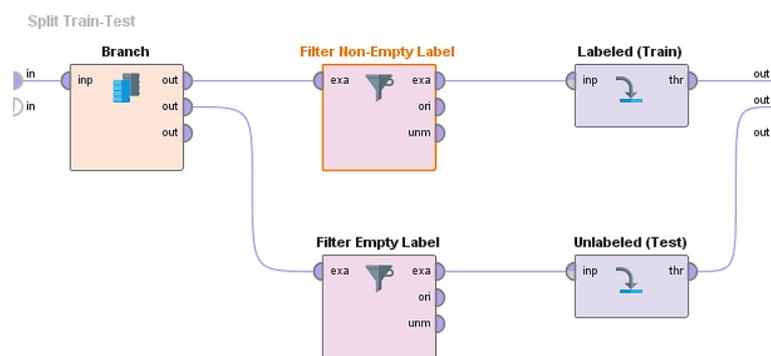


Gambar 5. Operator Proses *Transform* Data

Dari hasil pelabelan data yang ditunjukkan oleh gambar 5, bahwa proses di awal menggunakan 100 *record* data menghasilkan hasil analisis sentimen *tweet* yang menunjukkan bahwa sentimen positif sebanyak 53 *tweet*, yang diikuti oleh sentimen negatif sebanyak 47. Proses ini dilanjutkan ke tahapan pembagian data untuk diteruskan ke proses klasifikasi data.

3.3.2 Pembagian Data

Proses terakhir dalam tahapan *pre-processing* data ini adalah pembagian data menjadi data latih (*train*) dan data uji (*test*) yang ditunjukkan oleh Gambar 6. Data latih akan digunakan untuk melatih model yang akan dibuat, sedangkan data uji ditujukan sebagai dataset yang digunakan dalam penerapan model yang telah dibuat dalam melakukan klasifikasi atau prediksi. Berikut adalah operator yang digunakan pada proses pembagian data yang ditampilkan pada Gambar 6, yaitu :



Gambar 6. Operator Proses Pembagian Data

3.4 Klasifikasi Data

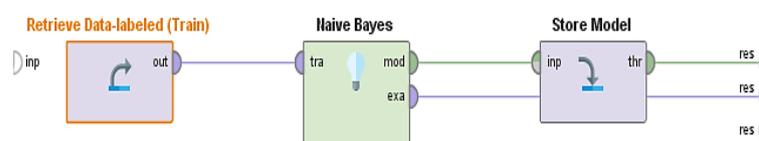
Metode atau teknik ini dicetuskan pertama kali pada 1770 oleh Thomas Bayes [13]. Metode ini bekerja dengan baik pada kasus atau situasi pada data yang sifatnya kompleks. Metode ini memiliki konsep probabilitas bersyarat yang dinotasikan pada persamaan 1.

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (1)$$

Notasi persamaan ini memiliki arti dimana X merupakan data dengan class yang belum diketahui. Sedangkan H merupakan hipotesis data pada suatu class yang spesifik. P(H|X) menggambarkan hipotesis H yang berdasar pada kondisi X begitupun sebaliknya pada notasi P(X|H) menggambarkan probabilitas X yang berdasarkan kondisi pada hipotesisi H. P(H) artinya probabilitas hipotesis H (prior probabilitas) dan P(X) artinya probabilitas X. Teknik klasifikasi ini digunakan untuk penentuan dari kelas sentimen yang akan diterapkan pada data *tweet* yang digunakan.

3.4.1 Pembuatan Model

Naïve Bayes bertujuan untuk membuat model klasifikasi atau prediksi yang sesuai. Proses ini ditampilkan pada gambar 7 yang menunjukkan proses pemodelan menggunakan *Naïve Bayes*. Dengan menggunakan data *training* yang dihasilkan sebelumnya untuk melatih algoritma *Naïve Bayes*, maka terbentuk model *Classifier* yang disimpan dan digunakan kemudian pada data *testing*. Proses ini ditampilkan pada Gambar 7 berikut ini :

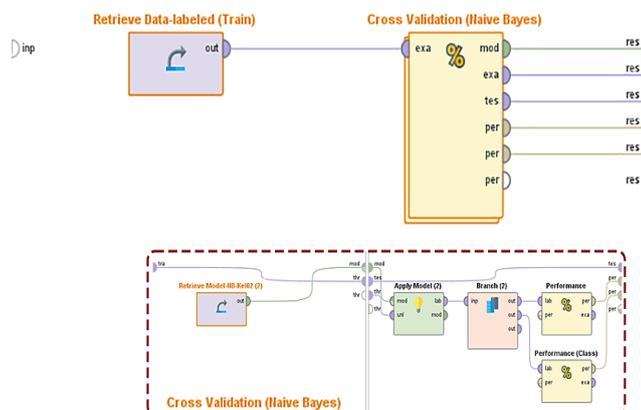


Gambar 7. Pemodelan *Naïve Bayes Classifier*

Data *training* yang berjumlah 100 *record* data dan data testing berjumlah 243 *record* data, kemudian akan di uji kembali melalui proses evaluasi model, ini untuk menghasilkan analisis data keseluruhan yang akurat.

3.4.2 Evaluasi Model

Setelah mendapatkan model klasifikasi atau prediksi *Naïve Bayes Classifier*, perlu dilakukan proses evaluasi terhadap model tersebut sebelum melanjutkan ke penerapan model pada data *testing* yang ditampilkan pada gambar 8. Evaluasi model dilakukan untuk mengetahui tingkat akurasi, presisi, maupun tingkat kesalahan klasifikasi dari model ini. Untuk melakukan evaluasi model, kami menggunakan operator *Cross Validation* (yang di dalamnya mencakup operator *Retrieve Model*, *Apply Model*, dan *Performa*) pada data latih. *Cross validation* digunakan untuk mengevaluasi kinerja algoritma dengan melakukan pemecahan data menjadi data latih dan data uji. Proses ini ditampilkan pada Gambar 8 berikut ini :



Gambar 8. Evaluasi Model menggunakan *Naïve Bayes Classifier*

3.5 Pengukuran Performa

Pengukuran performa pada penelitian ini menggunakan *Confusion Matrix* yang digunakan untuk melihat tingkat keberhasilan suatu algoritma yang diterapkan pada sebuah model. *Confusion matrix* dilakukan untuk mencari beberapa nilai yang digunakan, seperti nilai *precision*, *recall*, *accuracy* dan *F-Measure*. Perhitungannya dilakukan dengan menghitung nilai dari *true positif* (TP), *false positif* (FP), *true negative* (TN), dan *false negative* (FN). Notasi dijabarkan dalam persamaan 2-5.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+TN} \quad (4)$$

$$F\ Measure = \frac{2*recall*precision}{recall+precision} \quad (5)$$

Notasi persamaan ini memiliki arti yaitu TP yang menggambarkan jumlah data yang terdeteksi benar, sementara TN yang menggambarkan jumlah data negative yang terdeteksi benar. Sedangkan FP artinya jumlah data positif yang terdeteksi salah dan FN yang artinya jumlah data negative yang terdeteksi salah [14].

4 Hasil dan Pembahasan

Penelitian ini menggunakan data dari sosial media *twitter* berupa *tweet*. Pengumpulan data *tweet* diambil, dengan limit 5000 data opini dengan kata kunci atau *query* yang digunakan yaitu “Jokowi 3 Periode”. Adapun total perolehan *tweet* yang dihasilkan dari teknik *crawling* data sebanyak 2465 *sample* data yang ditampilkan pada tabel 4. Data yang sudah diambil, disimpan di dalam *dataset* ke dalam bentuk *excel* menggunakan operator *write excel* sebagai *dataset* yang berikutnya akan dilanjutkan ke proses pembersihan dan pemilihan data. Berikut adalah *dataset* hasil *crawling* yang ditampilkan pada Tabel 4 berikut ini :

Tabel 4. Dataset Hasil Crawling Twitter

No.	Created-At	From-User-Id	To-User-Id	Language	Text	Id
1.	2022-11-02 06:56:54	2974828741	-1	in	RT @NegeriKolam: A1 Jokowi akan melakukan show of force dengan mengundang massa di akhir bulan November nanti. Peserta akan didatangkan dar...	1587594563 241541632
2.	2022-11-02 06:56:33	948535842657 599488	-1	in	RT @Ari3Pras: Isu 3 periode GAGAL, Isu ijazah palsu KANDAS. Gorenganpun diganti dg mengadu domba Pak @jokowi @ganjarpranowo dengan Bu Mega...	1587594474 380722177
3.	2022-11-02 06:56:00	157927388301 4041600	15028223669 46402304	in	@DokterTifa anda n kelompok anda slalu Dengungkan 3 periode, omongmu adalah doa. Dijabanin jokowi ngamuk	1587594334 463225857
4.	2022-11-02 06:02:51	167664742	-1	in	RT @ocongBoyy: pak Jokowi itu gak minat 3 periode krn visinya Internasional, bukan lagi nasional...ambisinya sudah tinggi, jadi Sekjen PBB...	1587580962 694893568
...						
2466	2022-10-24 20:07:26	151693308651 6477952	88774358757 9944960	in	@OposisiCerdas Jokowi marah ke SP / Nasdem bukan krn faktor anies nya...tp faktor bahwa itu menutup kemungkinan 3 periode.. Tq @NasDem pokoknya !!	1584532017 462792192

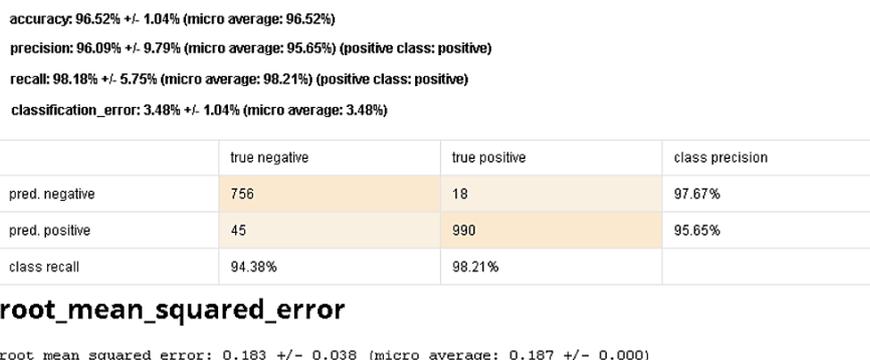
Selanjutnya dari hasil *crawling* atau pengumpulan data, melalui langkah-langkah yang dicantumkan pada metode penelitian yaitu persiapan data dan *pre-processing* data, maka dihasilkan data sebanyak 343 data yang terekam dan siap diolah. Data akhir ditunjukkan pada Tabel 5 berikut ini :

Tabel 5. Dataset Final

No.	Text
1	A1 Jokowi akan melakukan show of force dengan mengundang massa di akhir bulan November nanti Peserta akan didatangkan dar...
2	Isu 3 periode GAGAL Isu ijazah palsu KANDAS Gorenganpun diganti dg mengadu domba Pak dengan Bu Mega...
3	anda n kelompok anda slalu Dengungkan 3 periode omongmu adalah doa Dijabanin jokowi ngamuk
4	pak Jokowi itu gak minat 3 periode krn visinya Internasional bukan lagi nasionalambisinya sudah tinggi jadi Sekjen PBB...
...	
344	Jokowi marah ke SP / Nasdem bukan krn faktor anies nyatp faktor bahwa itu menutup kemungkinan 3 periode Tq pokoknya

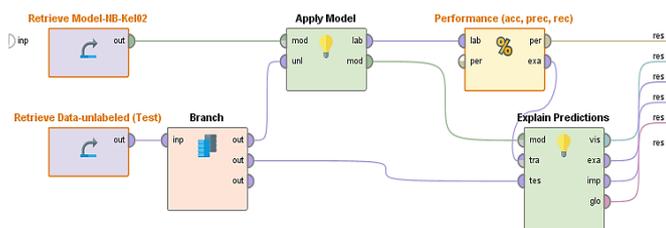
Setelah persiapan data selesai dan atribut *text* dalam dataset sudah berisi kalimat yang sesuai dengan bahasa alami (*natural language*), maka kami melanjutkan ke proses berikutnya, dimana proses ini menghasilkan dua dataset yaitu data latih atau *data training*, yang berisikan 100 *record* data dengan atribut *sentiment* yang telah terisi melalui proses *scoring & labeling* pada tahap *pre-processing* sebelumnya, dan data uji, yang berisikan 243 *record* data. Data uji ini menghasilkan sentimen negatif sebanyak 101 data (44%) dan sentimen positif sebanyak 142 (56%).

Selanjutnya data *training* digunakan untuk melatih algoritma *naïve bayes*, maka terbentuk model *naïve bayes classifier* yang akan digunakan pada data uji untuk mendapatkan model akhir. Namun sebelumnya harus dilakukan evaluasi terhadap model untuk mengetahui tingkat akurasi, presisi, maupun tingkat kesalahan klasifikasi dari model ini. Dan hasil dari evaluasi tersebut dapat dilihat pada Gambar 9 berikut ini :



Gambar 9. Metriks Hasil Evaluasi Model

Dari evaluasi model yang ditunjukkan pada gambar 9, maka didapatkan hasil yang cukup menjanjikan dengan skor akurasi dari model sebesar 96.52% dan skor presisi sebesar 96.09% serta skor *recall* sebesar 98.18%. Selain itu didapatkan juga skor tingkat kesalahan klasifikasi (*Classification Error Rate*) sebesar 3.48% dan skor RMSE (*Root Mean Square Error*) sebesar 0.183 (18%), yang terbilang relatif sangat kecil. RSME digunakan untuk mengukur kesalahan yang dapat dibuat oleh model dalam menerjemahkan hasil, yang dapat dilihat pada Gambar 10. Dengan demikian, model sudah memadai untuk digunakan pada tahap selanjutnya yaitu penerapan model pada data uji yang prosesnya ditampilkan pada Gambar 10 berikut ini:



Gambar 10. Penerapan Model pada Data Uji

Ketika model akhir di atas dilakukan proses *running*, maka muncul hasil dari pengukuran performa yang ditampilkan pada Gambar 11 berikut ini :

Naive Bayes - Performance

Criterion	Value	Standard Deviation
Accuracy	63.2%	± 7.1%
Classification Error	36.8%	± 7.1%
AUC	62.7%	± 18.2%
Precision	67.5%	± 11.8%
Recall	73.4%	± 10.2%
F Measure	69.4%	± 6.4%
Sensitivity	73.4%	± 10.2%
Specificity	49.1%	± 22.1%

Confusion Matrix

	true negative	true positive	class precision
pred. negative	12	9	57.14%
pred. positive	12	24	66.67%
class recall	50.00%	72.73%	

Gambar 11. Hasil Akhir Performance

Hasil penerapan algoritma *naïve bayes* terhadap data yang didapatkan seperti yang terlihat dari gambar 11, menunjukkan akurasi uji data sebesar 63.2% yang menggambarkan persentase keakuratan model klasifikasi yang dihasilkan. Dilanjutkan dengan nilai presisi yang menggambarkan akurasi data dengan hasil prediksi yang dihasilkan oleh model sebesar 67.5%, dan *recall* yang menggambarkan keberhasilan model dalam menghasilkan sebuah informasi, dengan presentase sebesar 73.4% (gambar 12). Mempertimbangkan jumlah nilai positif yang didapatkan dari model yang digunakan cenderung jauh lebih banyak dibandingkan dengan yang bersifat negatif, maka target acuan yang digunakan dalam evaluasi performa model adalah *true* positif. Oleh karena itu, yang menjadi poin penting sebagai acuan dalam menilai performa model adalah nilai *f1-score* (*f-measure*). Dalam hal ini, nilai *f1-score* menggambarkan perbandingan dari rata-rata presisi dan *recall* yang didapatkan adalah sebesar 69.4%. Hal ini menunjukkan metode klasifikasi dengan algoritma *naïve bayes* menunjukkan *performance* yang bagus diatas 50% yang artinya algoritma *naïve bayes* cukup berhasil untuk menilai kategori sentimen mengenai Jokowi dalam tiga periode.

5 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dengan metode klasifikasi menggunakan algoritma *naïve bayes* untuk analisis sentimen dengan pengolahan data *tweet* dari *twitter* tentang pencalonan kembali Jokowi dalam tiga periode, menunjukkan total hasil sentimen positif masyarakat sebesar 195 data yang menunjukkan dukungan kepada Jokowi untuk tiga periode dan sentimen negatif sebesar 148 dari 343 data akhir yang terekam. Darisini dapat disimpulkan jika sentimen positif lebih tinggi, jika dibandingkan sentimen negatif. Hal ini bisa diasumsikan bahwa masih banyaknya masyarakat yang menginginkan dan mempercayai Jokowi untuk terus melanjutkan kepemimpinannya menjadi tiga periode. Namun, keinginan masyarakat dari hasil penelitian ini bertentangan dengan Undang - Undang Dasar 1945 yang berlaku di Indonesia, dimana pada pasal 7

yang ditetapkan pada tanggal 19 Oktober 1999, yang berbunyi bahwa Presiden dan Wakil Presiden Indonesia memegang jabatan selama lima tahun, dan sesudahnya dapat dipilih kembali dalam jabatan yang sama, hanya dalam satu kali masa jabatan saja. Informasi ini dapat digunakan acuan bahwa kepemimpinan dari Jokowi mendapat banyak apresiasi masyarakat Indonesia yang digambarkan oleh hasil positif yang didapat dari analisis sentimen menggunakan *tools* rapidminer dan metode *Naïve Bayes*.

Referensi

- [1] K. Zuhri dan N. A. Saputri, "Analisis Sentimen Masyarakat terhadap Pilpres 2019 berdasarkan Opini dari Twitter menggunakan Metode Naive Bayes Classifier," *Journal of Computer and Information Systems Ampera*, pp. 185-199, 2019.
- [2] A. M. Priyatno dan L. Ningsih, "Pembobotan TF-IDF untuk Mendeteksi Akun Spammer di Twitter berdasarkan Tweet dan Representasi Retweet dari Tweet," *Jurnal Sistem Informasi*, pp. 614-622, 2022.
- [3] O. A. Shvetsova, "Smart Education in High School : New Perspective in Global World," dalam *Quality Management, Transport and Information Security, Information Technologies*, 2017.
- [4] I. Zukhrufillah, "Gejala Media Sosial Twitter sebagai Media Sosial Alternatif," *Jurnal Komunikasi dan Penyiaran Islam*, pp. 102-109, 2018.
- [5] M. A. Rizaty, 6 November 2022. [Online]. Available: <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022>.
- [6] UUD RI, "Undang-Undang Republik Indonesia," Jakarta, 2013.
- [7] F. Nurhuda, S. W. Sihwi dan A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Teknologi dan Informasi ITSmart*, 2016.
- [8] Y. Nurtikasari, S. Alam dan T. I. Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film pada Platform Twitter Menggunakan Algoritma Naive Bayes," *Jurnal Sains dan Teknologi*, pp. 411-423, 2022.
- [9] M. F. A. S. Mustofa dan A. R. Pratama, "Analisis Sentimen atas Wabah Penyakit Covid-19 di Indonesia dengan CrowdTangle di Media Sosial Facebook," *Jurnal Ilmiah Elektronika dan Komputer*, pp. 89-96, 2022.
- [10] A. Anjani, "Analisis Sentimen Kaum LGBT pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes," *Jurnal Teknik Informatika*, pp. 1-8, 2022.
- [11] C. Prianto, N. H. Harani dan I. Firmansyah, "Analisis Sentimen terhadap Kandidat Presiden Republik Indonesia pada Pemilu 2019 di Media Sosial Twitter," *Jurnal Media Informatika Budidarma*, pp. 405-413, 2019.
- [12] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter," *Journal of Information Technology*, 2017.
- [13] E. Sutoyo, A. P. Rifai, A. Risnumawan dan M. Saputra, "A Comparison of Text Weighting Schemes on Sentiment Analysis of Government Policies: A Case Study of Replacement of National Examinations," *Multimed. Tools Appl*, p. 6413–6431, 2022.
- [14] I. M. d. Diego, A. R. Redondo, R. R. Fernandez dan J. Navarro, "General Performance Score for classification problems," *Applied Intelligence*, 2022.

- [15] H. Tohari, Analisis Serta Perancangan Sistem Informasi Melalui Pendekatan UML, Yogyakarta: Penerbit Andi, 2014.
- [16] Kusrini dan A. Kristanto, Tuntunan Praktis Membangun Sistem Informasi Akuntansi dengan Visual Basic dan SQL Server, Yogyakarta: Penerbit Andi, 2009.
- [17] A. Kadir, Pengenalan Sistem Informasi Edisi Revisi, Yogyakarta: Penerbit Andi, 2013.
- [18] F. Ilmiajayanti dan D. I. K. Dewi, "Persepsi Pengguna Taman Tematik Kota Bandung Terhadap Aksesibilitas Dan Pemanfaatannya," *RUANG, Volume 1, Nomor 1, 2015 ISSN: 2356-0088*, p. 23, 2015.