

## PENERAPAN *NATURAL LANGUAGE PROCESSING* DALAM ANALISIS SENTIMEN CAWAPRES 2024 MENGGUNAKAN ALGORITMA *NAIVE BAYES*

Iin, Reza Supriatna, Mulyawan, Dede Rohman

Teknik Informatika, STMIK IKMI CIREBON

Jalan Perjuangan No. 10 B Majasem Cirebon, Indonesia

Iin@gmail.com

### ABSTRAK

Pemerintah Indonesia akan mengadakan pemilu tahun 2024 isu yang paling di sorot yaitu pemilihan calon presiden dan calon wakil presiden tahun 2024, keputusan Mahkamah Konstitusi (MK) tentang pemilihan calon presiden dan wakil presiden pada tahun 2024 menjadi isu yang sangat viral, terutama terkait CAWAPRES muda. Beberapa analisis sentimen di Twitter menunjukkan bahwa teks yang disebutkan masyarakat memiliki sentimen negatif, positif, atau netral. Dalam konteks penelitian ini, fokusnya adalah menilai apakah teks yang muncul di Twitter bersifat negatif, positif, atau bahkan bersifat netral. Metode yang digunakan dalam penelitian ini metode SEMMA (sample, explore, modify, assess) dengan menggunakan algoritma *Naive Bayes*, dengan pemodelan confusion matrix untuk prediksi sentimen menggunakan *Naive Bayes* dengan menghasilkan nilai presisi, akurasi dan recall. Hasil evaluasi dari algoritma *Naive Bayes* tersebut mencapai 73,98% untuk nilai recall, 77,37% untuk nilai akurasi dan 100% untuk nilai presisi.

**Kata kunci :** cawapres 2024, pemilu, *Naive Bayes*, SEMMA, *Natural Language Processing*

### 1. PENDAHULUAN

Pemerintah Indonesia akan mengadakan pemilu tahun 2024 isu yang paling di sorot yaitu pemilihan calon presiden dan calon wakil presiden tahun 2024, Ketua Komisi Pemilihan Umum (KPU) Hasyim Asy'ari mengumumkan bahwa pendaftaran calon presiden dan calon wakil presiden telah selesai dan resmi ditutup. Tiga pasangan calon sudah resmi mendaftar sebagai peserta Pilpres 2024. Sebelumnya keputusan Mahkamah Konstitusi(MK) dalam menentukan cawapres 2024 sangat viral terutama tentang CAWAPRES muda, publik dikejutkan oleh putusan MK yang dinilai kontroversi. Putusan Mahkamah Konstitusi (MK) mengabulkan perkara nomor 90/PUU-XXI/2023 yang menyangkut usia minimal calon presiden dan wakil presiden dalam Undang-undang Nomor 7 Tahun 2017 tentang pemilu.

Keputusan tersebut menyatakan bahwa calon presiden dan calon wakil presiden yang pernah terpilih melalui pemilu, entah sebagai anggota DPR/DPD, Gubernur, atau Walikota, diizinkan mencalonkan diri meskipun usianya belum mencapai 40 tahun. Pemberitaan dimasyarakat pelaksanaan pemilu 2024 yang akan datang sangat tersebar luas, seperti isu keputusan Mahkamah Konstitusi dan Cawapres Muda, menurut masyarakat keputusan mahkamah konstitusi dalam menentukan batasan umur untuk calon wakil presiden dinilai banyak ke ganjilan banyak masyarakat Indonesia terutama di Twitter beranggapan bahwa Gibran putra dari presiden Jokowi ini belum siap dan ada juga yang pro mendukung keputusan MK dan cawapres Gibran yang maju berpasangan dengan Prabowo, Oleh karena itu, tujuan dari penelitian ini adalah untuk menganalisis sentimen masyarakat, khususnya di platform Twitter.

Tujuan penelitian ini untuk menerapkan algoritma *Naive Bayes* dalam analisis sentimen

cawapres 2024 di media sosial twitter terhadap keputusan Mahkamah konstitusi tentang batasan usia cawapres dengan mengetahui tingkat akurasi dari penggunaan algoritma *Naive Bayes* dalam pendeteksian komentar masyarakat di media sosial twitter yang menghasilkan sentimen positif, negatif atau netral dalam teks berbahasa Indonesia pada media sosial twitter, yang akan Memberikan pemahaman yang mendalam tentang wawasan politik yang sangat berharga.[1]

Penelitian ini menurut Yuliana Romadhoni, Khadijah Fahmi Hayati Holle yang berjudul Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode *Naive Bayes* dan LSTM, dalam penelitian ini menggunakan 2 metode yang berasal dari kedua model tersebut yaitu metode *Naive Bayes* dan metode Long Short-Term Memory. Pengumpulan data dengan cara crawling data menggunakan API Twitter Hasil penelitian ini menyatakan bahwa metode LSTM mendapatkan nilai performa lebih tinggi yaitu nilai Akurasi sebesar 77%, Presisi sebesar 84%, Recall sebesar 75%, dan F1-Score sebesar 80%. pengujian metode *Naive Bayes* didapatkan hasil Akurasi sebesar 76%, Presisi sebesar 75%, Nilai Recall sebesar 75% dan F1-Score sebesar 75%.[2]

Penelitian ini penulis menerapkan sebuah metode klasifikasi dengan metode klasifikasi *Naive bayes*. Menurut M. Walid and F. Halimiyah yang berjudul "Klasifikasi Kemandirian Siswa SMA/MA Double Track Menggunakan Metode *Naive Bayes*,"[3] pemilihan metode *Naive bayes* sebagai pendekatan utama didasarkan pada kemampuannya dalam menangani klasifikasi teks, terutama pada data yang berasal dari media sosial. Metode ini memanfaatkan probabilitas kondisional untuk mengklasifikasikan sentimen menjadi kategori positif, negatif, atau netral,

memberikan fondasi yang kuat untuk menganalisis pandangan masyarakat secara sistematis. Metode *Naive bayes* akan diimplementasikan untuk mengklasifikasikan sentimen masyarakat terhadap Cawapres 2024 berdasarkan analisis teks yang telah diproses. Penelitian ini yang saya buat pengambilan data menggunakan google colab crawling di Twitter dan pengolahan data aplikasi Rapid minner studio yang menghasilkan nilai confusion matrix.

## 2. TINJAUAN PUSTAKA

Hasil literature review yang telah dilakukan pada jurnal-jurnal penelitian terkait topik Penerapan *Natural Language Processing* Dalam Analisis Sentimen Cawapres 2024 Menggunakan algoritma *Naive bayes* dapat dijabarkan sebagai berikut:

*Paper 1* Menurut T. A. Sundara, S. Ekaputri, and S. Sotar, tahun 2020 Pembahasan tentang kebijakan larangan penggunaan cadar dan celana cingkrang bagi Aparatur Sipil Negara (ASN) telah menimbulkan beragam tanggapan di kalangan masyarakat, terutama pengguna Twitter. Respon masyarakat terhadap kebijakan tersebut dapat dianalisis melalui metode analisis sentimen, yang merupakan disiplin ilmu yang mempelajari pendapat, sentimen, evaluasi, sikap, dan emosi individu yang diungkapkan melalui teks. Pendapat masyarakat, khususnya dalam bentuk tweet, dikumpulkan melalui teknik web crawling menggunakan antarmuka Aplikasi Program Twitter (API). Data yang terkumpul kemudian diproses menggunakan teknik pemrosesan bahasa alami (Natural Language Processing/NLP) dan bahasa pemrograman Python. Tujuan dari penelitian ini adalah untuk menghasilkan informasi mengenai sentimen masyarakat terkait kebijakan larangan penggunaan cadar dan celana cingkrang bagi Aparatur Sipil Negara. Dalam penelitian ini, Algoritma *Naïve Bayes Classifier* digunakan untuk mengklasifikasikan tweet menjadi positif dan negatif. Hasil pengujian akurasi penelitian ini menunjukkan tingkat akurasi sebesar 86%. [4].

*Paper 2* menurut Akbar & Sugiharto, 2023 membahas mengenai ChatGPT, sebuah produk yang dikembangkan oleh OpenAI, perusahaan riset kecerdasan buatan yang berbasis di San Francisco, California, Amerika Serikat. ChatGPT pertama kali diperkenalkan pada tahun 2021 setelah mengalami pengembangan sejak tahun 2020. Fokus penelitian ini adalah Analisis Sentimen Pengguna Twitter di Indonesia terhadap ChatGPT. Penelitian ini menggunakan dua algoritma utama, yaitu Algoritma C4.5 dan *Naïve Bayes*, untuk menganalisis data yang diperoleh dari pengguna Twitter di Indonesia. Algoritma C4.5, diterapkan melalui perangkat lunak RapidMiner, digunakan untuk mengidentifikasi pola penggunaan Twitter yang paling signifikan dalam interaksi dengan ChatGPT. Sementara itu, metode *Naïve Bayes* digunakan untuk mengklasifikasikan jenis interaksi yang paling umum terjadi antara pengguna Twitter dan ChatGPT. Hasil analisis

menunjukkan hasil yang positif, dengan tingkat akurasi sebesar 77.33%, presisi sebesar 100.00%, dan recall sebesar 30.18%. Kesimpulan dari temuan ini menandakan bahwa respons atau sentimen yang diungkapkan oleh pengguna Twitter, khususnya di Indonesia, menunjukkan umpan balik positif terhadap ChatGPT dalam konteks interaksi di platform tersebut. [5].

*Paper 3* menurut Samsir, Ambiyar, Unung Verawardina & Watrianthos, 2021 membahas mengenai Analisis pembelajaran daring, Strategi untuk mengatasi pandemi dengan menerapkan pembatasan sosial telah mengakibatkan semua institusi pendidikan terpaksa beralih ke pembelajaran daring. Meskipun pembelajaran daring pada awalnya dianggap sebagai suatu strategi, namun pergeseran mendadak dari pembelajaran tatap muka ke pembelajaran daring dalam skala besar menimbulkan kontroversi akibat singkatnya proses adaptasi. Perubahan ini menimbulkan berbagai respons di masyarakat. Penelitian ini bertujuan untuk menganalisis opini publik terkait pembelajaran daring selama masa pandemi COVID-19 di Indonesia pada awal November 2020. Metode penelitian ini melibatkan penambangan teks berbasis dokumen dari platform Twitter, yang kemudian dianalisis menggunakan algoritma *Naïve Bayes*. Temuan dari penelitian ini menunjukkan bahwa pembelajaran daring pada periode tersebut menciptakan sentimen positif sebesar 30%, sentimen negatif sebesar 69%, dan sentimen netral sebesar 1%. Tingginya tingkat sentimen negatif disebabkan oleh ketidakpuasan masyarakat terhadap implementasi pembelajaran daring. Beberapa tweet mengungkapkan rasa kekecewaan dengan kata-kata seperti 'stres' dan 'malas', yang secara signifikan muncul dalam percakapan yang dianalisis [6].

*Paper 4* menurut Syarifuddin, 2020 membahas mengenai Aspirasi masyarakat kadang-kadang menghadapi kendala dalam penyampaiannya secara langsung kepada pihak yang bertanggung jawab. Kondisi ini mendorong masyarakat untuk menyuarakan aspirasi, kritik, dan sejenisnya melalui media sosial, salah satunya platform yang populer saat ini, yaitu Twitter. Ansambel aspirasi atau tweet dari pengguna Twitter yang mengungkapkan pandangan terhadap dampak Pembatasan Sosial Berskala Besar (PSBB), misalnya, dapat dijadikan bahan analisis sentimen terhadap opini publik. Sebanyak 170 opini mengenai efek PSBB diperoleh dan kemudian diproses melalui teknik penambangan data, yang mencakup langkah-langkah seperti penambangan teks, tokenisasi, transformasi, klasifikasi, dan stemming. Tiga algoritma berbeda, yaitu Decision Tree, K-Nearest Neighbors (K-NN), dan *Naïve Bayes Classifier*, diterapkan untuk perbandingan, dengan tujuan mencari akurasi terbaik. Aplikasi RapidMiner Versi 7.1 juga digunakan untuk mempermudah proses pengolahan data. Hasil tertinggi dari penelitian ini ditemukan pada algoritma Decision Tree dengan nilai

akurasi sebesar 83,3%, presisi sebesar 79%, dan *recall* sebesar 87,17% [7].

*Paper 5* Sulistio dan teman-teman, 2023 membahas mengenai Tanggapan konsumen terhadap layanan situs online marketplace umumnya disampaikan melalui media sosial, termasuk di Twitter. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap opini pelanggan terkait situs online marketplace di Indonesia di platform Twitter. Analisis sentimen ini dapat berguna untuk menetapkan peringkat situs online marketplace, membantu masyarakat dalam memilih platform belanja online di masa depan. Setelah proses clustering, terbentuk dua kelas data, yaitu kelas data positif dan negatif, dengan total 1200 data. Data yang sudah dikelompokkan ini digunakan sebagai data pelatihan untuk mesin pengklasifikasi. Dalam penelitian ini, algoritma mesin pengklasifikasi yang digunakan adalah *Naïve Bayes*. Terakhir, kinerja mesin pengklasifikasi diukur menggunakan metode 10-fold cross validation. Kesimpulannya, hasil evaluasi menunjukkan rata-rata akurasi sebesar 93,33% [8].

*Paper 6* menurut Khaerullah dan teman-teman, 2023 membahas mengenai *cryptocurrency*, Belakangan ini, keberadaan tagar #crypto di platform Twitter telah menciptakan tren topik global yang ramai diperbincangkan. Fenomena ini menghasilkan beragam opini dan pandangan dari pengguna Twitter. Mengingat banyaknya opini yang muncul terkait tagar tersebut, menentukan sentimen positif, negatif, atau netral secara manual menjadi tugas yang sulit dilakukan. Oleh karena itu, penerapan machine learning menjadi suatu kebutuhan untuk mengotomatisasi proses pelabelan tersebut, dengan tujuan mengidentifikasi sentimen dalam skala positif, netral, dan negatif. Melalui penggunaan Python TextBlob, analisis otomatis dilakukan terhadap total 1032 tweet yang dihasilkan, mengungkapkan bahwa 632 tweet atau 61.24% menyatakan sentimen positif, sementara sentimen netral terdapat pada 296 tweet atau 28.68%, dan sentimen negatif terdapat pada 104 tweet atau 10.07%. Kesimpulannya dalam menguji hasil pelabelan, algoritma *Naïve Bayes* digunakan dengan pembagian data testing sebesar 0.2 dan data training sebesar 0.8. Hasil pengujian menunjukkan nilai akurasi sebesar 71.98%, presisi 83.04%, *recall* 60.88%, dan *f1\_score* 65.07% [9].

*Paper 7* Irvandi et al., 2023 membahas mengenai Saat ini, Kepala Negara Indonesia telah menjabat selama dua periode berturut-turut, sesuai dengan peraturan yang melarang pencalonan ulang sebagai kepala negara. Sebagai tanggapan terhadap batasan ini, lembaga survei telah merilis hasil survei tentang tokoh-tokoh yang memiliki elektabilitas sebagai calon kepala negara. Banyak warganet juga telah menyuarakan pendapat mereka, dan dari pendapat-pendapat tersebut dapat disimpulkan mengenai sentimen masyarakat terhadap tokoh-tokoh potensial. Dalam penelitian ini, metode Knowledge Discovery

from Data digunakan dengan algoritma *Naïve Bayes* dan skor sentimen untuk menganalisis pendapat warganet. Tujuan penelitian ini adalah untuk memberikan referensi kepada masyarakat dalam pemilihan kepala negara pada pemilihan kepala negara mendatang. Hasil dan kesimpulan penelitian menunjukkan bahwa warganet umumnya memiliki sentimen positif terhadap setiap tokoh yang berpotensi menjadi calon kepala negara. Evaluasi dari algoritma *Naïve Bayes* menunjukkan akurasi sebesar 73,68% dan AUC 0,74 pada lipatan ke-7 untuk dataset pertama, akurasi 71,43% dan AUC 1,0 pada lipatan ke-5 untuk dataset kedua, akurasi 60% dan AUC 0,92 pada lipatan ke-1 untuk dataset ketiga, dan akurasi 62,5% dan AUC 0,65 pada lipatan ke-3 untuk dataset terakhir. [10].

*Paper 8* Twitter tetap menjadi salah satu platform media sosial yang populer di kalangan masyarakat. Penelitian ini memfokuskan pada analisis percakapan terkait RUU Cipta Kerja di Twitter, mengingat isu ini menjadi topik hangat di kalangan masyarakat Indonesia di platform tersebut. Tujuan dari penelitian ini adalah untuk mengidentifikasi kecenderungan komentar atau tweet pengguna Twitter terkait RUU Cipta Kerja melalui analisis sentimen, dengan mengklasifikasikan sentimen menjadi label positif, negatif, dan netral. Algoritma yang digunakan untuk analisis sentimen adalah *Naïve Bayes* karena dikenal memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan sentimen. Kesimpulan dari penelitian ini menunjukkan bahwa rasio training dan testing sebesar 75:25 dipilih karena memberikan tingkat akurasi tertinggi, yakni 88%, dibandingkan dengan rasio lainnya. Hasil prediksi dari klasifikasi cenderung lebih banyak berlabel positif. Selain melakukan klasifikasi sentimen, penelitian ini juga mengevaluasi performa dari model yang dikembangkan. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* memiliki precision sebesar 92%, recall 84%, dan *f1-score* 86%. [3].

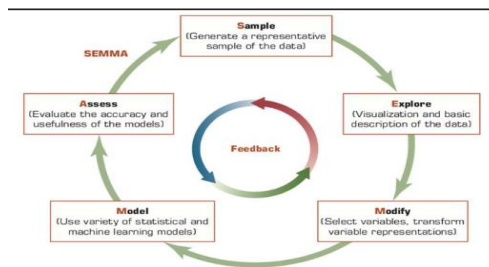
*Paper 9* membahas mengenai Tindakan body shaming dapat menurunkan tingkat kepercayaan diri bahkan menyebabkan korban melakukan hal ekstrim seperti bunuh diri akibat merasa tertekan. Body shaming sering terjadi di media sosial, salah satu media sosial yang banyak diminati adalah Twitter. Untuk mencegah tindakan yang berbahaya tersebut, diperlukan sebuah pendeteksi body shaming. Penelitian ini bertujuan untuk melakukan pendeteksian terhadap kata yang mengandung unsur penindasan body shaming pada tweet dan mengetahui akurasi model klasifikasi dengan menggunakan metode *Naïve Bayes*. Model yang dibangun menggunakan algoritma *Naïve Bayes*. Dari percobaan yang dilakukan diperoleh hasil akurasi sebesar 74,70% dengan komposisi data train 70% dan data test 30% [11].

*Paper 10* membahas mengenai Twitter tetap menjadi salah satu platform media sosial yang populer di kalangan masyarakat. Penelitian ini membahas perbincangan seputar RU Cipta Kerja di Twitter,

mengingat kasus ini menjadi perbincangan hangat di kalangan masyarakat Indonesia di platform tersebut. Tujuan dari penelitian ini adalah untuk mengidentifikasi kecenderungan komentar atau tweet pengguna Twitter terkait RU Cipta Kerja melalui analisis sentimen, dengan mengklasifikasikan sentimen menjadi label positif, negatif, dan netral. Algoritma yang digunakan untuk analisis sentimen adalah Naïve Bayes karena dikenal memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan sentimen. Kesimpulan dari penelitian ini menunjukkan bahwa rasio training dan testing sebesar 75:25 dipilih karena memberikan tingkat akurasi tertinggi, yakni 88%, dibandingkan dengan rasio lainnya. Hasil prediksi dari klasifikasi cenderung lebih banyak berlabel positif. Selain melakukan klasifikasi sentimen, penelitian ini juga mengevaluasi performa dari model yang dikembangkan. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes memiliki precision sebesar 92%, recall 84%, dan f1-score 86%. [12].

### 3. METODE PENELITIAN

Dalam penelitian ini, peneliti mencoba menerapkan beberapa teknik Natural Language Processing (NLP) dan Machine Learning. Secara keseluruhan, pendekatan metodologi yang digunakan dalam penelitian ini mengacu pada proses Data Mining SEMMA.[13]. Gambar 1 menjelaskan tahapan yang dilakukan dalam penelitian ini.



Gambar 1 Semma Data Mining Process

Sumber gambar 1 (Alizah et al., 2020) dalam jurnal *Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter* [13]

#### 3.1. Sample

Pada fase Pengumpulan Data, dilakukan pengumpulan data terkait tweets "Cawapres 2024" dari platform media sosial Twitter menggunakan API Twitter. Data yang berhasil terkumpul disimpan dalam format CSV pada sebuah file. Selanjutnya, data tersebut dibaca dan dipersiapkan untuk proses selanjutnya.

#### 3.2. Explore

Pada tahap ini, dijelaskan tentang deskripsi data dan visualisasi data. Deskripsi data akan memberikan gambaran umum tentang informasi yang terkandung dalam data yang digunakan. Sementara itu, visualisasi data akan menampilkan informasi secara visual dari data tersebut.

#### 3.3. Modify

Pada tahap ini dilakukan beberapa metode persiapan data diantaranya:

##### 3.3.1. Case Folding

*Transform case* atau *case folding* merupakan proses pemerataan huruf dari huruf kapital menjadi huruf kecil atau sebaliknya. Pada penelitian ini data set diubah menjadi huruf kecil semua karena tweet yang sudah diambil sebagian besar menggunakan huruf kecil.

##### 3.3.2. Pembersihan (URL, Username, Hashtag)

Pembersihan ini dilakukan untuk menghapus URL atau tautan link website, Username, dan Hashtag (#Cawapres2024) yang ada pada tweets yang ada pada data.

##### 3.3.3. Tokenize

Tokenisasi adalah proses pemecahan teks menjadi bagian-bagian yang lebih kecil, yang diartikan oleh token dalam bahasa pemrograman. Tokenisasi dilakukan untuk memecah teks menjadi kata-kata yang memiliki arti tertentu, sehingga kata-kata tersebut dapat diolah lebih lanjut dalam proses, seperti analisis sentimen atau proses selanjutnya.

##### 3.3.4. Filtering

Dalam proses ini dilakukan penghapusan kata yang memiliki huruf terlalu pendek atau yang disingkat dan huruf yang terlalu panjang.

##### 3.3.5. Stemming

Stemming adalah teknik yang digunakan untuk mereduksi kata-kata ke bentuk dasarnya. Tujuannya adalah untuk membersihkan kata-kata dari variasi pengejaan yang kurang tepat.

#### 3.4. Model

Pada tahap ini, data akan dipecah menjadi dua, yaitu data uji dan data latih dengan rasio 3:7. Setelah itu, dilakukan proses ekstraksi fitur dari tweets yang terambil sehingga selanjutnya tweets tersebut dapat dilakukan proses pemodelan machine learning menggunakan data latih untuk melakukan prediksi label pada sentimen tweets pada data uji.

#### 3.5. Assess

Pada tahap ini akan dilakukan evaluasi terhadap pemodelan yang telah dibuat. Evaluasi yang dimaksud adalah membandingkan hasil yang didapat dari model prediksi terhadap data uji dengan label sentimen pada data uji yang sebelumnya sudah diberi label. Hasil evaluasi yang dilakukan dihitung berdasarkan besaran dari precision, recall, f1-score dan akurasi.

### 4. HASIL DAN PEMBAHASAN

Dalam bab ini, akan diuraikan hasil dari penelitian yang dilakukan. Dalam eksperimen ini, peneliti menggunakan Python untuk pencarian data dan RapidMiner Studio untuk pengolahan data dalam menerapkan konsep yang diusulkan. Hasil penelitian ini akan disajikan dalam beberapa tahap sesuai dengan metodologi yang digunakan.

#### 4.1. Hasil pada tahap Sample

Dalam pengambilan dataset pada penelitian ini menggunakan teknik *Crawling* data, penelitian ini menggunakan tools google colab dengan bahasa pemrograman python yang disimpan dalam bentuk file csv. Peneliti menggunakan API Twitter dengan menggunakan modul *tweepy* pada bahasa pemrograman Python.

#### 4.2. Hasil pada tahap Explore

Hasil *Crawling* terdiri 3 kolom yang berisi id merupakan nomor id twitter, *created\_at* merupakan waktu tweet diunggah oleh pengguna, *from\_user\_id* merupakan id pengguna twitter, dan *text* merupakan isi tweet. Hasil *crawling* ada 1495 cuitan masyarakat. Agar mempermudah dalam menganalisis data, maka hanya atribut *Tweets* saja yang ditampilkan.

Tabel 1 Contoh Data set Hasil *Crawling Twitter*

| No | Tweets   |
|----|--|
| 1  | "@mirashahabudin @Chaves1305 Prabowo Subianto dan Gibran Rakabuming Raka menunjukkan ketulusan mereka dengan ikut maju dalam pemilihan capres dan cawapres tahun 2024  |
| 2  | "@sandiuno Setelah jadi cawagub dan sukses jadi Wagub Indonesia meloncat jadi cawapres di pilpres 2019... harapan saya Mas Menteri ikut berlayar jadi capres di pilpres 2024   |
| 3  | "PRABOWO SUDAH PASTI MASUK RONDE 2 PADA PILPRES 2024 ??? Selengkapanya hanya ada di youtube Skala Survei Indonesia <a href="https://t.co/3MaMCHhy3X">https://t.co/3MaMCHhy3X</a> #Pilpres #Cawapres #KPU #Parpol #PrabowoSubianto #GanjarPranowo #aniesbaswedan #SSI |
| 4  | "@jokowi Selamat Pak Prestasi mu tetap kami apresiasi namun kesalahan fatal terhadap putusan MK yg terindikasi ini dan itu dan tdk ada larangan dari bapak terhadap Putra sulung bapak menjadi cawapres jg akan kami hitung dan pertimbangkan tgl 14 Feb 2024        |
| 5  | "Bobby Nasution menyatakan akan mendukung Capres Prabowo Subianto dan Cawapres Gibran Rakabuming pada Pilpres 2024   |
| 6  | "Relawan optimistis Ganjar-Mahfud menang Pilpres 2024  |
| 7  | "Jadi sekarang ini yg menguasai pemerintahan adalah PKI  |
| 8  | "@Fahrihamzah Ratusan trilyun itu ngucur Khan krn ugal2an bapak si bocil ngutang dan kebijakannya itu cacat hukum. Sama halnya dgn cawapres jagoanmu yg cacat hukum. Ketika penantang jd pecundang kyk km.....   |
| 9  | "ini mah cawapres terganteng 2024 @gibran_tweet pak @mohmahfudmd cak @cakimiNOW lewat  |
| 10 | "Kesombongan partai yg merasa bisa sendirian ajukan capres/cawapres  |

#### 4.3. Hasil pada Modify

Pada tahap *modify*, Data set mengalami modifikasi melalui proses *text preprocessing*. Tujuan dari langkah ini adalah untuk meningkatkan

kemampuan sistem komputer dalam mengenali format data set. Selain itu, tahap ini dapat mengubah data yang tidak terstruktur menjadi data yang terstruktur dengan lebih teratur. Dalam *text preprocessing* terdapat beberapa proses persiapan data yaitu *Case folding*, *Pembersihan (URL, Username, Hashtag)*, *Tokenize*, *Filtering* dan *Stemming*.

| id | created_at           | text  |
|----|----------------------|---|
| 1  | 2024-01-15T10:00:00Z | "@mirashahabudin @Chaves1305 Prabowo Subianto dan Gibran Rakabuming Raka menunjukkan ketulusan mereka dengan ikut maju dalam pemilihan capres dan cawapres tahun 2024   |
| 2  | 2024-01-15T10:00:00Z | "@sandiuno Setelah jadi cawagub dan sukses jadi Wagub Indonesia meloncat jadi cawapres di pilpres 2019... harapan saya Mas Menteri ikut berlayar jadi capres di pilpres 2024  |
| 3  | 2024-01-15T10:00:00Z | "PRABOWO SUDAH PASTI MASUK RONDE 2 PADA PILPRES 2024 ??? Selengkapanya hanya ada di youtube Skala Survei Indonesia https://t.co/3MaMCHhy3X #Pilpres #Cawapres #KPU #Parpol #PrabowoSubianto #GanjarPranowo #aniesbaswedan #SSI                                |
| 4  | 2024-01-15T10:00:00Z | "@jokowi Selamat Pak Prestasi mu tetap kami apresiasi namun kesalahan fatal terhadap putusan MK yg terindikasi ini dan itu dan tdk ada larangan dari bapak terhadap Putra sulung bapak menjadi cawapres jg akan kami hitung dan pertimbangkan tgl 14 Feb 2024 |
| 5  | 2024-01-15T10:00:00Z | "Bobby Nasution menyatakan akan mendukung Capres Prabowo Subianto dan Cawapres Gibran Rakabuming pada Pilpres 2024  |
| 6  | 2024-01-15T10:00:00Z | "Relawan optimistis Ganjar-Mahfud menang Pilpres 2024   |
| 7  | 2024-01-15T10:00:00Z | "Jadi sekarang ini yg menguasai pemerintahan adalah PKI   |
| 8  | 2024-01-15T10:00:00Z | "@Fahrihamzah Ratusan trilyun itu ngucur Khan krn ugal2an bapak si bocil ngutang dan kebijakannya itu cacat hukum. Sama halnya dgn cawapres jagoanmu yg cacat hukum. Ketika penantang jd pecundang kyk km.....  |
| 9  | 2024-01-15T10:00:00Z | "ini mah cawapres terganteng 2024 @gibran_tweet pak @mohmahfudmd cak @cakimiNOW lewat   |
| 10 | 2024-01-15T10:00:00Z | "Kesombongan partai yg merasa bisa sendirian ajukan capres/cawapres   |

Gambar 2 hasil tahap *modify*

Setelah melakukan serangkaian *Text preprocessing* gambar diatas adalah hasil nya yang awal data 1945 menjadi 853 data dan siap ke tahap model.

#### 4.4. Hasil pada Proses Model

Pada tahap ini, data akan dibagi menjadi dua bagian, yaitu data latih dan data uji dengan rasio 70% untuk data latih dan 30% untuk data uji. Setelah itu, dilakukan proses ekstraksi fitur dari tweets yang terambil sehingga tweets tersebut dapat diproses menggunakan teknik *machine learning* dengan menggunakan data latih untuk memprediksi label sentimen pada data uji. Tahap pemodelan dalam penelitian ini mencakup pelabelan secara manual dan pemodelan klasifikasi dengan menggunakan metode *Naive Bayes*.

#### 4.5. Hasil pada Proses Assess

Tahap *assess* digunakan untuk mengevaluasi metode *Naive Bayes*. Evaluasi metode ini melibatkan nilai *confusion matrix* yang mencakup akurasi, presisi, dan *recall* yang diambil dari data tes.

```

PerformanceVector
PerformanceVector:
accuracy: 77.37%
ConfusionMatrix:
True: positif negatif neutral
positif: 381 0 0
negatif: 10 157 0
neutral: 124 59 122
kappa: 0.641
ConfusionMatrix:
True: positif negatif neutral
positif: 381 0 0
negatif: 10 157 0
neutral: 124 59 122

```

Gambar 3 *Confusion Matrix*

Algoritma *Naive Bayes* mencapai tingkat akurasi sebesar 77,37%, yang berarti sekitar 77,37% dari model *Naive Bayes* mampu mengklasifikasikan data secara benar. Jumlah sentimen positif sebanyak 515, sentimen negatif sebanyak 216, dan sentimen netral sebanyak 122. Presisi, yang merupakan perbandingan antara hasil prediksi dan data yang diminta, mencapai 73,98% untuk prediksi positif, 72,69% untuk prediksi negatif, dan 100,00% untuk prediksi netral. Nilai *recall* mencerminkan keberhasilan model dalam menemukan kembali informasi, dengan hasil 100,00%

untuk data positif, 94,01% untuk data negatif, dan 40,00% untuk data netral.

|              | true positif | true negatif | true netral | class precision |
|--------------|--------------|--------------|-------------|-----------------|
| pred positif | 381          | 0            | 0           | 100.00%         |
| pred negatif | 10           | 157          | 0           | 94.01%          |
| pred netral  | 124          | 59           | 122         | 40.00%          |
| class recall | 73.98%       | 72.69%       | 100.00%     |                 |

Gambar 4 akurasi, presisi, recall

Lalu untuk menghitung Accuracy, precision dan recall berikut cara dan rumusnya

Diketahui :  $Accuracy = (TP+TN+TNR) / (TP+PP+TN+PN+TNR+PNR)$

$Accuracy =$

$(381+157+122)/(381+0+0+$

$10+157+0+124+59+122)$

$= 0.7737$

$= 0.7737 * 100 \% = 77,37\%$

Bisa dilihat bahwa perhitungan *Accuracy Naive bayes* mendapatkan hasil dengan tingkat *Accuracy* sebesar 77,37%.

Setelah perhitungan nilai *Akurasi*, langkah berikutnya adalah menghitung nilai *recall*. *Recall*

merupakan perbandingan antara jumlah item relevan yang dipilih dengan total jumlah item relevan yang ada. *Recall* dapat dihitung menggunakan rumus berikut :

$Recall = TP / (TP+PN+PNR)$

$Recall = 381 / 381+10+124$

$Recall = 0.7398$

$Recall = 0.7398 * 100\% = 73,98\%$

Bisa dilihat bahwa perhitungan *Recall* pada *Naive bayes* mendapatkan 73,98%.

Setelah perhitungan nilai *Recall*, langkah berikutnya adalah menghitung nilai *precision*. *Precision* merupakan perbandingan antara jumlah item relavan yang dipilih dengan jumlah item yang terpilih. *Precision* dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban. *Precision* dapat dihitung dengan rumus sebagai berikut :

$Precision = TP / (TP + TP + TNR)$

$Precision = 381 / (381+0+0)$

$Class Precision Positif = 1 * 100\% = 100 \%$

Dapat dilihat bahwa perhitungan *Precision* pada *Precision Class Positif* sebesar 100 % dengan metode *naive bayes*.

Tabel 2 Tabel Hasil Prediksi Sentimen Rapid miner

| Rows No | Sentimen | Prediksi (sentimen) | Confiden (Positif) | Confiden (Negatif) | Confiden (netral) | Text   |
|---------|----------|---------------------|--------------------|--------------------|-------------------|--|
| 1       | positif  | positif             | 1                  | 0                  | 0                 | Tentukan pilihan anda  |
| 2       | negatif  | negatif             | 0                  | 1                  | 0                 | Strategi projokowi   |
| 3       | negatif  | negatif             | 0                  | 1                  | 0                 | Saya kira quote darijokowi sudah pak ngga usah ngamuk ngamuk menggantikan prabowo itu semua bias terjadi |
| 4       | positif  | positif             | 1                  | 0                  | 0                 | Jawabannya sederhana   |
| 5       | positif  | positif             | 1                  | 0                  | 0                 | Indonesia bangkit  |

Tabel diatas adalah hasil prediksi sentimen dari Rapid miner dengan secara otomatis menghasilkan sentimen yang menurut rapid miner tepat.

## 5. KESIMPULAN DAN SARAN

Hasil dari metode *Naive Bayes* dengan jumlah 515 sentimen positif, 216 sentimen negative dan 122 sentimen netral. Nilai akurasi 77.37%, presisi prediksi positif 100,00%, presisi prediksi negatif 94,01% dan presisi prediksi netral 40,00%, *recall* data positif 73,98%, *recall* data negatif 72,69% dan *recall* data netral 100,00%. Dengan merujuk pada nilai akurasi, presisi, dan *recall* metode *Naive bayes* mempunyai nilai akurasi yang tinggi. Berdasarkan hasil sentimen dari *Naive bayes*, sentiment positif berjumlah 515 dan sentimen negatif berjumlah 216 dan sentiment netral berjumlah 122. Dengan demikian, terdapat lebih banyak sentimen positif daripada sentimen negatif dan netral, yang menunjukkan bahwa pengguna Twitter cenderung memberikan pendapat positif terkait Pemilu pada Cawapres 2024. Untuk penelitian selanjutnya,

disarankan untuk memperluas cakupan dengan menggunakan platform media sosial lain seperti Facebook, Instagram, dan YouTube, termasuk konten video. Selain itu, dalam pengolahan data, dapat dipertimbangkan penggunaan aplikasi lain dan bahasa pemrograman seperti Python atau R. Selain itu, penelitian mendatang dapat mengeksplorasi penambahan algoritma dan metode lain seperti KDD, SVM, dan sebagainya. Selain Bahasa Indonesia, penelitian selanjutnya juga dapat memanfaatkan bahasa lain untuk analisis sentimen.

## DAFTAR PUSTAKA

- [1] . Diyank Shabira, Mufid Nilmada, "IMPLEMENTASI METODE NAÏVE BAYES PADA ANALISIS SENTIMEN BODY SHAMING DI MEDIA SOSIAL TWITTER".
- [2] Y. Romadhoni and K. F. H. Holle, "Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM," *J. Inform. J. Pengemb.*



- IT*, vol. 7, no. 2, pp. 118–124, 2022, doi: 10.30591/jpit.v7i2.3191.
- [3] M. Walid and F. Halimiyah, “Klasifikasi Kemandirian Siswa SMA/MA Double Track Menggunakan Metode Naïve Bayes,” *J. ICT Inf. Commun. Technol.*, vol. 22, pp. 190–197, 2022.
- [4] T. A. Sundara, S. Ekaputri, and S. Sotar, “Naïve Bayes Classifier untuk Analisis Sentimen Isu Radikalisme,” *Pros. SISFOTEK*, vol. 4, no. 1, pp. 93–98, 2020, [Online]. Available: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/159>
- [5] Y. Akbar and T. Sugiharto, “Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes,” *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 115–122, 2023, [Online]. Available: <https://doi.org/10.55338/saintek.v4i3.1368>
- [6] F. E. Samsir, Ambiyar, Unung Verawardina and R. Watrianthos, “Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [7] M. Syarifuddin, “Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Dan Naïve Bayes,” *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.
- [8] M. R. Sulistio, N. Suarna, and O. Nurdianan, “Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang,” *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, pp. 37–42, 2023, doi: 10.56854/jtik.v1i2.49.
- [9] R. R. Khaerullah, N. Suarna, and O. Nurdianan, “Analisa Pengelompokan Dataset Komputer Menggunakan Algoritma X-Means,” *J. Inform. dan Teknol. Inf.*, vol. 1, no. 2, pp. 125–132, 2023, doi: 10.56854/jt.v1i2.135.
- [10] I. Irvandi, B. Irawan, and O. Nurdianan, “Naive Bayes and Wordcloud for Sentiment Analysis of Halal Tourism in Lombok Island Indonesia,” *Innov. Res. Informatics*, vol. 5, no. 1, pp. 30–35, 2023, doi: 10.37058/innovatics.v5i1.6675.
- [11] O. Nurdianan, A. Irma Purnamasari, and I. Ali, “Analisa Penjualan Mobil Dengan Menggunakan Algoritma K-Means Di PT. Mulya Putra Kencana,” *J. Data Sci. dan Inform.*, vol. 1, no. 2, pp. 32–35, 2021.
- [12] T. Prasetyo, H. Zakaria, and P. Wiliantoro, “Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes Classifier,” *OKTAL J. Ilmu Komput. dan Sains*, vol. 1, no. 6, pp. 573–582, 2022, [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal>
- [13] M. D. Alizah, A. Nugroho, U. Radiyah, and W. Gata, “Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter,” *Indones. J. Softw. Eng.*, vol. 6, no. 2, pp. 223–229, 2020, doi: 10.31294/ijse.v6i2.8991.