

Analisis Sentimen Masyarakat Di Twitter Terhadap Pemerintahan Anies Baswedan Menggunakan Metode Naive Bayes Classifier

¹Febriyanti Rambu Bangu Kahi, ²Alfrian C. Talakua, ³Reynaldi Thimotius Abineno.

^{1, 2, 3} Universitas Kristen Wira Wacana Sumba, Indonesia

¹febryantimaglodz27@gmail.com, ²alfriantalakua@unkriswina.ac.id, ³reynaldi@unkriswina.ac.id

ABSTRAK

Twitter adalah platform media sosial yang banyak digunakan di Indonesia. Dengan adanya media sosial dapat dimanfaatkan untuk menggalang opini masyarakat di Twitter mengenai kinerja seorang pemimpin dengan kata-kata yang menyinggung dan provokatif. Banyak cuitan di Twitter yang menjadi berita hangat di Indonesia, salah satunya Anies Baswedan. Anies Baswedan adalah mantan Gubernur Daerah Khusus Ibu Kota Jakarta yang dikenal sikap kepemimpinannya tegas dan membawa perubahan. Saat ini banyak media yang memberikan informasi bahwa Anies Baswedan mencalonkan diri sebagai Presiden Republik Indonesia pada tahun 2024 mendatang. Pencalonan Anies Baswedan sebagai Presiden Indonesia, tidak bisa dipungkiri bahwa banyak masyarakat yang ingin mengetahui program – program apa saja yang sudah menjadi bukti nyata selama masa pemerintahan Anies Baswedan. Dengan demikian, banyak opini dari masyarakat tentang kinerja Anies Baswedan selama masa jabatannya, karena dengan masyarakat mengetahui program dan adanya bukti nyata yang sudah diwujudkan Anies Baswedan pada masa pemerintahannya. Sehingga karya Anies Baswedan menjadi tolak ukur bagi masyarakat untuk mempertimbangkan Anies Baswedan layak dipilih menjadi Presiden Republik Indonesia atau tidak. Algoritma yang digunakan dalam analisis ini adalah klasifikasi Naive Bayes. Naive Bayes Classifier merupakan salah satu metode machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Manfaat penelitian ini untuk mengetahui berapa banyak sentimen opini positif dan negatif masyarakat terhadap Pemerintahan Anies Baswedan selama menjabat sebagai Gubernur Daerah Khusus Ibu Kota Jakarta dan akan mencalonkan diri untuk menjadi caleg Presiden di tahun 2024 mendatang. Performa teknik klasifikasi tingkat akurasi 76,04%, presisi 71,43% *recall* 81,63 % dan *f-measure* 76,19%. Hasil akhir dari penelitian ini tanggapan masyarakat berkategori positif sebanyak 746 (47,7%), berkategori negatif lebih banyak dengan jumlah data sebanyak 817 (52,3%).

Kata kunci: Anies Baswedan, Analisis Sentimen, Naive Bayes, Twitter.

PENDAHULUAN

Seiring kemajuan teknologi yang berkembang sangat pesat saat ini, dan hadirnya berbagai media yang dapat digunakan masyarakat untuk melakukan komunikasi dan mengakses informasi secara cepat dengan memanfaatkan koneksi internet. Internet merupakan sarana informasi dan komunikasi yang dimiliki setiap orang, dimana saja dan kapan saja. Begitu pula penggunaan Internet. Indonesia merupakan salah satu negara dengan jumlah pengguna internet tertinggi di dunia. Berdasarkan hasil survei yang dilakukan APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) pada tahun 2023. Mencatat pengguna internet di Indonesia telah mencapai 78,19 persen pada 2023 atau menembus 215.626.156 jiwa dari total populasi yang sebesar 275.773.901 jiwa. Banyak media yang dapat digunakan untuk berinteraksi dengan sesama pengguna, misalnya Twitter, Instagram, Facebook, dan lain-lain.

Penggunaan internet dan jumlah platform yang menyediakan layanan bagi masyarakat untuk mengakses Internet semakin meningkat. Media sosial Twitter adalah salah satu platform yang

paling banyak digunakan di Indonesia untuk berkomunikasi dan mendapatkan informasi dengan cepat. Informasi di berbagai jejaring sosial, seperti komentar, review dan opini dari pengguna media sosial. Kasus ini menunjukkan banyak sekali dampak yang dilihat masyarakat Indonesia di platform twitter yang dapat digunakan untuk studi kasus. Twitter digunakan sebagai sarana untuk menyampaikan ide kepada para pemimpin dan untuk mengekspos peristiwa dan budaya yang sedang terjadi. Dengan adanya media sosial dapat dimanfaatkan untuk menggalang opini masyarakat di twitter mengenai kinerja seorang pemimpin dengan kata-kata yang menyinggung dan provokatif. Banyak cuitan di twitter yang menjadi berita hangat di Indonesia, salah satunya Anies Baswedan.

H. Anies Rasyid Baswedan, S.E.,M.P.P.,Ph.D lahir di Kuningan (Jawa Bar-at), 7 Mei 1969. Anies Baswedan adalah Mantan Gubernur Daerah Khusus Ibu Kota Jakarta yang dikenal sikap kepemimpinannya tegas dan membawa perubahan. Salah satu karya yang dilakukan Anies Baswedan selama masa jabatan yaitu pembangunan Stadion Jakarta Internasional Stadium (JIS), dirancang sesuai standar Federation Internasional de Football Association (FIFA) dan merupakan salah satu mega proyek berskala dunia yang di bangun diatas lahan seluas 22 hektare di Jakarta Utara. Anies Baswedan adalah salah satu alumnus, Fakultas Ekonomi, Universitas Gadjah Mada. Anies Baswedan telah menyelesaikan masa pemerintahannya sebagai Gubernur daerah khusus Ibu Kota Jakarta sejak 16 Oktober 2022, saat ini banyak media yang memberikan informasi bahwa Anies Baswedan mencalonkan diri sebagai Presiden Republik Indonesia pada tahun 2024 mendatang. Pencalonan Anies Baswedan sebagai Presiden Indonesia pada tahun 2024 mendatang tidak terlepas dari opini masyarakat tentang masa jabatannya sebagai Gubernur Daerah Khusus Ibu Kota Jakarta, tidak bisa dipungkiri bahwa banyak masyarakat yang ingin mengetahui program – program apa saja yang sudah menjadi bukti nyata selama masa pemerintahan Anies Baswedan. Dengan demikian, banyak opini dari masyarakat tentang kinerja Anies Baswedan selama masa jabatannya, karena dengan masyarakat mengetahui program dan adanya bukti nyata yang sudah diwujudkan Anies Baswedan pada masa pemerintahannya. Sehingga karya Anies Baswedan menjadi tolak ukur bagi masyarakat untuk mempertimbangkan Anies Baswedan layak dipilih menjadi Presiden Republik Indonesia atau tidak. Masyarakat sangat ingin tahu latar belakang seorang politikus yang ingin menjadi pemimpin di Indonesia untuk mengetahui tujuan perubahan yang akan kerjakan nantinya oleh pemimpin ketika masyarakat memberikan kepercayaan un-tuk memberikan perubahan yang baru.

Opini adalah pola pikir seseorang dalam memberi tanggapan saat terjadi permasalahan di mana suatu opini dapat mengandung kalimat positif dan negatif. Dari opini masyarakat tentang masa pemerintahan Anies Baswedan digunakan sebagai sumber data analisis sentimen untuk mengetahui berapa banyak opini positif dan negatif.

Berdasarkan permasalahan di atas, penulis melakukan analisis sentimen dengan menggunakan opini masyarakat terhadap masa pemerintahan Anies Baswedan. Pada penelitian ini klasifikasi analisis sentimen dibagi menjadi dua konsep yaitu sentimen positif dan sentimen negatif menggunakan algoritma Naive Bayes. Naive Bayes Classifier merupakan salah satu metode machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

Tujuan dari penelitian ini adalah untuk melakukan analisis sentimen terhadap pemerintahan Anies Baswedan menggunakan algoritma naive Bayes classifier.

TINJAUAN PUSTAKA

Tabel 1 Penelitian terdahulu

No.	Judul dan Penulis	Metode	Variabel	Hasil
1	Analisis Sentimen dengan SVM, NAIVE BAYES Dan KNN untuk Studi	Svm, Naive Bayes Dan KNN	Twitter membahas Tanggapan Masyarakat	Kesimpulan bahwa algoritma SVM memiliki akurasi yang lebih tinggi daripada Naive Bayes dan

No.	Judul dan Penulis	Metode	Variabel	Hasil
	Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter, Fajar Sodik Pamungka, dkk (2021)		Indonesia Terhadap Pandemi Covid-19	<i>KNN</i> dengan rata-rata akurasinya sebesar 90,01% pada <i>SVM</i> dengan <i>kernel linear</i> , 79,20% pada <i>Naive Bayes</i> dengan jumlah <i>laplace</i> adalah 1, dan 62,10% pada <i>KNN</i> dengan jumlah <i>K</i> adalah 20 dan menggunakan <i>kernel optimal</i> .
2	Analisis Sentimen Tanggapan Masyarakat Terhadap bantuan Sosial pemerintah Di Masa Pandemi Covid-19 Pada Platform <i>Twitter</i> , Melani Asta Rosari, dkk (2022)	<i>Naive Bayes Classification</i>	<i>Twitter</i> membahas Analisis Sentimen Tanggapan Masyarakat Terhadap bantuan Sosial pemerintah Di Masa Pandemi Covid-19	Klasifikasi ini menghasilkan 2.382 atribut kata atau vektor kata dari 747 data, dengan 370 data sampel untuk pengujian model yang menghasilkan nilai akurasi sebesar 24,32%, nilai <i>recall</i> yang benar-benar netral sebesar 100%, dan nilai presisi netral sebenarnya sebesar 24,32%. Kata yang paling sering muncul dari hasil ini analisis sentimen adalah kata “bantuan”.
3	Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada <i>Twitter</i> Menggunakan Metode <i>Naive Bayes Classifier</i> , Dianati Duei, dkk (2022)	Menggunakan Metode <i>Naive Bayes</i>	Kinerja Dewan Perwakilan Rakyat (DPR) Pada <i>Twitter</i>	Hasil dari penelitian ini didapatkan bahwa DPR mendapatkan 95 <i>tweet</i> positif dengan polaritas 0.75 atau 75% sentimen positif, 693 <i>tweet</i> netral dengan polaritas 0.79 atau 79% sentimen netral dan 758 <i>tweet</i> negatif dengan polaritas 0.82 atau 82% sentimen negatif dengan <i>accuracy score</i> 0.8 atau 80% berdasarkan data testing sebanyak 20%.
4	Analisis Sentimen Masyarakat Terhadap Prabowo Subianto Menggunakan Support Vector Machine, Nabila Tasya Amalia, (2023)	<i>Support Vector Machine</i>	Analisis Sentimen Masyarakat Terhadap Prabowo	Hasil dari pembobotan dan labelling, di klasifikasi menggunakan <i>Support Vector Machine</i> Dari penelitian ini diperoleh hasil performa sistem yaitu akurasi sebesar 94,72%, <i>recall</i> sebesar 98,15%, dan presisi sebesar 96,02%.

No.	Judul dan Penulis	Metode	Variabel	Hasil
				Berdasarkan data twitter yang diperoleh, pada tahun 2022 masyarakat cenderung menilai Prabowo secara positif daripada negatif.
5	Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM, Mohammed Hafizh Al- Areef, ddk(2023)	Menggunakan Algoritma LSTM	Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024	Hasil evaluasi performa untuk setiap model yang telah dilatih adalah 82% akurasi, 86% presisi, 92% <i>recall</i> , dan 89% <i>f1-score</i> pada model Ganjar Pranowo. 82% akurasi, 82% presisi, 96% <i>recall</i> , dan 89% <i>f1-score</i> pada model Prabowo Subianto. 87% akurasi, 91% presisi, 95% <i>recall</i> , 93% <i>f1-score</i> pada model Ridwan. 87%.

Twitter

Twitter adalah jejaring sosial yang digunakan untuk berbagi pesan teks, foto, video, berita, informasi, dll. Berdasarkan laporan Kominfo, pengguna Twitter berdasarkan data PT Bakrie Telecom merupakan pengguna di seluruh dunia. Twitter telah menjadi salah satu jejaring sosial terbesar di dunia, dengan pendapatan \$145 juta. Pengguna Twitter di Indonesia sebagian besar merupakan konsumen, yaitu masyarakat yang tidak memiliki blog atau belum mengunggah video ke YouTube namun mengupdate statusnya di Twitter dan Facebook.

Analisis Sentimen

Analisis sentimen adalah alat pengorganisasian hasil pencarian untuk menganalisis karakteristik produk (kualitas, fitur, dan lain-lain) dan proses memperoleh opini.

Tugas utama analisis sentimen adalah membedakan polaritas teks dalam dokumen, apakah pendapat yang dikemukakan dalam dokumen tersebut positif, negatif, atau netral. Penelitian dan analisis emosi berkembang sejak tahun 2003 dan merupakan bagian dari Text Mining yang merupakan analisis matematis berdasarkan emosi, emoticon, pikiran, ekspresi dan setiap kata yang diungkapkan melalui teks.

Analisis sentimen berfokus pada evaluasi klasifikasi berdasarkan polaritas. Berdasarkan klasifikasinya, analisis sentimen dibagi menjadi dua jenis utama. Yaitu pembagian dokumen menjadi ide atau fakta, atau dikenal dengan klasifikasi subjektivitas, dan klasifikasi dokumen menjadi positif atau negatif, atau dikenal dengan analisis sentimen. Ini adalah proses penting untuk memusatkan dokumen mana yang mempunyai opini dan dokumen mana yang menyimpulkan opini tersebut positif, negatif, atau netral.

Python

Python adalah bahasa pemrograman tingkat tinggi yang mendukung pemrograman berorientasi objek. Python berbeda dengan bahasa pemrograman lain terutama pada scriptingnya. Dalam bahasa pemrograman Python, terdapat berbagai library dan framework yang digunakan untuk analisis data.

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Python juga didukung oleh

komunitas yang besar.

Python merupakan salah satu dari bahasa pemrograman yang sering digunakan oleh programmer atau pembuat program dalam membuat program mereka. Python memiliki karakteristik sintaks yang tidak terlalu rumit. Dalam menulis sebuah kode program menggunakan bahasa pemrograman Python, terdapat beberapa aturan yang harus dipenuhi. Hal ini untuk mengantisipasi terjadinya error atau masalah pada program yang dibuat. Aturan sintaks Python yang pertama adalah dalam penulisan Statement atau perintah.

Text Mining

Text mining merupakan suatu pengetahuan yang didalamnya terkait dengan proses hubungan antara pengguna dengan dokumen-dokumen yang telah ada sebelumnya, menggunakan tool analisis (Feldman & Sanger, 2007). Dilakukannya Text Mining adalah untuk melakukan analisis dan menemukan pola-pola yang menarik dan relevan dari dokumen yang digunakan.

Text mining dan data mining di persepsikan dalam keilmuan yang sama di karenakan algoritma sama dapat digunakan dalam kedua konsep mining tersebut (Eroglu, 2018). Namun, keduanya berbeda dalam arti bahwa data mining melibatkan data terstruktur, sementara teks berkaitan dengan fitur-fitur tertentu dan relatif tidak terstruktur dan memerlukan preprocessing. Text mining adalah bidang yang saling terkait dengan Natural Language Processing (NLP).

Text mining adalah proses penggalian informasi dengan mengidentifikasi dan menganalisis pola-pola menarik dari sumber data berupa data teks tidak terstruktur. Pengolahan data dilakukan untuk mengubah data yang tidak terstruktur menjadi data terstruktur sehingga data tersebut siap untuk diolah. Preprocessing yang dilakukan adalah case folding, filtering, stemming, tokenizing kata. Preprocessing Case folding adalah tahap yang untuk mengubah huruf kapital menjadi huruf kecil. Selain itu, pada tahap juga ini dilakukan penghapusan tanda baca. Filtering adalah tahap membuang kata-kata yang tidak penting dalam dataset. Stemming adalah tahap menghilangkan imbuhan depan dan belakang menjadi kata dasar pada dataset. Tokenizing adalah tahap membagi kalimat menjadi kata per kata.

Pembobotan Term Frequency-Inverse Document Frequency (TF-IDF)

Term frequency (tf) merupakan sistem pembobotan yang mengukur frekuensi kemunculan istilah dalam dokumen. Semakin tinggi tf berarti term sering muncul maka dapat dianggap term umum sehingga tidak penting nilainya. Inverse Document Frequency (idf) merupakan frekuensi berbanding terbalik.

Analisis data teks memerlukan pembobotan pada setiap kata. Hal ini dikarenakan kata yang ada pada satu kalimat dengan kalimat lainnya bisa saja bermakna berbeda. Pembobotan kata pada data teks dapat dilakukan dengan TF-IDF (Term Frequency Inverse Document Frequency). TF-IDF adalah metode pengukuran statistik yang mengevaluasi seberapa relevan suatu kata dengan suatu dokumen dalam kumpulan dokumen (Riego and Villarba, 2023). Perhitungannya dilakukan dengan mengalikan 2 buah matriks yang berisi banyaknya suatu kata yang muncul pada suatu dokumen, serta invers frekuensi dokumen dari kata pada satu set dokumen. Fungsi dari TF-IDF sendiri yaitu untuk analisis teks secara otomatis, dan sangat berguna dalam pembobotan kata dalam algoritma machine learning untuk Natural Language Processing (NLP) (Kim and Gil, 2019). Persamaan TF-IDF (Christian et al., 2016).

Naïve Bayes Classifier

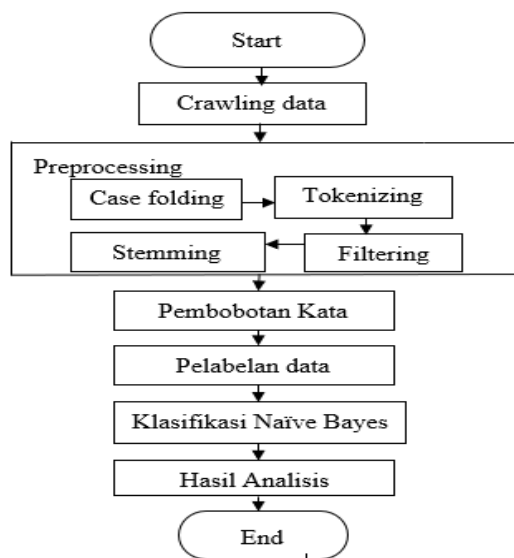
Naïve Bayes Classifier merupakan sebuah metode klasifikasi dengan probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan (independen) yang tinggi. Penggunaan metode Naive Bayes Classifier pada penelitian ini didasarkan pada banyaknya dataset yang dipakai sehingga membutuhkan suatu metode yang mempunyai performansi yang cepat dalam pengklasifikasian serta keakuratan yang cukup tinggi. Naive Bayes classifier adalah Suatu algoritma yang sering dipakai untuk keperluan data mining karena memudahkan penggunaannya, waktu pemrosesan yang cepat, mudah diterapkan dengan bentuknya yang cukup sederhana dengan tingkat efektifitas yang tinggi. Naïve Bayes Classifier merupakan salah satu metode machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Data yang digunakan pada penelitian ini merupakan data kualitatif, yang berupa sentimen opini masyarakat terhadap masa pemerintahan Anies Baswedan. Data yang dikumpulkan adalah data teks berupa tweets dari berbagai akun twitter. Banyaknya data yang digunakan sebanyak 1.563 tweets dari tahun Januari 2019 – Oktober 2022. Data twitter di filter, kemudian jumlah data hasil filter data dibagi dalam data training (70%) dan data testing (30%) data dipilah secara random. Kemudian data diberi label dengan dua kelas sentimen positif untuk opini yang bersifat negatif dan sentimen negatif untuk opini yang bersifat negatif, dengan jumlah data yang seimbang antara label positif dan label negatif.

Alur Penelitian



Gambar 1. Alur Penelitian

1. Crawling Data

Data yang diperoleh dari platform *twitter* dengan menggunakan teknik data *SNScrape* dan *tools* yang digunakan *python* pada *google colab* yang merupakan sebuah *library python* untuk mengumpulkan data *tweet* dari *twitter*. Data *Crawling* adalah proses otomatis untuk mengumpulkan dan mengindekskan data dari berbagai sumber seperti situs *web*, dataset, atau dokumen. Pada penelitian ini pengumpulan data dengan cara *crawling* data opini masyarakat di *twitter*. Pengumpulan data dengan cara menginputkan “#AniesBaswedan” sebagai kata kunci yang akan dicari. Untuk menganalisis sentimen terhadap kinerja Anies Baswedan selama masa jabatan sebagai Gubernur Daerah Khusus Ibu Kota Jakarta, dengan meninjau kembali program – program yang sudah diwujudkan oleh Anies Baswedan. Ada beberapa program yang dijadikan sampel penelitian ini, kata kunci program adalah “#JakLingko”, “#Tugusepatu”, “#bambugetahgetih”, “#formulaE”, “#Sumurserapan”, akan tetapi pada penelitian ini hanya menggunakan satu sampel untuk simulasi. Pengambilan data dari tahun Januari 2019 – Oktober 2022. Kemudian hasil *crawling* di simpan dalam format *ms. Excel* untuk dilakukan *pre-processing*. Pada tahap berikutnya dilakukan klasifikasi dengan perbandingan

2. Data Preprocessing

Pada tahap *pre-processing* data akan diproses menjadi lebih sistematis sebelum dilakukan proses klasifikasi. Tujuan dari *preprocessing* adalah untuk menyiapkan data agar digunakan dengan membersihkan kutipan yang dianggap tidak relevan. Tabel 1 tahap *Pre-processing*. Adapun tahapan-tahapan analisis yang dilakukan:

- I. *Case folding* adalah tahap yang untuk mengubah huruf kapital menjadi huruf kecil. Selain itu, pada tahap juga ini dilakukan penghapusan tanda baca.
- II. *Tokenizing* adalah tahap membagi kalimat menjadi kata per kata
- III. *Filtering* adalah tahap membuang kata - kata yang tidak penting dalam dataset.

IV. *Stemming* adalah tahap menghilangkan imbuhan depan dan belakang menjadi kata dasar pada dataset

3. Pembobotan Kata

Tahap pembobotan kata *TF-IDF*. Pembobotan kata dengan metode *TF*, jika dalam satu dataset muncul 2 kali teks maka akan diberi skor 2 akan tetapi jika dalam term hanya satu kali teks muncul maka akan diberi skor 1.

4. Pelabelan Data

Setelah tahap *tokenizing*, kemudian akan dilakukan pelabelan data karena hasil *tokenizing* merupakan data yang belum di beri label. Maka tahap ini dilakukan proses pelabelan data untuk menentukan kata – kata yang memiliki arti positif dan negatif. Selanjutnya akan dihitung seluruh jumlah kata positif dan negatif yang ada pada dataset. Klasifikasinya adalah sebagai berikut: label 1 jika jumlah kata positif melebihi kata negatif, label -1 jika jumlah kata negatif melebihi kata positif.

5. Klasifikasi Naïve Bayes

Naïve Bayes Classifier merupakan salah satu algoritma yang terdapat pada teknik klasifikasi data uji yang digunakan untuk mencari nilai probabilitas tertinggi pada kategori yang paling tepat, yang memanfaatkan perhitungan probabilitas dan statistik. Dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Tahap selanjutnya akan dilakukan klasifikasi menggunakan metode *Naïve Bayes*. Proses ini dilakukan untuk membuat cara klasifikasi yang nantinya akan digunakan untuk mengklasifikasikan data uji X agar mendapatkan label Y.

Secara umum proses yang dilakukan dalam klasifikasi naïve bayes adalah:

1. Mencari nilai peluang dari setiap kategori dataset sentimen positif dan sentimen negatif.
2. Mencari nilai peluang muncul pada masing-masing kategori dataset.
3. Menentukan kategori dataset berdasarkan perhitungan dari proses pertama dan kedua.
4. Perhitungan performa sistem.

Setelah mengolah data uji, kinerja sistem dihitung berdasarkan akurasi, presisi, *recall*, dan *f-measure* menggunakan *confusion matrix*. Pada tahap pembuatan sistem klasifikasi menggunakan *Naïve Bayes*, dilakukan dua langkah. Yang pertama adalah proses pelatihan data dan yang kedua adalah proses analisis data.

$$\text{Acc} = \frac{TN+TF}{FN+FP+TN+TP} \quad (1)$$

$$\text{Pre} = \frac{TP}{FP+TP} \quad (2)$$

$$\text{Rec} = \frac{TP}{FN+TP} \quad (3)$$

$$\text{F1} = 2 \frac{\text{pre} \cdot \text{rec}}{\text{pre}+\text{rec}} \quad (4)$$

a. Langkah-Langkah Pelatihan (*Training*) Langkah-langkah dalam proses pelatihan menggunakan NB adalah:

- 1) Menentukan model probabilitas dari data pelatihan.
- 2) Menghitung probabilitas kata (*TF-IDF*) dari data pelatihan
- 3) Menghitung probabilitas setiap kelas data pelatihan

$$P(w_k) = \frac{p^{k+1}}{nI+vacob} \quad (5)$$

Dimana:

P(w_k) : Probabilitas *term*

P_{k+1} : Query + 1 agar hasil tidak nol

n : jumlah kata yang muncul di dataset

|vocab| : total kata keseluruhan

b. Tahap *Testing* (Pengujian)

Selanjutnya tahap *testing* menggunakan NB adalah:

- 1) Menemukan Vmap untuk setiap data uji

$$Vmap = P(w_k|c). P(c) \quad (6)$$

- 2) Menentukan Vmap maksimum
- 3) Diketahui model klasifikasinya
- 4) Diperoleh kelas prediksi

6. Hasil Analisis

Setelah melakukan proses klasifikasi data, maka hasil analisis akan di sajikan dalam bentuk diagram batang sehingga dapat diketahui presentase berapa banyak opini positif dan opini negatif terhadap masa pemerintahan Anies Baswedan.

HASIL DAN PEMBAHASAN

Crawling Data

Data *crawling* merupakan langkah awal yang dilakukan peneliti untuk mengambil atau mengumpulkan data *tweet* dari server *Twitter* dengan menggunakan teknik data *SNScrape* dan *tools* yang digunakan *python* pada *google colab*. Pada tahap *crawling* data jumlah data *tweet* yang berhasil dikumpulkan sebanyak 1.563 rentang waktu dari Januari 2019 – Oktober 2022.

Preprocessing

Pemrosesan data adalah proses mengubah data mentah menjadi data bersih untuk proses klasifikasi. Tabel 2 tahap *Pre-processing*. Adapun tahapan-tahapa analisis yang dilakukan:

Tabel 2 Data Awal

	Hasil
Data awal	Selamat ulang tahun @pt_transjakarta! Apa kamu salah satu pengguna bus Transjakarta? Pengalaman unik apa yang pernah rasakan ketika menggunakan bus transjakarta? Yuk cerita di bawah ini! #DKIJakarta #Transjakarta #Jaklingko #NaikTJLebihOK #SmartMobility https://t.co/28ZUUhJRRE0 Anies Doakan Arifin Ilham Segera Kembali Berdakwah #KlikRMOL #aniesbaswedan https://t.co/FHI7aL6GJI Formula E musim 2018-2019 akan berlanjut Round 3 di Santiago Chile. Saksikan keseruan pertandingannya Minggu pukul 16.00 WIB via streaming di Vidio. Klik: https://t.co/XkUK20BpvW - #Sports #Racing #FormulaE https://t.co/ZMz3AcLC3g

Case Folding

Case folding adalah tahap yang untuk mengubah huruf kapital menjadi huruf kecil. Selain itu, pada tahap juga ini dilakukan penghapusan tanda baca. Dapat dilihat pada tabel 3 proses *case folding* di bawah:

Tabel 3 Proses *Case Folding*

Sebelum case folding	Sesudah Case Folding
Selamat ulang tahun @pt_transjakarta! Apa kamu salah satu pengguna bus Transjakarta? Pengalaman unik apa yang pernah rasakan ketika menggunakan bus transjakarta? Yuk cerita di bawah ini! #DKIJakarta #Transjakarta #Jaklingko #NaikTJLebihOK #SmartMobility https://t.co/28ZUUhJRRE0	selamat ulang tahun apa kamu salah satu pengguna bus transjakarta pengalaman unik apa yang pernah rasakan ketika menggunakan bus transjakarta yuk cerita bawah ini dkijakarta transjakarta jaklingko naiktjlebihok smartmobility
Anies Doakan Arifin Ilham Segera Kembali Berdakwah #KlikRMOL #aniesbaswedan https://t.co/FHI7aL6GJI	anies doakan arifin ilham segera kembali berdakwah klikrmol aniesbaswedan
Formula E musim 2018-2019 akan berlanjut Round 3 di Santiago Chile. Saksikan keseruan	formula musim akan berlanjut round santiago chile saksikan keseruan

<i>Sebelum case folding</i>	<i>Sesudah Case Folding</i>
pertandingannya Minggu pukul 16.00 WIB via streaming di Vidio. Klik: https://t.co/XkUK20BpvW - #Sports #Racing #FormulaE https://t.co/ZMz3AcLC3g	pertandingannya minggu pukul wib via streaming vidio klik sports racing formulae

Tokenizing

Tokenizing adalah tahap membagi kalimat menjadi kata per kata. Dapat dilihat pada tabel 4 proses Tokenizing dibawa:

Tabel 4 Proses *Tokenizing*

<i>Sebelum Tokenizing</i>	<i>Sesudah Tokenizing</i>
selamat ulang tahun apa kamu salah satu pengguna bus transjakarta pengalaman unik apa yang pernah rasakan ketika menggunakan bus transjakarta yuk cerita bawah ini dkijakarta transjakarta jaklingko naiktjlebihok smartmobility	['selamat', 'ulang', 'tahun', 'apa', 'kamu', 'salah', 'satu', 'pengguna', 'bus', 'transjakarta', 'pengalaman', 'unik', 'apa', 'yang', 'pernah', 'rasakan', 'ketika', 'menggunakan', 'bus', 'transjakarta', 'yuk', 'cerita', 'bawah', 'ini', 'dkijakarta', 'transjakarta', 'jaklingko', 'naiktjlebihok', 'smartmobility']
anies doakan arifin ilham segera kembali berdakwah klikrmol aniesbaswedan	['anies', 'doakan', 'arifin', 'ilham', 'segera', 'kembali', 'berdakwah', 'klikrmol', 'aniesbaswedan']
formula musim akan berlanjut round santiago chile saksikan keseruan pertandingannya minggu pukul wib via streaming vidio klik sports racing formulae	['formula', 'musim', 'akan', 'berlanjut', 'round', 'santiago', 'chile', 'saksikan', 'keseruan', 'pertandingannya', 'minggu', 'pukul', 'wib', 'via', 'streaming', 'vidio', 'klik', 'sports', 'racing', 'formulae']

Filtering

Filtering adalah tahap membuang kata - kata yang tidak penting dalam *dataset*. Dapat dilihat pada tabel 5 proses *filtering* dibawah:

Tabel 5 Proses *Filtering*

<i>Sebelum Filtering</i>	<i>Sesudah Filtering</i>
['selamat', 'ulang', 'tahun', 'apa', 'kamu', 'salah', 'satu', 'pengguna', 'bus', 'transjakarta', 'pengalaman', 'unik', 'apa', 'yang', 'pernah', 'rasakan', 'ketika', 'menggunakan', 'bus', 'transjakarta', 'yuk', 'cerita', 'bawah', 'ini', 'dkijakarta', 'transjakarta', 'jaklingko', 'naiktjlebihok', 'smartmobility']	['selamat', 'ulang', 'salah', 'pengguna', 'bus', 'transjakarta', 'pengalaman', 'unik', 'rasakan', 'bus', 'transjakarta', 'yuk', 'cerita', 'dkijakarta', 'transjakarta', 'jaklingko', 'naiktjlebihok', 'smartmobility']
['anies', 'doakan', 'arifin', 'ilham', 'segera', 'kembali', 'berdakwah', 'klikrmol', 'aniesbaswedan']	['anies', 'doakan', 'arifin', 'ilham', 'berdakwah', 'klikrmol', 'aniesbaswedan']
s['formula', 'musim', 'akan', 'berlanjut', 'round', 'santiago', 'chile', 'saksikan', 'keseruan', 'pertandingannya', 'minggu', 'pukul', 'wib', 'via', 'streaming', 'vidio', 'klik', 'sports', 'racing', 'formulae']	['formula', 'musim', 'berlanjut', 'round', 'santiago', 'chile', 'saksikan', 'keseruan', 'pertandingannya', 'minggu', 'wib', 'via', 'streaming', 'vidio', 'klik', 'sports', 'racing', 'formulae']

Stemming

Stemming adalah tahap menghilangkan imbuhan depan dan belakang menjadi kata dasar pada dataset. Dapat dilihat pada tabel 6 proses *Stemming* dibawah:

Tabel 6 Proses *Stemming*

<i>Sebelum Stemming</i>	<i>Sesudah Stemming</i>
['selamat', 'ulang', 'salah', 'pengguna', 'bus', 'transjakarta', 'pengalaman', 'unik', 'rasakan', 'bus', 'transjakarta', 'yuk', 'cerita', 'dkijakarta',	['selamat', 'ulang', 'salah', 'guna', 'bus', 'transjakarta', 'alam', 'unik', 'rasa', 'bus', 'transjakarta', 'yuk',

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
'transjakarta', 'jaklingko', 'naiktjlebohok', 'smartmobility']	'cerita', 'dkijakarta', 'transjakarta', 'jaklingko', 'naiktjlebohok', 'smartmobility']
['anies', 'doakan', 'arifin', 'ilham', 'berdakwah', 'klikrmol', 'aniesbaswedan']	['anies', 'doa', 'arifin', 'ilham', 'dakwah', 'klikrmol', 'aniesbaswedan']
['formula', 'musim', 'berlanjut', 'round', 'santiago', 'chile', 'saksikan', 'keseruan', 'pertandingannya', 'minggu', 'wib', 'via', 'streaming', 'vidio', 'klik', 'sports', 'racing', 'formulae']	['formula', 'musim', 'lanjut', 'round', 'santiago', 'chile', 'saksi', 'seru', 'tanding', 'minggu', 'wib', 'via', 'streaming', 'vidio', 'klik', 'sports', 'racing', 'formulae']

Pembobotan Kata

Tahap pembobotan kata *TF-IDF*. Pembobotan kata dengan metode *TF*, jika dalam satu dataset muncul 2 kali teks maka akan diberi skor 2 akan tetapi jika dalam *term* hanya satu kali teks muncul maka akan diberi skor 1.

Term 0 = ['kartu', 'jaklingko', 'klikrmol', 'jaklingko']

Term 1 = ['jak', 'lingko', 'kelola', 'join', 'venture', 'jaklingko']

Tabel 7 Hasil TF-IDF

Term	Kata	TF	IDF	TF-IDF
0	Kartu	0,25	0,80	0,2
	Jaklingko	0,5	0,50	0,25
	Klikrmol	0,25	1,53	0,4
1	Jak	0,16	0,550	0,08
	Lingko	0,16	0,60	0,09
	Kelola	0,16	0,90	0,18
	Join	0,16	1,04	0,00
	Venture	0,16	1,11	0,18
	Jaklingko	0,18	0,18	0,05

Pelabelan Data

Setelah melewati tahap *Stemming* maka akan dilanjutkan proses pelabelan data. Pelabelan data pada tahap ini akan dibagi menjadi 2 kelas yaitu kelas positif dan kelas negatif. Dapat dilihat pada tabel 8 Pelabelan Data dibawah:

Tabel 8 Pelabelan Data

Tweet	Label
saveaniesbaswedan guncang medsos guncang bangun istana negara lutut hun getar otak kepala didih aniesbaswedan erdogan umat rakyat 'jelata pasang badan beliau	Positif
sport formulae balap prix santiago tantang	Negatif
terima kasih pt transjakarta hubung hidup jakarta moga nadi utama sistem transportasi jakarta integrasi jaklingko harap teman teman transjakarta connectingthelifeofjakarta nai	Positif
gubernur jakarta pantau sebab banjir gubernur dki jakarta tinjau titik banjir kota salur air sumbat bangun lrt jabodetabek kontraktor lrt pompa permanen newsone viralnews jakarta banjirjakarta aniesbaswedan	Negatif
btw dasar bikin tugu sepatu jakarta pecak jakarta tutu sepatu uang bingung pakai tugusepatu jakarta	Negatif
sumuresapanggaran dki jakarta tugusepatu	Negatif
klaim tugusepatu bambumesum resap sulap kayumonas tugusepeda transportasikaki penghargaanwtp rumahsehatjakarta kalee	Negatif

Klasifikasi Naïve Bayes

Klasifikasi adalah langkah yang akan dilakukan dalam proses pengujian dataset yang sudah melewati beberapa tahap sebelumnya. Tahapan yang telah diselesaikan yaitu dari *crawling* data, *preprocessing*, pembobotan *TF-IDF*, pelabelan data. Kemudian langkah selanjutnya akan dibuatkan *Confusion matrix* yang digunakan untuk menguji tingkat akurasi, bukan hanya akurasi, akan tetapi juga *confusion matrix* dapat menguji presisi, *recall* dan *fmeasure*. Dalam *confusion matrix* 4 istilah dimana 10 diantaranya *True Positive*, *True Negative*, *False Positive* dan *False Negative*. Berikut adalah tabel *confusion matrix*nya dengan menggunakan metode *Naïve Bayes Classifier* perbandingan dataset 70:30. Dapat dilihat pada tabel 9 prediksi awal dibawah:

Tabel 9 Prediksi Aktual

Prediksi	Aktual	
	TP 1 120	FP 1 48
	TN 0 118	FN 0 27

Dari tabel diatas dapat dilihat *true positive* (TP) adalah data kategori positif yang berhasil diprediksi positif yaitu sebanyak 120 data. *True negative* (TN) adalah semua data kategori negatif yang berhasil prediksi negatif yaitu sebanyak 118 data. Adapun *false positive* (FP) yang artinya semua data yang kategori negatif tapi diprediksi positif yaitu sebanyak 48 data. Sedangkan *false negative* (FN) yang artinya semua data yang berkategori positif tapi diprediksi negatif yaitu sebanyak 27 data.

$$\text{Acc} = \frac{TP+TN}{FN+FP+TN+TP}$$

$$\text{Acc} = \frac{120+118}{27+48+118+120}$$

$$\text{Acc} = \frac{238}{313} \times 100\%$$

$$= 76,04 \%$$

$$\text{Pre} = \frac{TP}{FP+TP}$$

$$\text{Pre} = \frac{120}{48+120}$$

$$\text{Pre} = \frac{120}{168} \times 100\%$$

$$= 71,43\%$$

$$\text{Recc} = \frac{TP}{FN+TP}$$

$$\text{Recc} = \frac{120}{27+120}$$

$$\text{Recc} = \frac{120}{147} \times 100\%$$

$$= 81,63\%$$

$$F1 = 2 \frac{\text{pre} \times \text{rec}}{\text{pre} + \text{rec}}$$

$$F1 = 2 \frac{71,43 \times 81,63}{71,43 + 81,63}$$

$$F1 = 2 \frac{5830,84}{153,06}$$

$$= 76,19\%$$

Akurasi adalah banyaknya seluruh data yang diklasifikasikan menjadi tipe data positif atau negatif, kemudian dibagi dengan total seluruh data dan diperoleh hasil sebesar 76,04%. Presisi adalah jumlah data positif dibagi seluruh data yang tergolong positif, sehingga menghasilkan hasil sebesar 71,43%. *Recall* adalah banyaknya data yang tergolong baik, dibagi seluruh data yang tergolong baik, baik positif maupun salah, hasilnya 81,63%. Selanjutnya menghitung *f-measure*

dengan membagi hasil perkalian presisi dan recall dengan hasil penjumlahan presisi dan *recall*, lalu dikalikan 2, diperoleh hasil 76,19%. Kemudian dari hasil yang didapat dari jumlah data 746 teks hasil klasifikasi ke dalam sentimen positif dan 817 data hasil klasifikasi sentimen negatif. Jadi untuk persentase sentimen positif adalah sebagai berikut:

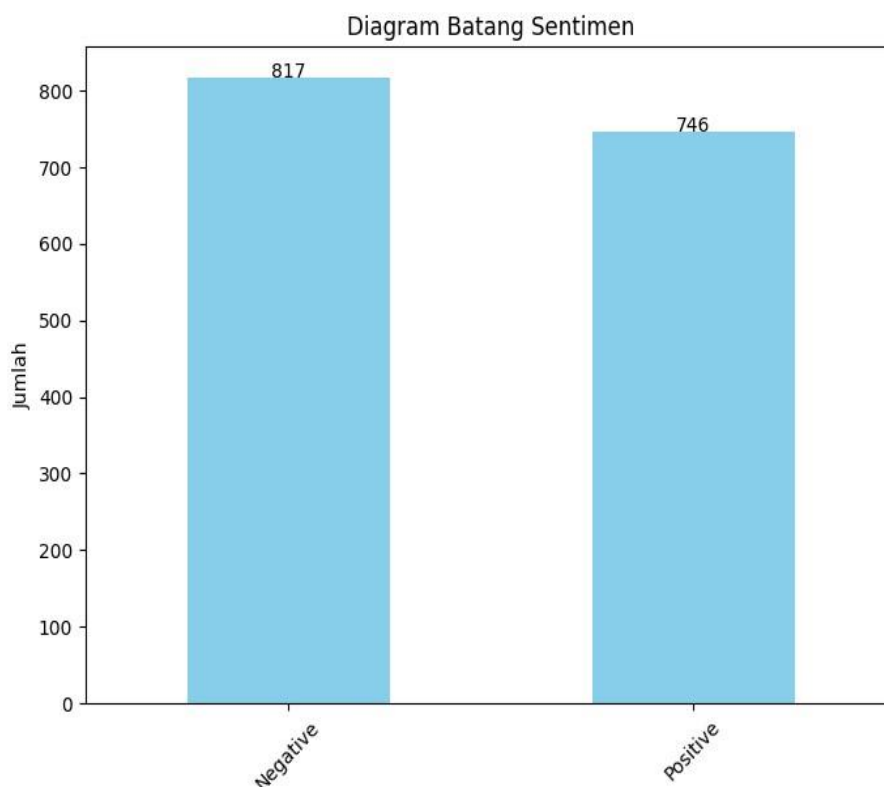
$$\frac{746}{1563} \times 100\% = 47,7\%$$

Untuk persentase kategori negatif adalah :

$$\frac{817}{1563} \times 100\% = 52,3\%$$

Hasil Analisis

Dari proses yang sudah dilakukan untuk mendapatkan hasil tujuan dari penelitian ini, maka diperoleh respon dari masyarakat selama masa pemerintahan Anies Baswedan sebagai Gubernur Daerah Khusus Ibu Kota Jakarta. Dimana hasil klasifikasi menunjukkan bahwa jumlah respon yang berkategori negatif lebih banyak dengan total data sebanyak 817 (52,3%), dan respon berkategori positif sebanyak 746 (47,7%). Maka hasil akhir dari penelitian ini dapat dilihat pada gambar diagram batang pada gambar 2 di bawah ini.



KESIMPULAN

Dari hasil penelitian ini didapatkan kesimpulan klasifikasi sistem menggunakan metode Naïve Bayes Classifier. Klasifikasi dibagi ke dalam dua sentimen kategori positif dan negatif pada kasus tanggapan masyarakat terhadap masa pemerintahan Anies Baswedan sebagai Gubernur Daerah Khusus Ibu Kota Jakarta, setelah proses *crawling data*, *preprocessing*, pelabelan data, maka dapat diketahui performa tingkat akurasi 76,04%, *presisi* 71,43, *recall* 81,63 % dan *f-measure* 76,19%. Data diperoleh dengan melakukan *crawling data* dari *Twitter* rentang waktu Januari 2019 – Oktober 2022 dengan jumlah data 1563 Tweet, dapat diketahui hasil tanggapan masyarakat kategori positif sebanyak 746 (47,7%). Akan tetapi respon masyarakat pada masa pemerintahan Anies Baswedan berkategori negatif lebih banyak dengan jumlah data sebanyak 817 (52,3%). Maka hasil akhir dari penelitian ini negatif.

REFERENSI

- Dianati Duei P., Gigih Forda N., Wahyu Eko S. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (Dpr) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan (Jitet)*, 2303-0577.
- Fauzi A., Akbar Fattullah, M., Asmawan, Andri, Ferdi Y. (2019). Sentimen Analisis Berinternet Pada Media Sosial Dengan Menggunakan Algoritma Bayes. *Jurnal Informatika*, 1(1), 77-83.
- Kusuma Sinta W., Sari Arum Y., Indriati. (2021). Analisis Sentimen menggunakan Metode Naive Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan Document Frequency Thresholding. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(12), 5582-5590.
- Melani A., R., Wasino., Tony. (2022). Analisis Sentimen Tanggapan Masyarakat Terhadap bantuan Sosial Pemerintah Di Masa Pandemi Covid-19 Pada Platform Twitter. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 10(1).
- Mohammed Hafizh Al-A., Kana Saputra S. (2023). Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM. *Jurnal Saintikom (Jurnal Sains Manajemen Informatika Dan Komputer)*, 22(2) 70-279..
- Pamungkas Sodik F., Kharisudin I. (2021). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. *E- Prosiding Seminar Nasional Matematika*, 4, 628-634.
- Putu Pasek O., M., Gusti Arya S., I Putu Agus E., P., (2022). Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap “Figure Pemimpin” Menggunakan Python. *Jitter- Jurnal Ilmiah Teknologi Dan Komputer*, 3, 810-820.
- Shania Gabriela K., Daniel R. K., Vivi P. R. (2021). Analisis Sentimen New Normal Pada Masa Covid-19 Menggunakan Algoritma Naive Bayes. *Journal Of Informatics Engineering*.
- Sulaeman Riva K., Setianingsih C., Saputra Erfa R. (2022). Analisis Algoritma Support Vector Machine Dalam Klasifikasi Penyakit Stroke Support Vector Machine Algorithm Analysis In Stroke Disease Classification. *E-Proceeding Of Engineering*, 9(3), 922.
- Zulqornain Alfiah J., Indriati., Adikara Pandu P. (2021). Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naive Bayes dan Categorical Proportional Difference (CPD). *Jurnal Pengembangan*.