



ANALISIS SENTIMEN TERHADAP CHATGPT PLUS MENGGUNAKAN NAÏVE BAYES DI TWITTER

Putri Yani¹⁾ Habi Baturohmah²⁾

^{1, 2)} Program Studi Sistem Informasi, Universitas Nusa Putra

Jln Raya Cibatu Cisaat No.21 Cibolang Kaler, Kec Cisaat Sukabumi Regency, Jawa Barat

¹⁾putriyani_si20@nusaputra.ac.id ²⁾habi.baturohmah@nusaputra.ac.id

* Korespondensi: e-mail: putriyani_si20@nusaputra.ac.id

ABSTRAK

Dalam era digital yang terus berkembang, media sosial telah menjadi pusat interaksi dan pertukaran informasi antara individu di seluruh dunia. Salah satu platform media sosial yang paling populer adalah Twitter, di mana pengguna dapat berbagi pikiran, pendapat, dan perasaan mereka dalam bentuk tweet yang singkat banyaknya opini masyarakat berpendapat yang bersifat positif, negatatif dan netral adapun Penelitian ini bertujuan untuk melakukan analisis sentimen pada data Twitter menggunakan carwling twitter dengan kata kunci "ChatGPT Plus" dan "chatGPT-4" dari OpenAI menggunakan metode klasifikasi Naive Bayes. Untuk mencari seberapa akurasi impelementasi naive bayes. Hasil dari penelitian yang dilakukan terdapat Accuracy yang di hasilkan 84.60% sedagkan untuk precisionnya untuk positif 85.62%, negatig 95.43%, netral 62.77% dan untuk recallnya 94.24% positif, 76.61%, negatif, 90.77% negati

Kata Kunci: chatgpt plu, naiv bayes, twitter.

ABSTRACT

In the ever-evolving digital era, social media has become a center for interaction and exchange of information between individuals around the world. One of the most popular social media platforms is Twitter, where users can share their thoughts, opinions and feelings in the form of short tweets. Many public opinions are positive, negative and neutral. This research aims to analyze sentiment on Twitter data. using Twitter carwling with the keywords "ChatGPT Plus" and "chatGPT-4" from OpenAI using the Naive Bayes classification method. To find how accurate is the implementation of Naive Bayes. The results of the research conducted showed that the accuracy was 84.60%, while the precision was 85.62% positive, 95.43% negative, 62.77% neutral and 94.24% positive, 76.61%, negative, 90.77% negative.

Keywords: chatgpt plu, naiv bayes, twitter.

I. Pendahuluan

Dalam era digital yang terus berkembang, media sosial telah menjadi pusat interaksi dan pertukaran informasi antara individu di seluruh dunia. Salah satu platform media sosial yang paling populer adalah Twitter, di mana pengguna dapat berbagi pikiran, pendapat, dan perasaan mereka dalam bentuk tweet yang singkat. Pengguna internet di Indonesia pada bulan maret 2021 mencapai 76,8 persen yang berarti sebanyak 212,35 juta dari total populasi sebanyak 276,2 juta jiwa. Selain itu, Indonesia merupakan urutan ke-15 di Asia pada tahun 2021 dengan penetrasi internet tertinggi. Salah satu aplikasi media sosial yang populer di Indonesia yaitu twitter.[1] Di Indonesia, twitter tidak hanya digunakan sebagai media untuk mengungkapkan perasaan tetapi juga sebagai media dalam memberikan saran maupun keritik terhadap suatu objek, layanan atau kebijakan pemerintah. Pengguna twitter juga memberikan opini mereka terhadap suatu kebijakan dan dijadikan trending topic.[2]

Dalam konteks ini, ChatGPT plus telah menjadi salah satu topik yang menarik perhatian di kalangan pengguna Twitter. ChatGPT plus merupakan versi terbaru dari model generatif bahasa yang dikembangkan oleh OpenAI. sebelumnya juga open AI telah mengeluarkan OpenAI Gym, Universe, cahatgpt-3 chatgpt-3,5, OpenAI memperkenalkan DALL-E, memperkenalkan Generative Pre-trained. Transformer 4 (GPT-4), Dengan kemampuannya dalam memahami konteks dan menghasilkan teks yang lebih manusiawi ChatGPT Plus telah mempercepat interaksi manusia-mesin di platform seperti Twitter. Layanan berlangganan dua



puluh dolar per bulan ini memungkinkan pengguna mengakses ChatGPT selama jam sibuk, memberikan waktu respons yang lebih cepat, dan memberikan akses awal ke fitur baru. Analisis sentimen, sebagai cabang penting dari pemrosesan bahasa alami, bertujuan untuk memahami dan mengekstrak emosi atau pendapat yang terkandung dalam teks. Dalam konteks ini, pendekatan menggunakan algoritma Naive Bayes telah menjadi salah satu metode yang populer dalam melakukan analisis sentimen. Algoritma ini mengasumsikan independensi fitur dalam teks dan menghitung probabilitas terjadinya sentimen tertentu berdasarkan kemunculan kata-kata kunci.[3]

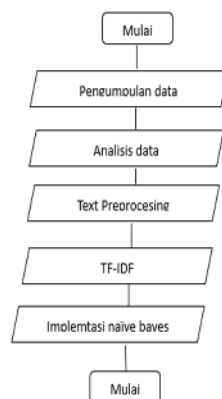
Algoritma Naive Bayes, yang diciptakan oleh ilmuwan Inggris Thomas Bayes, menggunakan teknik probabilitas dan statistik untuk menyaring spam. Dasar dari pengklasifikasi Naive Bayes, digunakan untuk memprediksi peluang di masa depan berdasarkan pengalaman masa lalu. Pengklasifikasi, juga dikenal sebagai pengklasifikasi, adalah kelompok algoritma klasifikasi yang bekerja berdasarkan gagasan bahwa setiap fitur yang diklasifikasikan tidak bergantung satu sama lain. Salah satu ciri utama klasifikator

Naive Bayes ini merupakan keyakinan yang kuat (naif) bahwa setiap kondisi atau kejadian tidak bergantung pada yang lain. Studi sebelumnya tentang analisis sentimen menurut penelitian yang dilakukan amaia N, supraptiT, dan dwilestari G, penelitian tersebut menyatakan bahwa Hasil Menurut Dalam Algoritma Decision Tree memiliki nilai akurasi 68,19%, recall 96,33%, dan ketepatan 37,83%, sedangkan algoritma Naive Bayes memiliki nilai akurasi 81,15%, recall 75,98%, dan ketepatan 91,10% dalam penelitian ini.[2] Menurut hasil pengujian model algoritma Naive Bayes oleh irvandi, irwan B, nurdiwan o, memiliki tingkat akurasi sebesar 74,75%.[4] dan menurut penelitian yang dilakukan oleh halim anshor dan safuwana tentang prediksi virus covid-19 dengan naïve bayes dengan Uji Nave Bayes + Particle Swarm Optimization menunjukkan akurasi 88,01%, sedangkan uji analisis sentimen Naive Bayes menunjukkan 67,11%. dengan peningkatan kinerja AUC (0,26).[5] Selain itu penelitian dengan metode naïve bayes juga digunakan untuk menganalisis penilaian aplikasi brimo hasil klasifikasi. Metode ini menunjukkan akurasi sebesar 84,52%. [6]

Berdasarkan uraian di atas, peneliti akan menganalisis sentimen opini publik yang netral, positif, dan negatif dengan menggunakan algoritma naïve bayes classifier dan data tweet chatgpt plus. Untuk menentukan tingkat akurasi prosedur dan prosentase masing-masing emosi yang dapat digunakan untuk meningkatkan kualitas openAI pada fitur chatgpt plus yang baru dirilis, tujuan penelitian ini adalah untuk menghasilkan hasil ini.

II. METODE

Penelitian ini dilakukan dengan menggunakan metode naïve bayes adapun rangkain dan langkah-langkahnya sebagai berikut:

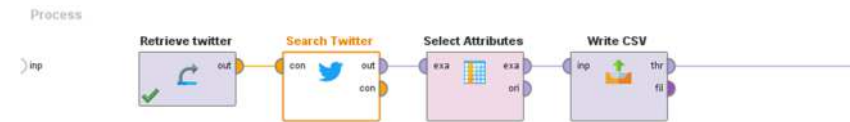


Gambara 2.1 Flowchart Penelitian



1. Pengumpulan data

Pengumpulan data yang dilakukan untuk riset ini dari berbagai sumber yang pertama dengan melakukan carweling pada twitter dengan bantuan menggunakan rapidminer dengan versi 10.0 menghasilkan 500 data lebih, selama 7 hari yang akan kemudian di sortir kembali, dan melakukan literatur riviweu pada jurnal-jurnal sebelumnya untuk mendapatkan metode yang sesuai dengan yang dibutuhkan, serta dari berbagai sumber lain seperti internet mengenai chatgpt plus



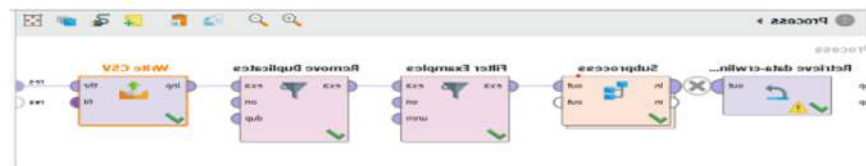
Gambar 2.2 Proses Crawling Twitter

	Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text	Geo-Locatio...
965333425...	Jun 4, 2023, ...	virus girfriend	1586611567...	?	-1	in	<a href="http://...	Aku ada...	?
965253387...	Jun 4, 2023, ...	Westphoenzl...	50332464	?	-1	in	<a href="http://...	Kun #S...	?
965110672...	Jun 4, 2023, ...	Hamiddi	1059972402	AmanullMu_min	220861088	in	<a href="http://...	@Amiru...	?
964882331...	Jun 3, 2023, ...	luturpremium1	1563376817...	?	-1	in	<a href="http://...	Aku ada...	?
964629091...	Jun 2, 2023, ...	hansil ipi dia...	1367754563...	?	-1	in	<a href="http://...	RT @zof...	?
964628831...	Jun 2, 2023, ...	jenny	1087910559...	?	-1	in	<a href="http://...	makin L...	?
964571132...	Jun 2, 2023, ...	x86/gorgon.exe	711939499	naufab85	711939499	in	<a href="http://...	@Hamk...	?
964483379...	Jun 2, 2023, ...	Algi G.Dharm...	317717924	?	-1	in	<a href="http://...	Kepikira...	?
964224930...	Jun 1, 2023, ...	mijuu.open ...	1374511477...	@KissMajaHa...	1572888014...	in	<a href="http://...	@Kiss...	?
964224786...	Jun 1, 2023, ...	mijuu.open ...	1374511477...	rainboububbl...	1137560449...	in	<a href="http://...	@rainb...	?
964224746...	Jun 1, 2023, ...	mijuu.open ...	1374511477...	adimasw12	7463821240...	in	<a href="http://...	@adim...	?
964224718...	Jun 1, 2023, ...	mijuu.open ...	1374511477...	munstale	1699116990	in	<a href="http://...	@muns...	?
964224543...	Jun 1, 2023, ...	mijuu.open ...	1374511477...	droleo	1366173501...	in	<a href="http://...	@drole...	?
964224463...	Jun 1, 2023, ...	mijuu.open ...	1374511477...	caramoel	1377163127...	in	<a href="http://...	@cara...	?

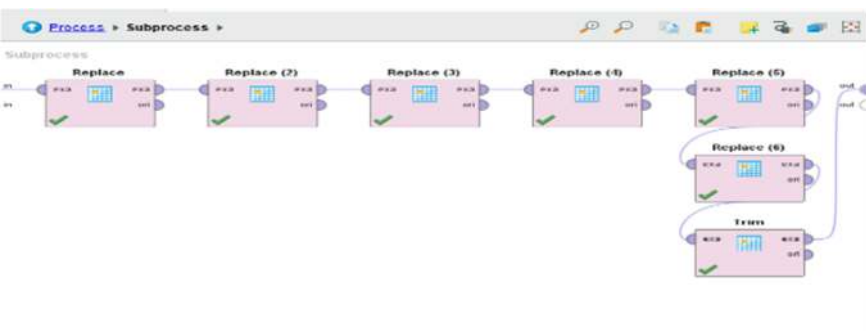
Gambar 2.3 hasil Proses Crawling Twitter

2. Analisis data

Pada tahapan ini data yang sudah di dapat data dari carwling twitter tersebut di seleksi lagi hanya beberapa artribut yang di gunakan atau yang di perlukan dan kata-kata yang berulang2, berikut tahapan untuk menganalisis data tersebut menggunakan rapidminer



Gambar 2.4 Proses Filtering



Gambar 2.5 Proses subprocesses



3. Text preprocessing

Dataset yang dihasilkan dari proses text mining pada dasarnya memiliki struktur yang tidak beraturan dan tidak teratur, jadi sebelum data dimasukkan ke dalam model, tahap preprocessing dilakukan untuk membersihkan noise dan mengubah data menjadi data yang terstruktur.

1. Tokenisasi: Memecah kumpulan karakter atau kalimat menjadi potongan kata atau karakter sesuai kebutuhan.
2. Transform Cases: Proses transformasi case (to Lowercase) digunakan pada tahap preprocessing untuk mengubah teks menjadi huruf kecil semua.
3. Filter Tokens (By Length): adalah proses menghapus kata dengan jumlah huruf tertentu dengan menggunakan fungsi bersih satu karakter atau lebih dengan parameter min chars 4 dan max chars 25, yang membatasi jumlah huruf dalam teks menjadi minimal 4 dan maksimal 25 huruf.
4. Filter Stopword: Istilah ini digunakan untuk kalimat yang mengandung kata-kata yang sering keluar dan dianggap tidak penting, seperti waktu atau penghubung.
5. Stemming: Menghapus semua imbuhan adalah proses menemukan kata dasar (awalan, sisipan, akhiran, dan confixes) dari kata turunan, termasuk kombinasi awalan dan akhiran. Stemming adalah metode untuk mengubah kata dari bentuknya menjadi kata dasar yang memiliki struktur morfologi yang tepat dalam Bahasa Indonesia.[7]



Gambar 2.6 proses text-preprocessing

4. TF-IDF

Setelah tahap preprocessing selesai, metode TF-IDF digunakan untuk menghitung seberapa signifikan kata-kata dalam dokumen dan korpus. Metode ini menentukan seberapa penting setiap kata atau bobot kata pada dokumen terhadap seluruh korpusnya. Tingkat frekuensi-balik dokumen (TF-IDF) Ada kemungkinan untuk menghitungnya dengan menggunakan rumus berikut: [1] $TF-IDF = TF * IDF = TF * \log(n/df)$, di mana:

Teks frekuensi = TF

Teks frekuensi dokumen dikodekan dengan IDF.

n adalah jumlah dokumen

5. Metode Naive Bayes

Kalkulasi naif Bayes merupakan algoritma yang mudah digunakan dan sederhana yang menganalisis sentimen dengan menggunakan konsep probabilitas atau peluang dalam klasifikasi. Selain itu, algoritma ini dapat dengan akurat memperkirakan suatu kejadian berdasarkan hasil klasifikasi [7]. Persamaan probabilitas metode Naive Bayes adalah sebagai berikut:.[8]

$$P(X|Y) = \frac{P(X|Y) \cdot P(X)}{1P(Y)}$$

Di sini, X adalah asumsi sementara data kelas tertentu, dan Y adalah data kelas yang belum diketahui. Peluang taksiran X adalah p(X) dengan syarat Y adalah P(X|Y).

P(Y|X) adalah peluang taksiran Y dengan X.



Potensi $Y = P(Y)$.

Keterangan:

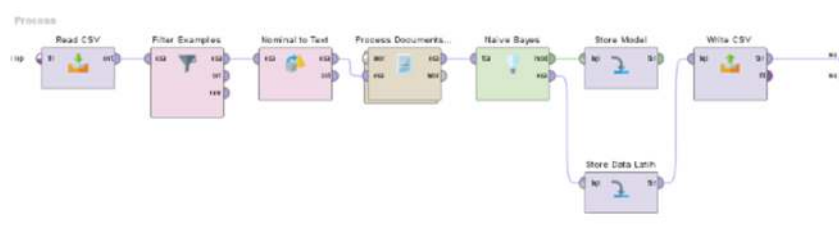
Potensi belakang: kemungkinan bahwa kelas X ada

Ada kemungkinan besar bahwa sampel awal kelas Y akan terjadi.[8]

III. HASIL DAN PEMBAHASAN

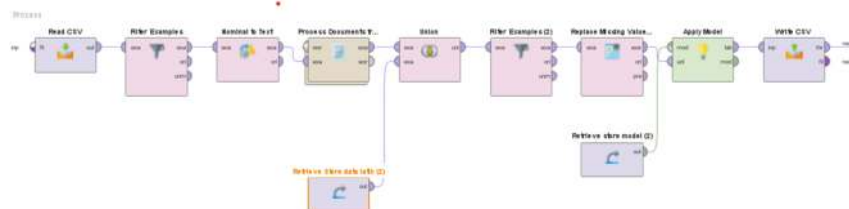
Darai hasil penelitan sentimen analisis untuk mendapatkan keaccruracy opini tentang chatgpt plus di twitter menggunakan metode naïve bayes yang sebelum nya sudah di jelakan dan sudah melewati beberapa tahapan pengumpulan data dan tex-preprocessing dari 689 menjadi 422 data yang akan dilakuakn pelabelen.

Setehal melakukan pelabelan pada cuitan data, tahapan selanjutnya yaitu membuat data latih untuk mentukan data latih yang akan di gunakan naïve bayes adapun untuk tahapan proses pembuatan data latih menggunakan bantuan tools rapidminer seperti yang ditunjukkan pada gambar berikut:



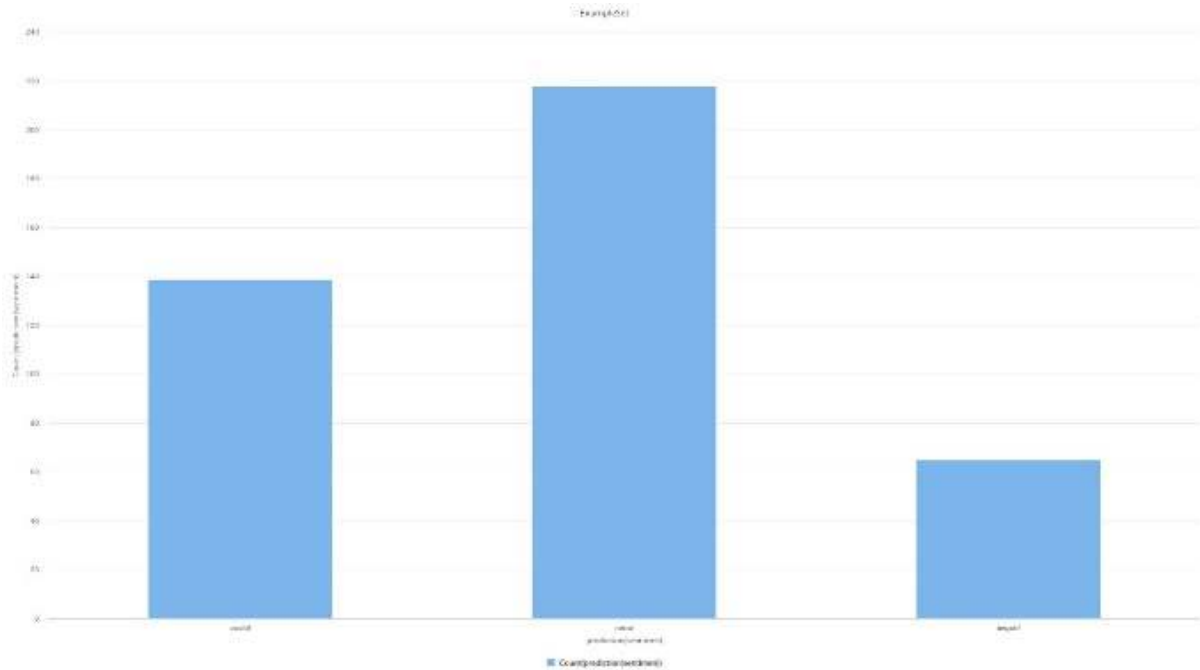
Gambar 3.1 Proses data latih

Setelah tahapan pemerosesan data latih di maka lanjutkan untuk pemerosan data uji untuk mendapatkan ke accuracy sebelum pada tahapan naïve bayes adapun proses tahapan data uji sebagai berikut:



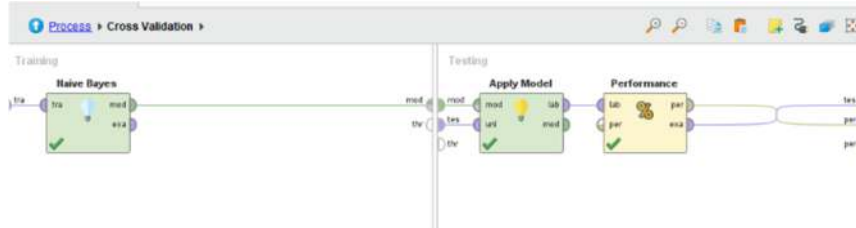
Gambar3.2 Proses data uji

Data latih dan uji sudah di dapatkan makan selanjutnya ke tahapan impelentasi naïve bayes dan cross validasion untuk mendapatkan hasil yang di inginkan berikut merukan diagram hasil dari data uji.



Gambar 3.3 Prediksi sentimen

Prediksi sentimen analisis setelah melakukan data uji sentimen yang bersifat netral sebesar 218 untuk fraction 0.517, beraifat positif 139 untuk fraction 0.329, bersifat negatif 65 untuk fraction 0.154



Gambar 3.4 Pemerosesan metode naïve bayes

Ada pun hasil dari implementasi naïve bayes ker accuracy dengan bantuan tools rapidminer sebagai berikut

accuracy: 84.60%

	true positif	true netral	true negatif	class precision
pred. positif	131	19	3	85.62%
pred. netral	5	167	3	95.43%
pred. negatif	3	32	59	62.77%
class recall	94.24%	76.61%	90.77%	

Gambar 3.5 Hasil accuracy

Accuracy yang di dihasilkan 84.60% sedangkan untuk precisionnya untuk positif 85.62%, negatif 95.43%, netral 62.77% dan untuk recallnya 94.24% positif, 76.61%, negatif, 90.77% negati

IV. KESIMPULAN

Jadi dapat di simpulkan dari hasil sentimen analisis chatGPT plus menggunakan metode naivi bayes ini menghasil kan accuracy 84.60% sedangkan untuk precisionnya untuk positif 85.62%, negatif 95.43%, netral 62.77% dan untuk recallnya 94.24% positif, 76.61%, negatif, 90.77% negatif/ untuk sentimen terbesar terhadap opini chatgpt plsu di tuwitter berssifat netral



Daftar Pustaka

- [1] E. Febriyani dan H. Februariyanti, “Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter,” vol. 17, no. 1.
- [2] N. Amalia *dkk.*, “ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAKSANAAN KURIKULUM MBKM,” vol. 18, no. 1, 2023.
- [3] P. F. Supriyadi dan Y. Sibaroni, “Xiaomi Smartphone Sentiment Analysis on Twitter Social Media Using IndoBERT,” *Jurnal Riset Komputer*, vol. 10, no. 1, hlm. 2407–389, 2023, doi: 10.30865/jurikom.v10i1.5540.
- [4] Irvandi, B. Irawan, dan O. Nurdiawan, “NAIVE BAYES DAN WORDCLOUD UNTUK ANALISIS SENTIMEN WISATA HALAL PULAU LOMBOK,” *INFOTECH journal*, vol. 9, no. 1, hlm. 236–242, Mei 2023, doi: 10.31949/infotech.v9i1.5322.
- [5] A. Halim Anshor dan A. Safuwani, “ANALISIS SENTIMEN OPINI WARGANET TWITTER TERHADAP TES SCREENING GENOSE PENDETEKSI VIRUS COVID-19 MENGGUNAKAN METODE NAÏVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION,” 2023.
- [6] M. Khoirul, U. Hayati, dan O. Nurdiawan, “ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES,” 2023.
- [7] I. Tri Julianto, D. Kurniadi, Y. Septiana, dan A. Sutedi, “Alternative Text Pre-Processing using Chat GPT Open AI,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 12, no. 1, hlm. 67–77, Mar 2023, doi: 10.23887/janapati.v12i1.59746.
- [8] D. Darwis, N. Siskawati, dan Z. Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” vol. 15, no. 1.