

# Klasifikasi Sentimen Masyarakat di Twitter terhadap Ganjar Pranowo dengan Metode Naïve Bayes Classifier

Sinta Wahyuni Ritonga\*, Yusra, Muhammad Fikry, Eka Pandu Cynthia

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia  
Jl. H.R. Soebrantas no. 155 KM. 18 Simpang Baru, Pekanbaru, 28293

Email: <sup>1\*</sup>11950121741@students.uin-suska.ac.id, <sup>2</sup>yusra@uin-suska.ac.id, <sup>3</sup>muhammad.fikry@uin-suska.ac.id,  
<sup>4</sup>eka.pandu.cynthia@uin-suska.ac.id

Email Penulis Korespondensi: 11950121741@students.uin-suska.ac.id

Submitted: 29/05/2023; Accepted: 27/06/2023; Published: 29/06/2023

**Abstrak**—Indonesia adalah negara dengan sistem politik demokrasi. Masyarakat diberikan kebebasan bersuara, berkolaborasi dan memberikan kritik secara publik. Pada era modern pemanfaatan media sosial sangat berkembang pesat di lapisan masyarakat. Trend sosial media di Indonesia salah satunya Twitter yang dimanfaatkan untuk menyampaikan aspirasi kepada pemerintah dan sebagai sarana untuk menyampaikan kegiatan sehari-hari, pendapat, budaya serta mendapatkan informasi atau berita terbaru Indonesia maupun luar negeri. Opini publik yang diambil dari Twitter dapat bersifat positif, negatif dan netral. Banyaknya tweet pada Twitter salah satu topik trend di Indonesia adalah Ganjar Pranowo, dapat digunakan sebagai sumber data dalam pengkajian klasifikasi sentimen yang diproses untuk menghasilkan nilai akurasi. Penelitian ini bertujuan untuk mengklasifikasikan opini publik di media sosial Twitter tentang Ganjar Pranowo dengan menggunakan metode Naïve Bayes Classifier. Dalam pemrosesan klasifikasi menggunakan dataset 4000 data tweet dengan dua kelas labeling, positif dan negatif untuk mengetahui efisiensi kinerja NBC yang dikombinasikan dengan pembobotan TF-IDF, feature selection menggunakan teknik pendekatan supervised learning. Hasil dari pengujian pada penelitian klasifikasi sentimen masyarakat di Twitter terhadap Ganjar Pranowo dengan menggunakan metode NBC menggunakan 10% data uji dari dataset yang digunakan sehingga menghasilkan nilai akurasi 83.0%.

**Kata kunci:** Ganjar Pranowo; Klasifikasi sentimen; Naïve Bayes Classifier; Preprocessing; Twitter

**Abstract**—Indonesia is a country with a Democratic political system. The public is given freedom of speech, collaboration and public criticism. In the modern era, the use of social media is growing rapidly at the community level. One of the social media trends in Indonesia is Twitter which is used to convey aspirations to the government and as a means to convey daily activities, opinions, culture and get the latest information or news from Indonesia and abroad. Public opinion taken from Twitter can be positive, negative and neutral. The number of tweets on Twitter one of the trend topics in Indonesia is Ganjar Pranowo, can be used as a source of data in the assessment of sentiment classification which is processed to produce accuracy values. This study aims to classify public opinion on social media Twitter about Ganjar Pranowo using Naïve Bayes Classifier method. In the classification processing using a dataset of 4000 tweet data with two labeling classes, positive and negative to determine the efficiency of NBC performance combined with TF-IDF weighting, feature selection using supervised learning approach techniques. The results of the test on the classification of public sentiment research on Twitter about Ganjar Pranowo using NBC method using 10% of the test data from the dataset used to produce an accuracy value of 83.0%.

**Keywords:** Ganjar Pranowo; Sentiment classification; Naïve Bayes Classifier; Preprocessing; Twitter

## 1. PENDAHULUAN

Indonesia adalah negara dengan sistem politik demokrasi. Masyarakat diberikan kebebasan bersuara, berkolaborasi dan memberikan kritik secara publik. Pada era modern pemanfaatan media sosial sangat berkembang pesat di lapisan masyarakat bahkan pejabat pemerintah pun memanfaatkan perkembangan teknologi. *Trend* sosial media di Indonesia salah satunya Twitter dan dimanfaatkan untuk menyampaikan gagasan mereka pada kegiatan sehari-hari, ekonomi, budaya, politik dan aspirasi kepada pemerintah [1].

Pemerintah merupakan suatu lembaga yang sudah memanfaatkan sosial media untuk mendapat masukan atau gagasan dari masyarakat. Twitter adalah publik prasarana online dan *microblogging* yang dapat dimanfaatkan konsumen bertukar pesan teks sampai 140 karakter, yang biasa disebut sebagai *tweet*. Aplikasi Twitter dirancang oleh Jack Dorsey tepatnya bulan Maret 2006 dan situs media sosial yang dirilis bulan Juli. Semenjak permulaan, *platform* Twitter merupakan salah satu media yang telah berkembang dari situs-situs lain yang lebih sering diakses di internet [2]. Sejauh ini pada tahun 2023 terdapat 368,4 juta akun Twitter yang aktif dari 1,3 miliar pengguna. Oleh karena itu, hal ini mendukung penelitian untuk memanfaatkan media sosial Twitter sebagai sumber data untuk dianalisis dalam bentuk klasifikasi sentimen dengan proses pemahaman, penggalian dan pengolahan data untuk memperoleh informasi opini publik yang tercantum dalam kalimat opini [3].

Opini adalah pendapat dan pola pikir seseorang dalam merespon suatu permasalahan atau pembahasan di mana pada sebuah opini terdapat positif dan negatif. Akhir-akhir ini salah satu topik *trend* Twitter di Indonesia adalah sosok Ganjar Pranowo. H. Ganjar Pranowo, S.H, M.I.P lahir di Karanganyar, 28 Oktober 1968, Jawa Tengah. Ganjar Pranowo merupakan seorang Gubernur Jawa Tengah yang dikenal sikap kepemimpinannya cerdas dan tegas. Ganjar salah satu tamatan Universitas Gajah Mada tepatnya Fakultas Hukum dan Pascasarjana Ilmu Politik dari Universitas Indonesia. Profesi seorang Ganjar dalam bidang politik bermula pada tahun 2009 sampai saat ini. Rekam jejaknya cukup memuaskan sehingga beliau banyak disenangi oleh masyarakat karena beliau merupakan salah satu pejabat pemerintah yang merakyat. Pengalaman seorang Ganjar dalam dunia politik bisa dilihat sudah sangat matang [4]. Dari

seiring banyaknya *tweet* tentang Ganjar Pranowo, pengguna Twitter atau masyarakat dapat memberikan komentar melalui *tweet* yang mana berisi persepsi tentang Ganjar Pranowo secara publik. Tanggapan yang dibagikan dapat berupa opini positif dan opini negatif [5]. Tidak salah jika terdapat beberapa pendapat dari kalangan masyarakat umumnya terutama yang berkaitan dengan pemerintahan. Dan dari beberapa perbedaan pendapat bisa disimpulkan dan menghasilkan sebuah solusi dari permasalahan. Ekstraksi data ini memerlukan teknik penyelidikan yang tepat untuk mendapatkan informasi yang bermanfaat bagi berbagai pihak yaitu pemanfaatan klasifikasi sentimen [6].

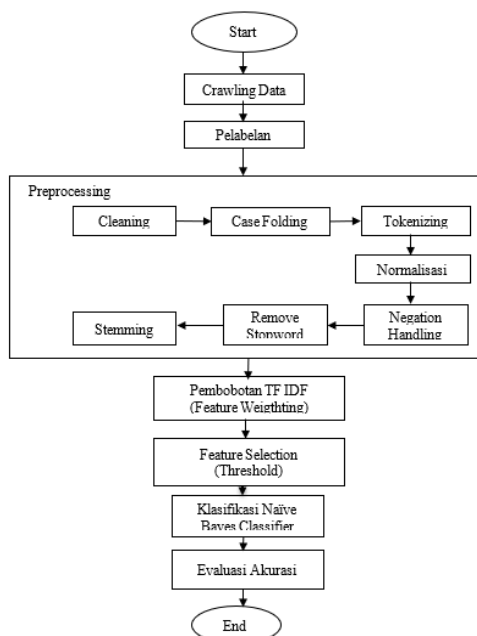
Studi terdahulu terkait klasifikasi sentimen di antaranya: penelitian yang dilakukan oleh Kararisma, Lestanti, dan Chulkamdi tentang aplikasi klasifikasi sentimen pada ulasan *smartphone* di situs jual beli online berbasis web menggunakan *Naïve Bayes* dengan TF-IDF diperoleh hasil 93% [7]. Zhafira, Rahayudi, dan Indriati melakukan penelitian analisis sentimen kebijakan kampus merdeka menggunakan *Naïve Bayes* dan pembobotan TF-IDF berdasarkan komentar pada *youtube* dengan hasil akurasi sebesar 91.8% [8]. Hasri dan Alita melakukan penelitian penerapan metode *Naïve Bayes Classifier* dan *Support Vector Machine* pada analisis sentimen terhadap dampak virus Covid'19 di Twitter dengan hasil akurasi NBC 81.07% dan SVM 79.96%, di mana nilai akurasi metode *Naïve Bayes Classifier* lebih tinggi [9].

Dari beberapa penelitian terdahulu metode *Naïve Bayes Classifier* memperoleh nilai akurasi yang tinggi. Penelitian ini menyesuaikan metodologi yang diterapkan dalam penelitian sebelumnya. dengan metode *Naive Bayes* dalam uji klasifikasi. Pro dan kontra yang ada di masyarakat khususnya kalangan pengguna Twitter terhadap Ganjar Pranowo dapat diketahui melalui klasifikasi sentimen dengan menerapkan metode *Naïve Bayes Classifier* [10]. Teknik pendekatan *supervised learning* diaplikasikan dalam penelitian ini. *Supervised learning* atau teknik pembelajaran adalah teknik yang mengklasifikasi setiap objek dalam data menjadi beberapa kelas dalam penelitian Dougherty. Dalam *supervised learning*, setiap objek dalam data memiliki properti, yaitu sifat dari setiap entitas. Setiap entitas data mempunyai jumlah properti yang sama. *Atribut* yang diterapkan sebagai *input* untuk menetapkan kelas dari objek [11]. Pada teknik pendekatan yang digunakan peneliti kategori setiap item dimengerti. Oleh sebab itu, persoalan yang ditangani *supervised learning* menetapkan entitas ke dalam kelas yang sesuai dengan berdasarkan karakteristik dari setiap objek atau entitas.

Berdasarkan kerangka yang telah dijabarkan maka sebuah penelitian dilakukan untuk mengklasifikasikan opini masyarakat di Twitter terhadap Ganjar Pranowo. Tujuan penelitian ini untuk melakukan klasifikasi sentimen masyarakat di media sosial Twitter terhadap Ganjar Pranowo dengan menggunakan metode *Naïve Bayes Classifier* dan melakukan pengujian akurasi metode NBC dalam mengklasifikasi sentimen tersebut. Analisis sentimen dapat digunakan untuk mengevaluasi efektivitas suatu kebijakan yang diterapkan oleh pemerintah kepada masyarakat dalam memungkinkan pemerintah mengatasi kekurangan kebijakan [12]. Hasil dari penelitian akan mendapatkan nilai akurasi sentimen masyarakat terhadap Ganjar Pranowo.

## 2. METODE PENELITIAN

Beberapa langkah yang diterapkan dalam menyelesaikan penelitian ini untuk mencapai hasil yang efisien. Langkah-langkah tersebut ialah *crawling* data, *preprocessing* data, pembobotan data, *feature selection*, klasifikasi *Naïve Bayes Classifier* dan evaluasi nilai akurasi. Langkah-langkah ini ditunjukkan secara lebih rinci pada Gambar 1.



Gambar 1. Flowchart Penelitian

## 2.1 Crawling

*Crawling* data merupakan proses pengumpulan data dari suatu *database*. Dataset yang akan diproses dalam penelitian ini diunduh dari *server* Twitter dalam bentuk data *tweet* dan user serta atribut-atributnya yang membahas Ganjar Pranowo. *Crawling* atau pengumpulan dataset dengan cara memanfaatkan *Application Programming Integration* (API) Twitter sebanyak 4000 *tweet* yang dikumpulkan dari tanggal 29 November 2022 sampai 03 Maret 2023 menggunakan bahasa *python* [13]. Dataset yang digunakan dalam penelitian ini dapat diakses pada google sheet berikut:

<https://docs.google.com/spreadsheets/d/15xEJ8kzdXA3UnJj0AZvQwx3cSAmDQvpD/edit?usp=sharing&ouid=115369590705216222738&rtpof=true&sd=tru>

## 2.2 Pelabelan

Data yang akan digunakan selanjutnya masuk ke tahap pelabelan untuk menentukan sentimen masing-masing data dengan kelas positif dan negatif oleh validator [14]. Pada penelitian ini penulis dibantu oleh seorang validator untuk melabelkan data yang digunakan yaitu bapak Revi Kasman, S.Pd.

## 2.3 Preprocessing

*Preprocessing* merupakan proses mengolah dataset awal menjadi dataset yang lebih mudah dipahami [15]. Tahapan ini bertujuan untuk meningkatkan nilai akurasi klasifikasi sentimen Ganjar Pranowo. Pemrosesan dataset dalam *preprocessing* digunakan melalui tujuh tahapan.

### a. Cleaning

Cleaning diterapkan untuk membersihkan tweet tentang kata-kata yang kurang penting seperti *URL*, tagar '#' dan nama pengguna '@'. Meskipun *URL* Format *URL* antara lain *http*, *https* atau *www*. Proses ini bertujuan untuk meminimalisir *noise* atau data *tweet* yang tidak mendukung [16].

### b. Case Folding

*Case folding* ialah proses untuk menukar huruf atau tulisan besar menjadi tulisan kecil [17]. Tujuan dari tahapan ini ialah untuk menyederhanakan proses selanjutnya.

### c. Tokenizing

*Tokenizing* merupakan tahapan pemecahan kalimat untuk membentuk suatu token atau potongan kata [18].

### d. Normalisasi

Normalisasi tahapan untuk memperbaiki ejaan yang salah pada setiap token atau mengubah menjadi kata baku [19].

### e. Negation Handling

Upaya meningkatkan akurasi pengklasifikasian, dapat menggunakan tahapan *negation handling* pada proses ketika ada token negasi antara lain enggak, tidak, bukan, jangan, belum dan lain-lain [20]. Berikut adalah beberapa langkah untuk menangani *negation handling* dari Vivek Narayanan.

1. Inisialisasi negasi = false
2. For each kata in dok  
If negasi = True  
Ubah kata menjadi = "kata\_negasi" + "kata"  
If kata = "tidak" or "jangan" or "enggak" or "belum"  
Negasi = True  
Elsenegasi = false

Tahapan *negation handling* merupakan ketika ditemukan kata negasi digabungkan dengan kata berikutnya kemudian diganti dengan memakai antonim sesudah kata negasi [21].

### f. Remove Stopword

*Remove stopword* merupakan langkah untuk mengekstrak token-token penting pengembalian token. Tahapan ini menerapkan model *stoplist* (menghilangkan token yang kurang efisien) atau *word list* (mengambil token yang efisien) [22].

### g. Stemming

*Stemming* adalah upaya untuk mengolah ragam morfologi kata yang berbeda menjadi satu bentuk baku. *Stemming* adalah pengurangan variasi morfologi menjadi satu bentuk kata dasar (akar). *Stemming* merupakan tahapan di mana varian morfologi kata yang berbeda dipetakan ke bentuk dasar/batangnya [23].

## 2.4 Pembobotan TF-IDF

Pembobotan TF-IDF adalah salah satu metode untuk menentukan nilai frekuensi dari setiap token dalam suatu dokumen. Jumlah token menunjukkan tingkat penting sebuah token dalam sebuah dokumen. Algoritma *feature*

*weighting* diterapkan untuk memberi nilai bobot kata masing-masing kategori agar sesuai dengan kata kunci kategori yang ada [24]. Berikut rumus untuk menghitung TF-IDF.

$$w_{ij} = tf \times \log \log \left( \frac{D}{df_j} \right) \quad (1)$$

## 2.5 Feature Selection

*Feature Selection* adalah langkah proses yang penting dan meningkatkan nilai akurasi (Putra, 2017). Pemilihan fitur melayani dua tujuan utama. Pertama, tingkatan data pelatihan yang digunakan dalam pengklasifikasian dengan mengurangi ukuran kosakata efektif. Kedua, pemilihan fitur secara umum dapat meningkatkan akurasi klasifikasi dengan menghilangkan fitur-fitur yang mengganggu [25].

## 2.6 Naïve Bayes Classifier

Pengklasifikasi *Naïve Bayes* menganalisis sentimen data menggunakan klasifikasi tingkat Kalimat. Klasifikasi tingkat kalimat mengklasifikasi setiap kata dalam kalimat dan dikenakan dua kelas (positif dan negatif) [26]. Tingkat analisis pertama adalah menambang pendapat orang tentang Ganjar Pranowo. Data diklasifikasikan berdasarkan bobot data menggunakan korpus emosi dan korpus yang dibangun. Untuk mengklasifikasikan sentimen (Positif dan negatif), probabilitas data positif, data negatif dihitung dan dibandingkan untuk memprediksi sentimen orang. Polaritas pernyataan (positif dan negatif) dihitung menggunakan Eq. berdasarkan fitur yang dipilih menggunakan *n* gram (unigram, Bi gram, Tri gram) [27].

*Naïve Bayes Classifier* adalah suatu algoritma yang paling marak atau populer diterapkan dalam penambangan data disebabkan kemudahan dalam penggunaan dan tempo pemrosesan yang lebih cepat, membuatnya simpel dan sangat efisien (Hamzah, 2012). Metode *Naïve Bayes* menggunakan konsep probabilitas yang dijabarkan oleh British Thomas Bayes, yang dapat memperhitungkan masa mendatang berlandaskan pengalaman masa lampau [28]. Berikut rumus *Naïve Bayes Classifier*.

a. Perhitungan Probabilitas Prior

$$p(P/N) = \frac{d(P/N)}{|jumlah\ data|} \quad (2)$$

Penjelasan :

$p(P/N)$  : Kemungkinan dokumen muncul dari kategori positif dan negatif.

$d(P/N)$  : Total semua dokumen pada tiap kategori

$|jumlah\ data|$  : Total semua data training

b. Perhitungan Probabilitas Setiap Fitur (Term)

$$P(X_i | positif/negatif) = \frac{(n_{k, (positif/negatif)} + 1)}{(n, positif/negatif + |kosakata|)} \quad (3)$$

Penjelasan :

$(X_i | positif/negatif)$  : Kemungkinan kata muncul pada kategori.

$X_i$  : Token yang muncul pada sebuah kategori

$(n_{k, (positif/negatif)} + 1)$  : Total frekuensi yang muncul setiap token pada kategori.

$|kosakata|$  : Total seluruh token dari semua sentimen

c. Menentukan Kategori

$$p(P/N|d) = p(P/N) * \prod_i P(a_i|P/N) \quad (4)$$

Penjelasan :

$p(N/P)$  : Kemungkinan kata muncul pada kategori dengan dokumen data testing.

$d$  : Dokumen data baru

$\prod_i P(a_i|N/P)$  : Kemungkinan  $a_i$  muncul pada tiap kategori

$a_i$  : Token baru yang akan di testing.

## 3. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini dirangkum sesuai susunan alur penelitian. Susunan alur penelitian ini mulai dari *crawling* data, *preprocessing*, pembobotan TF-IDF, *feature selection* dan klasifikasi sebagai berikut.

### 3.1 Crawling Data

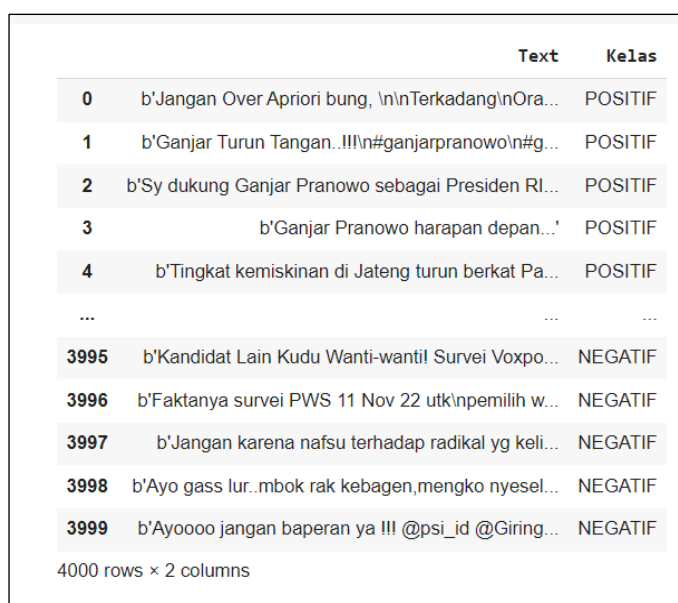
*Crawling* data merupakan tahapan yang dilakukan peneliti untuk mengambil atau mengumpulkan dari server Twitter dengan memanfaatkan API yang tersedia. Pada tahapan ini menghasilkan 4.000 data *tweet* yang berhasil dikumpulkan mulai dari tanggal 29 November 2022 sampai 03 Maret 2023 oleh peneliti untuk digunakan dalam penelitian ini yang bertopik Ganjar Pranowo, dan dikelompokkan secara manual oleh seorang ahli atau validator berdasarkan label yaitu

2.000 data label positif dan 2.000 data label negatif. Setelah ditentukan label pada setiap data akan digunakan untuk tahap selanjutnya. Berikut contoh data *tweet* yang akan digunakan, Tabel 1.

**Tabel 1.** Contoh Data *Tweet*

Label	<i>Tweet</i>
Positif	b'RT @Widyarenee: Keberhasilan alias prestasi Ganjar Pranowo selama 9thn memimpin Jateng ?\nUp lagi ajaaah... kalo ada yg ngga suka fakta ini\`xe2\x80\xa6'
Negatif	b'@BNurvia @ganjarpranowo Kalau Ganjar Pranowo presiden saya yakin Indonesia akan hancur berantakan.'
Positif	b'Tidak bergeser sedikit pun Ganjar pranowo the next presiden RI 2024.'
Negatif	b'Bisa kerja atau tidak anda @ganjarpranowo kau ajarin dulu ya @Dennysiregar7 gimana kerja yg bener.'

Tahapan pertama untuk melakukan klasifikasi dengan menggunakan *source code* bahasa python yaitu *load* data untuk menampilkan semua data yang akan diproses. Dari proses ini peneliti akan menyelesaikan setiap tahapnya dengan menggunakan 4.000 data, di mana dalam *output* terdapat 4.000 *rows* dan 2 *columns*. Gambar 2 tampilan *load* data.



	Text	Kelas
0	b'Jangan Over Apriori bung, \n\nTerkadang\nOra...	POSITIF
1	b'Ganjar Turun Tangan...!!!\n#ganjarpranowo\n#g...	POSITIF
2	b'Sy dukung Ganjar Pranowo sebagai Presiden RI...	POSITIF
3	b'Ganjar Pranowo harapan depan...'	POSITIF
4	b'Tingkat kemiskinan di Jateng turun berkat Pa...	POSITIF
...	...	...
3995	b'Kandidat Lain Kudu Wanti-wanti! Survei Voxpo...	NEGATIF
3996	b'Faktanya survei PWS 11 Nov 22 utk\npilih w...	NEGATIF
3997	b'Jangan karena nafsu terhadap radikal yg keli...	NEGATIF
3998	b'Ayo gass lur..mbok rak kebagen,mengko nyesel...	NEGATIF
3999	b'Ayooooo jangan baperan ya !!! @psi_id @Giring...	NEGATIF

4000 rows x 2 columns

**Gambar 2.** Tampilan *Load* Data

### 3.2 Preprocessing Data

*Preprocessing* data merupakan proses untuk mengubah data mentah menjadi data yang bersih untuk melakukan klasifikasi. Tahapan *preprocessing* data menggunakan beberapa tahapan yaitu:

- Cleaning*, langkah untuk pembersihan data pada setiap tweet dari *URL*, tanda baca, emot ikon dan lain-lain
- Case Folding*, menjadikan semua dataset menjadi huruf kecil
- Tokenizing*, membagi suatu kalimat menjadi kumpulan-kumpulan token
- Normalisasi, proses untuk memperbaiki ejaan yang salah
- Negation Handling*, mengubah kata nangasi beserta kata selanjutnya menjadi satu kata baru
- Remove Stopword*, menghapus token atau kata yang tidak penting (kata penghubung)
- Stemming*, tahapan untuk menghilangkan kata imbuhan menjadi kata dasar

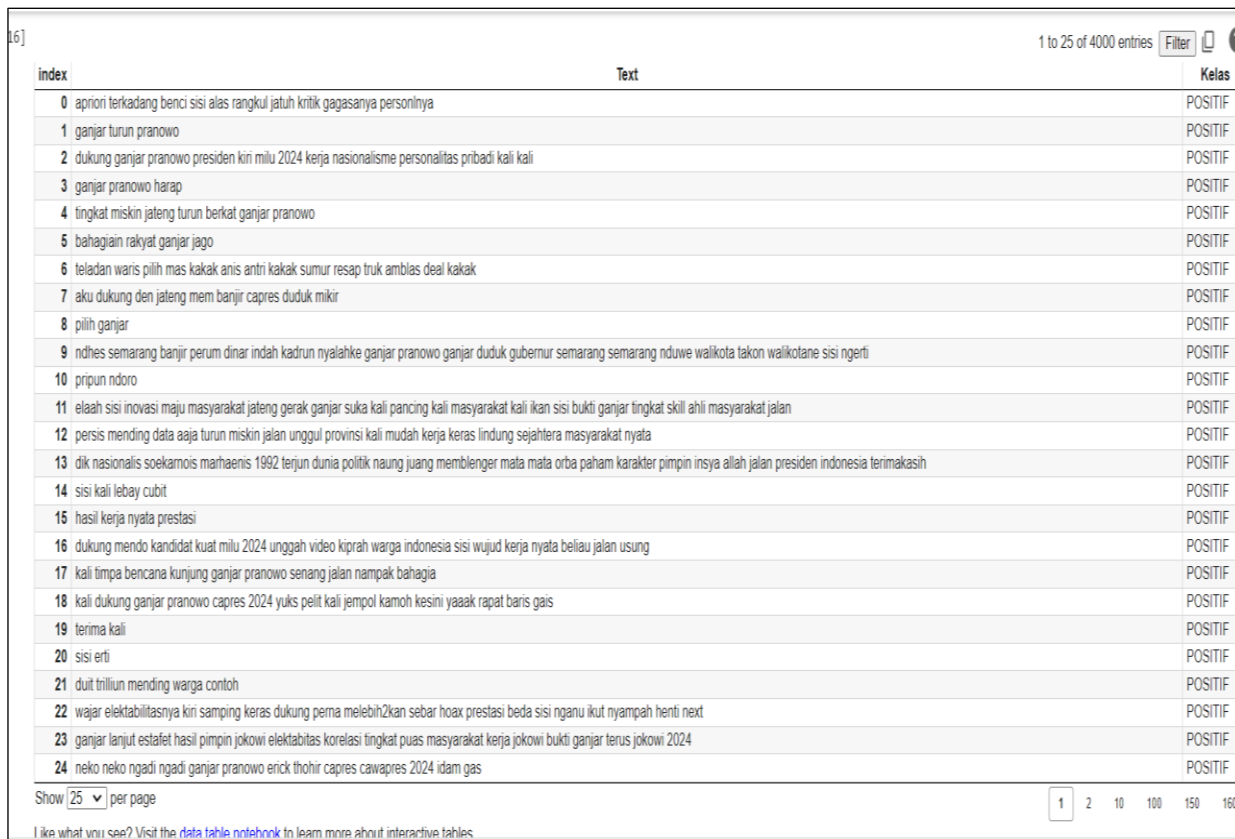
Berikut Tabel 2 contoh tahapan *preprocessing* dataset.

**Tabel 2.** Contoh Tahapan *Preprocessing* Data

Proses	Contoh
Data mentah	b'RT @Widyarenee: Keberhasilan alias prestasi Ganjar Pranowo selama 9thn memimpin Jateng ?\nUp lagi ajaaah... kalo ada yg ngga suka fakta ini\`xe2\x80\xa6'
<i>Cleaning</i>	Keberhasilan alias prestasi Ganjar Pranowo selama 9thn memimpin Jateng Up lagi ajaaah kalo ada yg ngga suka fakta ini
<i>Case Folding</i>	keberhasilan alias prestasi ganjar pranowo selama 9thn memimpin jateng up lagi ajaaah kalo ada yg ngga suka fakta ini
<i>Tokenizing</i>	“keberhasilan”, “alias”, “prestasi”, “ganjar”, “pranowo”, “selama”, “9thn”, “memimpin”, “jateng”, “up”, “lagi”, “ajaaah”, “kalo”, “ada”, “yg”, “ngga”, “suka”, “fakta”, “ini”

Normalisasi	“keberhasilan”, “alias”, “prestasi”, “ganjar”, “pranowo”, “selama”, “9tahun”, “memimpin”, “jateng”, “naik”, “lagi”, “ajaaah”, “kalau”, “ada”, “yang”, “enggak”, “suka”, “fakta”, “ini”
Negation Handling	“keberhasilan”, “alias”, “prestasi”, “ganjar”, “pranowo”, “selama”, “9tahun”, “memimpin”, “jateng”, “naik”, “lagi”, “ajaaah”, “kalau”, “ada”, “yang”, “benci”, “fakta”, “ini”
Remove	“keberhasilan”, “alias”, “prestasi”, “ganjar”, “pranowo”, “9tahun”, “memimpin”, “jateng”, “ajaaah”, “benci”, “fakta”
Stopword	“ajaaah”, “benci”, “fakta”
Stemming	“hasil”, “alias”, “prestasi”, “ganjar”, “pranowo”, “9tahun”, “pimpin”, “jateng”, “aja”, “benci”, “fakta”

Seluruh tahapan *preprocessing* diterapkan sesudah pengumpulan data. Sesudah *pretreatment* selesai, data siap untuk digunakan dalam tahap klasifikasi. Berikut Gambar 3 tampilan hasil *preprocessing* pada data yang akan digunakan yang di screenshot dari google collab.



index	Text	Kelas
0	apriani terkadang benci sisi alas rangkul jatuh kritik gagasannya personilnya	POSITIF
1	ganjar turun pranowo	POSITIF
2	dukung ganjar pranowo presiden kiri milu 2024 kerja nasionalisme personalitas pribadi kali kali	POSITIF
3	ganjar pranowo harap	POSITIF
4	tingkat miskin jateng turun berkat ganjar pranowo	POSITIF
5	bahagiain rakyat ganjar jago	POSITIF
6	teladan waris pilih mas kakak anis antri kakak sumur resap truk ambles deal kakak	POSITIF
7	aku dukung den jateng mem banjar capres duduk mikir	POSITIF
8	pilih ganjar	POSITIF
9	ndhes semarang banjar perum dinar indah kadrun nyalahke ganjar pranowo ganjar duduk gubernur semarang semarang nduwe walkota takon walkotane sisi ngerti	POSITIF
10	pripun ndoro	POSITIF
11	elaah sisi inovasi maju masyarakat jateng gerak ganjar suka kali pancing kali masyarakat kali ikan sisi bukti ganjar tingkat skill ahli masyarakat jalan	POSITIF
12	pemis mending data aja turun miskin jalan unggul provinsi kali mudah kerja keras lindung sejahtera masyarakat nyata	POSITIF
13	dik nasionalis soekarno marhaenis 1992 terjun dunia politik naung juang memblenger mata mata orba paham karakter pimpin insya allah jalan presiden indonesia terimakasih	POSITIF
14	sisi kali lebay cubit	POSITIF
15	hasil kerja nyata prestasi	POSITIF
16	dukung mendo kandidat kuat milu 2024 unggah video kiprah warga indonesia sisi wujud kerja nyata beliau jalan usung	POSITIF
17	kali timpa bencana kunjung ganjar pranowo senang jalan nampak bahagia	POSITIF
18	kali dukung ganjar pranowo capres 2024 yuks pelit kali jempol kamoh kesini yaaak rapat baris gais	POSITIF
19	terima kali	POSITIF
20	sisi erti	POSITIF
21	duit triliun mending warga contoh	POSITIF
22	wajar elektabilitasnya kiri samping keras dukung perma melebihi2kan sebar hoax prestasi beda sisi nganu ikut nyamph henti next	POSITIF
23	ganjar lanjut estafet hasil pimpin jokowi elektabilitas korelasi tingkat puas masyarakat kerja jokowi bukti ganjar terus jokowi 2024	POSITIF
24	neko neko ngadi ngadi ganjar pranowo erick thohir capres cawapres 2024 idam gas	POSITIF

Gambar 3. Tampilan Hasil *Preprocessing*

### 3.3 Pembobotan TF-IDF

Pembobotan TF-IDF adalah tahapan menentukan nilai bobot untuk setiap token pada dokumen yang digunakan untuk pengujian klasifikasi. Dalam penelitian ini, seluruh proses analisis dilakukan dengan menggunakan alat bahasa *python*. Pada tahap sebelumnya telah melakukan *preprocessing* data sebanyak 4.000 dataset yang akan diproses. Langkah berikutnya pembobotan untuk semua dataset dengan menerapkan kalkulasi TF-IDF (*Term Frequency Inverse Document Frequency*) secara *feature weighting*. Konsep kalkulasi TF-IDF yang berkaitan dari langkah *preprocessing* sebelumnya di mana setelah hasil *preprocessing* dihitung frekuensi munculnya token pada dokumen yang ditunjukkan, arti token pada dokumen. Berikut langkah-langkah untuk mendapatkan nilai bobot pada setiap kata.

- Menetapkan TF (Term Frequency) yaitu Q, D1, D2, D3

$$Q = \text{Query}$$

$$D1 = \text{Doc 1} \quad D3 = \text{Doc 3}$$

$$D2 = \text{Doc 2} \quad D4 = \text{Doc 4}$$

- Menjumlahkan df (data frekuensi)

$$df = Q + D1 + D2 + D3 + D4$$

- Menjumlahkan  $Idf = \log(D/df)$

- Menetapkan W (bobot kata) yaitu Q, D1, D2, D3, D4

$$Q = \log(D/df) \times Q$$

Berikut Gambar 4 tampilan hasil TF-IDF yang di screenshot dari google collab.



	00	000	01	02	03	04	05	056	06	07	...	ziarah	zikmal	zim	zodiak	zombie	zon	zoom	zubair	zulkifli	Kelas		
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3995	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3996	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3997	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3998	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3999	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF

4000 rows x 5256 columns

Gambar 4. Tampilan Hasil TF-IDF

### 3.4 Feature Selection

Feature selection merupakan tahapan yang bertujuan untuk menyeleksi jumlah properti atau fitur yang terlibat dalam menentukan nilai kelas target dengan menggunakan nilai batas ambang. Fitur yang diabaikan merupakan fitur yang tidak relevan dan data yang berlebihan. Berikut ini tampilan *feature selection* dengan menggunakan nilai *thresholding* = 0.001, dimana hasilnya terdapat 4.000 rows dan 150 columns, seperti Gambar 5 yang di screenshot dari google collab.

	2024	aamin	ajar	alam	alhamdulillah	anak	anies	apresiasi	atas	bagus	...	ubah	udah	ulama	umkm	untung	upaya	urus	usaha	warga	Kelas			
0	0.000000	0.0	0.0	0.0		0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	
1	0.000000	0.0	0.0	0.0		0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF
2	0.224144	0.0	0.0	0.0		0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF
3	0.000000	0.0	0.0	0.0		0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF
4	0.000000	0.0	0.0	0.0		0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3995	0.000000	0.0	0.0	0.0		0.0	0.0	0.172854	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3996	0.000000	0.0	0.0	0.0		0.0	0.0	0.133145	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3997	0.000000	0.0	0.0	0.0		0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3998	0.000000	0.0	0.0	0.0		0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF
3999	0.000000	0.0	0.0	0.0		0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF

4000 rows x 150 columns

Gambar 5. Tampilan Feature Selection

### 3.5 Klasifikasi

Klasifikasi adalah langkah yang akan diterapkan dalam pengujian dataset yang sudah selesai melakukan beberapa tahapan sebelumnya. Tahapan yang telah diselesaikan yaitu dari *crawling* data, pelabelan, *preprocessing*, pembobotan TF-IDF dan *feature selection*. Maka tahap selanjutnya yaitu melakukan klasifikasi dan menguji akurasi dengan menerapkan metode *Naïve Bayes Classifier*. Berikut tampilan klasifikasi menggunakan jumlah data uji 10% dari 4.000 data yang digunakan atau perbandingan 90:10. Hasil klasifikasi terdapat 400 data, hasilnya seperti Gambar 6 yang di screenshot dari google collab.

Hasil Klasifikasi																							
	2024	aamin	ajar	alam	alhamdulillah	anak	anies	apresiasi	atas	bagus	...	ubah	udah	ulama	umkm	untung	upaya	urus	usaha	warga	Kelas Target	Hasil Klasifikasi	
0	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	NEGATIF
1	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.533887	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	POSITIF
2	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	POSITIF
3	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF	NEGATIF
4	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	NEGATIF
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
395	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF	NEGATIF
396	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NEGATIF	POSITIF
397	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	POSITIF
398	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	POSITIF
399	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	POSITIF	POSITIF

400 rows x 151 columns

Gambar 6. Tampilan Hasil Klasifikasi Naïve Bayes Classifier

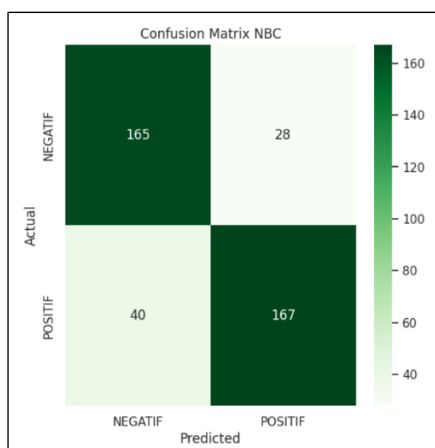
### 3.6 Akurasi

Tahap selanjutnya menguji akurasi dari pengklasifikasian dengan metode *Naïve Bayes Classifier*. Pengklasifikasian menghasilkan *confussion matrix* kelas negatif dengan hasil klasifikasi negatif sebanyak 165 data, kelas negatif memiliki nilai positif sebanyak 28 data dan kelas positif dengan hasil klasifikasi positif sebanyak 167 data dan dengan hasil negatif 40 data. Berikut *confussion matrix* dari uji coba pada Tabel 3

**Tabel 3.** *Confussion Matrix*

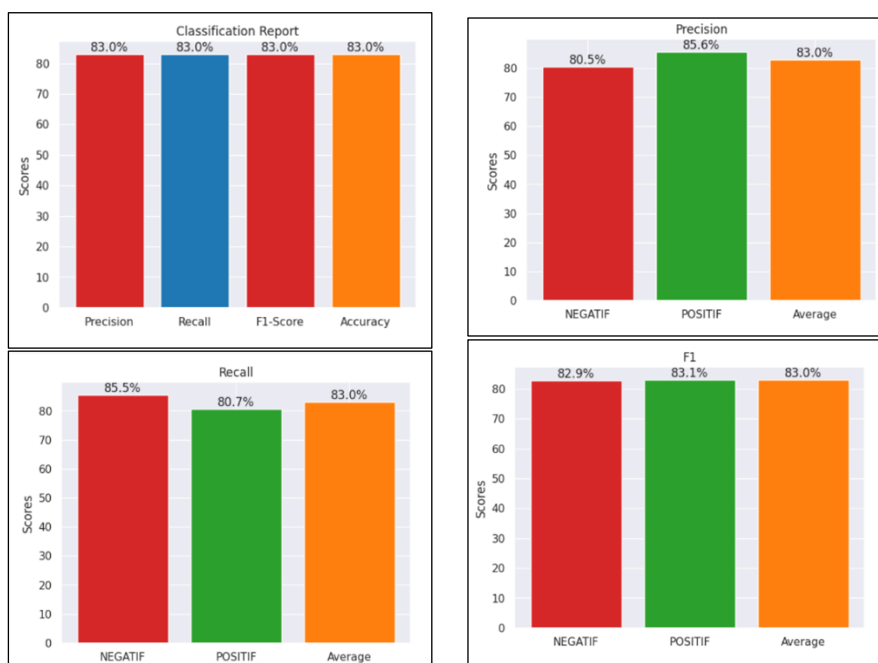
<i>Confussion Matrix</i>		
	Negatif	Positif
Negatif	165	28
Positif	40	167

Dari Tabel 3 *confussion matrix* di atas dapat disimpulkan False Positif (FP) dan False Negatif (FN). TP merupakan banyak data positif yang terklasifikasi oleh model dengan benar, maka TN ialah banyak data negatif pengklasifikasian dengan benar. Sementara itu, FP merupakan banyak data positif yang dianggap meleset oleh model, dan FN merupakan banyak data negatif yang meleset diklasifikasikan menurut model klasifikasi masing-masing sentimen. Berikut Gambar 7 grafik *confussion matrix* dari hasil pengujian.



**Gambar 7.** Grafik *Confussion Matrix*

Setelah hasil pengujian didapatkan maka diperoleh *classification report* dari yang dihasilkan. Berikut tampilan *classification report* pada Gambar 8.



**Gambar 8.** *Classification Report*

Dari hasil perhitungan nilai akurasi dengan melakukan klasifikasi sentimen masyarakat di Twitter terhadap Ganjar Pranowo menerapkan metode *Naïve Bayes Classifier* memperoleh akurasi yang efisien atau kategori baik, *accuracy* 83.0% dengan *precision* 83.0% dan *recall* 83.0%.

#### 4. KESIMPULAN

Berdasarkan dataset yang dikumpulkan, penelitian ini menggunakan 4.000 data *tweet* dengan kelas positif 2.000 data dan kelas negatif 2.000 data yang dilabelkan oleh seorang validator secara manual. Data yang digunakan diperoleh dari *database* Twitter dengan memanfaatkan API yang tersedia. Tahapan yang dilalui untuk melakukan pengklasifikasian melalui tahapan *preprocessing*, pembobotan TF-IDF dan *feature selection*. Pengujian klasifikasi sentimen masyarakat di Twitter terhadap Ganjar Pranowo dengan metode *Naïve Bayes Classifier* menggunakan 10% data uji dari jumlah data yang digunakan dan nilai *Threshold* 0.001. Pengujian klasifikasi dengan menerapkan metode *Naïve Bayes Classifier* dengan hasil akurasi 83.0%. dari hasil akurasi dapat disimpulkan bahwasanya metode NBC dapat diimplementasikan dalam pengklasifikasian sentimen masyarakat terhadap Ganjar Pranowo di Twitter berdasarkan data *tweet*. Saran untuk penelitian selanjutnya dapat dikembangkan dengan model pembobotan menggunakan fitur lainnya untuk mendapatkan nilai pembobotan antara lain *Chi-Square*, *Particle Swarm Optimization* dan lainnya.

#### REFERENCES

- [1] N. M. A. J. Astari, G. H. D. Dewa, and I. Gede, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," *J. Sist. dan Inform.*, vol. 15, no. 1, pp. 27–29, 2020.
- [2] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018.
- [3] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019.
- [4] R. Milla and M. Sholehkah, "Sosok ganjar pranowo sebagai pemimpin jawa tengah," *Univ. Muhammadiyah Yogyakarta*, no. June, 2021.
- [5] E. S. Romaito, M. K. Anam, Rahmadden, and A. N. Ulfah, "Perbandingan Algoritma Svm Dan Nbc Dalam Analisa Sentimen Pilkada Pada Twitter," pp. 169–179, 2021.
- [6] S. A. Sutresno, "Analisis Sentimen Masyarakat Indonesia Terhadap Dampak Penurunan Global Sebagai Akibat Resesi di Twitter," vol. 4, no. 4, pp. 1959–1966, 2023, doi: 10.47065/bits.v4i4.3149.
- [7] R. Karisma, S. Lestanti, and M. T. Chulkamdi, "Aplikasi Klasifikasi Sentimen Pada Ulasan Smartphone Di Situs Jual Beli Online Berbasis Web Menggunakan Naive Bayes Dengan Tf-Idf," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 1, pp. 31–37, 2021.
- [8] D. F. Zhafira, B. Rahayudi, and I. Indriati, "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube," *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 2, no. 1, pp. 55–63, 2021.
- [9] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022.
- [10] H. Malik, E. M. Shakshuki, and A. U. H. Yasar, "Approximating Viewership of Streaming T.V Programs Using Social Media Sentiment Analysis," *Procedia Comput. Sci.*, vol. 198, no. 2021, pp. 94–101, 2021.
- [11] E. Hari, A. Prastyo, P. Studi, and R. Wiratsongko, "Implementasi Teknik Web Scraping Pada Situs Berita Menggunakan Metode Supervised learning Implementasi Web Scraping Pada Situs Berita Menggunakan Metode Supervised learning IGL Putra Eka Prisma."
- [12] W. P. Anggraini, M. S. Utami, J. M. Berlianty, and E. Sellya, "Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Prakerja di Indonesia," vol. 13, no. 4, 2021.
- [13] A. C. Khotimah and E. Utami, "Perbandingan Algoritma Naïve Bayes Classifier, K-Nearest Neighbor Dan Support Vector Machine Dalam Klasifikasi Karakter Individu Pada Akun Twitter," vol. 3, no. 3, pp. 673–680, 2022.
- [14] R. Indransyah, "Klasifikasi Sentimen Pergelaran Motogp Di Indonesia Menggunakan Algoritma Correlated Naïve Bayes Clasifier," pp. 60–66, 2022.
- [15] A. Syahputri, M. Zarlis, and Suherman, "Analisis Klasifikasi Sentimen Mahasiswa Terhadap Strategi Pembelajaran Online Pada Media Sosial Twitter Menerapkan Metode Naïve Bayes," *Konf. Nas. Teknol. Inform. dan Komput.*, vol. 4, pp. 18–21, 2020.
- [16] H. S. Ginting, K. M. Lhaksana, and D. T. Murdiansyah, "Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter Menggunakan Naive Bayes," vol. 5, no. 1, pp. 1793–1802, 2018.
- [17] S. Lestari and S. Saepudin, "Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naïve Bayes," 2021.
- [18] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," vol. 10, pp. 71–76, 2020.
- [19] W. Widyatama, "Klasifikasi Komentar Sentimen Pada Celebrity Endorsement Online Shop di Instagram Menggunakan Metode Naive Bayes Classifier," *Tugas Akhir UIN*, 2021.
- [20] T. S. Az-Zahra, "Analisis sentimen terhadap belajar daring menggunakan optimasi naive bayes classifier dengan adaboost," *Tazkia Shabrina*, 2021.
- [21] M. D. Saputra, "Klasifikasi Sentimen Pada Video Channel Youtube Raditya Dika Menggunakan Metode Naive Bayes Classifier," *Tugas Akhir UIN*, 2021.
- [22] M. A. Djamaludin, A. Triayudi, and E. Mardiani, "Analisis Sentimen Tweet KRI Nanggala 402 di Twitter menggunakan



- Metode Naïve Bayes Classifier,” J. JTİK (Jurnal Teknol. Inf. dan Komunikasi), vol. 6, no. 2, pp. 161–166, 2022.
- [23] Y. Harfian, “Klasifikasi Sentimen Aplikasi Dompot Digital Dana Pada Komentar Di Instagram Menggunakan Naive Bayes Classifier,” Tugas Akhir UIN, 2021.
- [24] K. Azmi, “Klasifikasi Sentimen Terhadap Video Dakwah Ulama Islam Berdasarkan Komentar Di Youtube Menggunakan Naive Bayes Classifier,” J. Ekon. Vol. 18, Nomor 1 Maret201, vol. 2, no. 1, pp. 41–49, 2020.
- [25] A. Herdhianto, Sentiment Analysis Menggunakan Naive Bayes Classifier (NBC) Pada Tweet Tentang Zakat. 2020.
- [26] A. Rachmat C and Y. Lukito, “Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes,” J. Inform. dan Sist. Inf. Univ. Ciputra, vol. 02, no. 02, pp. 26–34, 2016.
- [27] J. Ramsingh and V. Bhuvanawari, “An efficient Map Reduce-Based Hybrid NBC-TFIDF algorithm to mine the public sentiment on diabetes mellitus – A big data approach,” J. King Saud Univ. - Comput. Inf. Sci., vol. 33, no. 8, pp. 1018–1029, 2021.
- [28] M. H. Asnawi, I. Firmansyah, R. Novian, and R. S. Pontoh, “Perbandingan Algoritma Naive Bayes , K-NN , dan SVM dalam Pengklasifikasian Sentimen Media Sosial,” 2021.