

## **Analisis Sentimen Kebijakan Pembelian Gas 3 Kg dengan KTP Menggunakan Naïve Bayes**

**Muhammad Ridwan Pratama<sup>1\*</sup>, Ahmad Fauzi<sup>2</sup>, Deden Wahiddin<sup>3</sup>, Adi Rizky Pratama<sup>4</sup>**  
 Teknik Informatika, Universitas Buana Perjuangan Karawang, Karawang, Indonesia  
 \*e-mail *Corresponding Author*: if20.muhammadpratama@mhs.ubpkarawang.ac.id

### **Abstract**

*The government introduced a policy on January 1 2024 regarding the purchase of 3 KG LPG which requires user registration using a Resident Identity Card (KTP), where residents are required to show their KTP and family card (KK) to register via the merchant application at the gas station. This raises problems regarding the availability, distribution of gas, and precisely the recipients of subsidies. The public responds whether they accept or not according to the policy on one of YouTube's social media. The Naïve Bayes algorithm is used to analyze public sentiment regarding the policy of purchasing 3 KG LPG using KTP, data obtained from social media YouTube with crawling techniques using Google Colab, with a collection period of April 2023 to January 2024. Analysis stages include text preprocessing, TF-IDF, labeling uses the Tex Blob library, and algorithm calculations use Naïve Bayes. Visualization in the form of a word cloud with terms that often appear. Classification reports are used to evaluate algorithms with Naive Bayes algorithm accuracy of 84%, precision of 85%, recall of 84%, and f1-score of 83%.*

**Keywords:** Resident Identity Card; Policy; LPG 3 Kg; Naive Bayes; Sentiment.

### **Abstrak**

Pemerintah memperkenalkan kebijakan pada tanggal 1 Januari 2024 terkait pembelian LPG 3 KG yang memerlukan pendaftaran pengguna memakai Kartu Tanda Penduduk (KTP), dimana warga wajib menunjukkan KTP serta Kartu keluarga (KK) untuk mendaftar melalui aplikasi *merchant* pada pangkalan gas. Hal ini menimbulkan permasalahan ketersediaan, distribusi gas, dan tepatnya penerima subsidi, Masyarakat memberikan respon menerima atau tidak dengan kebijakan pada salah satu media sosial *youtube*. Algoritma Naïve Bayes dipergunakan untuk menganalisis sentimen masyarakat terkait kebijakan pembelian LPG 3 KG dengan menggunakan KTP, data diperoleh dari media sosial *Youtube* dengan teknik *crawling* menggunakan *Google Colab*, dengan periode pengumpulan April 2023 hingga Januari 2024. Tahapan analisis meliputi *text preprocessing*, *TF-IDF*, pelabelan menggunakan *library Tex Blob*, dan perhitungan algoritma menggunakan *Naïve Bayes*. *Visualisasi* berupa *word cloud* dengan istilah-istilah yang seringkali muncul. *Classification report* digunakan untuk mengevaluasi algoritma dengan akurasi algoritma *Naive Bayes* sebesar 84%, *presisi* 85%, *recall* 84%, dan *f1-score* 83%.

**Kata kunci:** Kartu Tanda Penduduk; Kebijakan; LPG 3 Kg; Naive Bayes; Sentimen.

### **1. Pendahuluan**

Sejak tanggal 1 Januari 2024, pemerintah mengumumkan kebijakan baru terkait akuisisi LPG 3 Kg yang melibatkan registrasi pengguna dengan menggunakan Kartu Identitas Penduduk (KTP). Proses pendaftaran dilaksanakan dipangkalan resmi yang dimiliki oleh Pertamina, dimana warga diharuskan menunjukkan KTP untuk mendaftar melalui aplikasi *merchant* yang diklaim pihak pangkalan[1]. Kebijakan pemerintah dalam mengatur distribusi LPG 3 kg, khususnya melalui persyaratan penggunaan Kartu Tanda Penduduk (KTP) untuk pembelian, merupakan isu penting yang perlu diteliti. LPG 3 kg adalah kebutuhan dasar bagi banyak masyarakat Indonesia, dan kebijakan ini bertujuan untuk memastikan subsidi yang diberikan tepat sasaran serta mengurangi penyalahgunaan[2]. Oleh sebab itu, memahami reaksi dan persepsi masyarakat terhadap kebijakan ini sangat penting untuk menilai efektivitas serta dampak sosial dan ekonominya.

Tingginya penggunaan media sosial dikalangan masyarakat indonesia, dimana media sosial tidak hanya berfungsi sebagai sarana hiburan, melainkan juga sebagai *platform* untuk

menyampaikan pandangan terkait topik politik diIndonesia[3]. masyarakat dapat dengan mudah mengetahui isu-isu yang terkait dengan kebijakan yang akan dibahas oleh pemerintah, Fenomena ini terlihat pada konteks pembelian gas LPG 3 Kg yang harus menggunakan KTP, dimana masyarakat banyak mengungkapkan reaksinya melalui media sosial, termasuk *youtube*, sebagai saluran untuk mengekspresikan pendapat mereka[4]. Melalui *platform* media sosial *youtube* yang semakin populer, beragam pandangan masyarakat juga semakin meluas. Khususnya dalam peraturan pembelian gas LPG 3 Kg. Dengan pengumuman kebijakan pemerintah tentang pembelian gas LPG 3 Kg yang harus menggunakan KTP, dapat memunculkan berbagai respons yang luas dari masyarakat, mulai dari tanggapan positif hingga negatif. Maka dari itu perlu dilakukan analisis sentimen pada kebijakan pembelian gas LPG menggunakan Kartu Tanda Penduduk (KTP).

Teknik yang disebut analisis sentimen digunakan untuk mengidentifikasi dan mengklasifikasikan opini yang dinyatakan dalam teks, khususnya untuk menentukan apakah tanggapan komentar mempunyai sikap positif, negatif, atau netral[5]. Pada bidang analisis sentimen, terdapat berbagai prosedur penjabaran yang tersedia, dan salah satunya adalah pengklasifikasi *Naïve Bayes*, Herlambang et. al. pada tahun 2021 dalam penelitiannya yang berjudul "Sentimen Analisis Pengguna *Twitter* Terhadap Kebijakan Merdeka Belajar Menggunakan Algoritma *Naïve Bayes*". Penelitian ini menggunakan algoritma *Naïve Bayes*, yang telah terbukti efektif dalam klasifikasi teks, untuk menganalisis sentimen masyarakat dari platform media sosial seperti *YouTube*. Pendekatan ini merupakan kontribusi dalam memahami persepsi publik terhadap kebijakan publik[6].

Berdasarkan latar belakang pada penelitian ini bertujuan untuk menganalisis sentimen kebijakan pembelian gas LPG 3 Kg menggunakan KTP dengan menerapkan algoritma *Naïve Bayes* terhadap data online. Pemilihan algoritma *Naïve Bayes* dalam penelitian ini dikarenakan memiliki taraf akurasi tinggi serta bisa membuat hasil analisis dengan cepat serta efektif. Hasil dari penelitian ini diyakini akan memberikan data berharga dalam menilai reaksi masyarakat terhadap strategi pembelian gas LPG 3 kg yang memerlukan penggunaan KTP. Oleh karena itu, diharapkan pemerintah bisa mempertimbangkan atau mengambil keputusan yang lebih baik, mengurangi dampak negatif, dan memaksimalkan manfaat bagi masyarakat.

## 2. Tinjauan Pustaka

Beberapa penelitian tentang sentimen analisis telah banyak dilakukan, penelitian sebelumnya tentang. Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Prakerja Indonesia (Winda, 2020), penelitian menggunakan metode klasifikasi *Naïve Bayes* dengan menggunakan data dari komentar masyarakat di *Twitter*. Akurasi sebesar 91,06% dan presisi sebesar 90,35%. Analisis tersebut mengungkapkan bahwa 52,87% dari komentar memiliki sentimen negatif terhadap kebijakan tersebut[7].

Rahmat, (2024), dalam penelitian Analisis Sentimen Aplikasi Identitas Kependudukan Digital (IKD) Menggunakan Metode *Naïve Bayes*, menggunakan metode *Naïve Bayes* dengan data yang dipakai dalam penelitian ini yakni data komentar dari masyarakat di media sosial *Twitter*, penelitian ini menghasilkan nilai akurasi sebesar 91.9%. Dari hasil pemeriksaan opini dengan metode *Naïve Bayes* cenderung diduga budaya Indonesia memiliki pandangan positif terhadap penggunaan aplikasi Identitas Kependudukan Digital (IKD)[8].

Dalam penelitian Kurnia, (2024), Dengan metode *Naïve Bayes*, data penelitian ini diperoleh dari *Twitter*. Analisis Sentimen Opini Masyarakat Terkait Pemindahan Ibu Kota Indonesia. Hasil penelitian menunjukkan bahwa mayoritas masyarakat Indonesia memberikan respons positif terhadap rencana pemindahan ibu kota negara Indonesia. Dari 1449 data yang dianalisis, sebanyak 1073 data (74%) menunjukkan sentimen positif, sedangkan 376 data (26%) menunjukkan sentimen negatif[9].

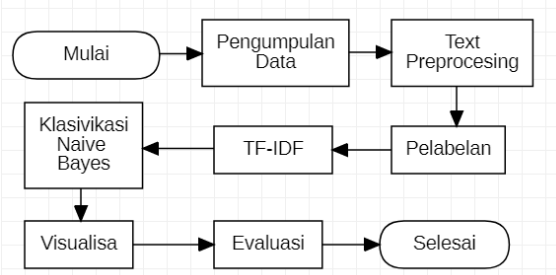
Sementara Nanis mengenai penelitian, (2022). *Analysis of the effect of the MBKM internship Program and certified independent study (MSIB) on university and performance Universitas 17 Agustus 1945 Surabaya*, Kebijakan MBKM, magang, dan belajar mandiri bersertifikat, dengan kontribusi sebesar 75,4%, memberikan dampak positif terhadap kinerja Universitas 17 Agustus 1945 Surabaya secara berbarengan. Menggunakan metode analisis deskriptif dan inferensial untuk menganalisis kesiapan dan dampak program MBKM. Data mencakup hasil survei yang dipelopori oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, melibatkan responden dari berbagai segmen, termasuk mahasiswa, dosen, dan tenaga kependidikan, Hasil survei yang dilakukan menunjukkan bahwa

secara umum mereka siap mengikuti program MBKM khususnya Magang dan Belajar Mandiri Bersertifikat (MSIB)[10]. Pada penelitian ini, analisis sentimen akan diuji dengan menggunakan metode Naive Bayes. proses ekstraksi untuk mendapatkan informasi sebanyak mungkin.

### 3. Metode Penelitian

#### 3.1 Diagram Alur Penelitian

Riset ini dilaksanakan dengan meliputi serangkaian langkah-langkah yang dijelaskan untuk memudahkan pemahaman proses penelitian. Penelitian ini bertujuan merancang sistem, dengan fokus pada sentimen analisis terkait kebijakan pemerintah mengenai pembelian gas LPG 3 Kg yang mewajibkan penggunaan KTP. Langkah-langkah yang diambil dalam penelitian sebelumnya dapat diilustrasikan melalui gambar 1 dibawah ini.



Gambar 1. Alur Penelitian

#### 3.2 Pengumpulan Data

Metode yang dipakai untuk mendapatkan atau memperoleh data dengan maksud mencapai tujuan penelitian tertentu[11]. Data berupa komentar yang diberikan masyarakat pada kolom komentar berasal dari media sosial *youtube*, dengan mengambil pada kanal berita yang ada, data tersebut *dicrawling* dengan menggunakan bantuan *google collab* menggunakan bahasa pemrograman *Python*.

#### 3.3 Text Preprocessing

*Text preprocessing* adalah tahapan krusial dalam pengolahan data menggunakan alat *text mining*[12]. Pada proses ini, data mentah dibersihkan untuk diubah menjadi data yang berkualitas. Beberapa teknik dalam *text preprocessing* mencakup *Cleansing Data*, *Case Folding*, *Word Normalization*, *Tokenization*, *Stopword*, dan *Stemming*.

##### 3.3.1 Cleansing Data

Data *cleansing* atau data *cleaning* ialah langkah yang melibatkan perbaikan atau penghapusan kesalahan, ketidaksesuaian, dan ketidakakuratan dalam suatu set data. Mulai dari menghapus URL, menghapus angka, menghapus karakter tanda baca, menghapus emoji, menghapus spasi, dan karakter aneh[13].

##### 3.3.2 Case Folding

Perubahan huruf besar menjadi huruf kecil, merupakan bagian dari transformasi data yang disebut dengan *case folding*. bertujuan untuk memastikan bahwa perbedaan huruf besar dan kecil tidak mempengaruhi perbedaan antar teks[14].

##### 3.3.3 Word Normalisasi

*Word Normalization* mengubah istilah yang salah eja menjadi kata yang benar, seperti yg menjadi yang. Manfaatkan *Sastrawi stemmer* untuk mengganti kata menjadi bentuk normal. Kamus *sastrawi* akan digunakan untuk mencocokkan setiap kata dalam teks, dan jika ada yang cocok, maka kata baku yang cocok dengan kamus yang digunakan akan menggantikannya[15].

##### 3.3.4 Tokenisasi

Proses mengubah teks menjadi daftar kata yang disebut token dikenal sebagai tokenisasi. Fungsi tokenisasi memanfaatkan modul NLTK untuk memisahkan sekelompok kata menjadi token (kata), atau potongan kata tunggal[16].

### 3.3.5 Stopword

penghapusan teks dimana kata-kata dianggap tidak memberikan informasi disebut penghapusan *stopword*. Setelah proses tokenisasi, dimana setiap token atau kata pada teks dibagi lagi untuk proses penghapusan *stopword*. Langkah ini bertujuan untuk meningkatkan kualitas pada teks[17].

### 3.3.6 Stemming

*Stemming* menghilangkan imbuhan suatu kata sehingga menghasilkan kata dasar. Menempatkan bentuk kata ke dalam bentuk dasarnya dan menguraikan bentuk tersebut ke dalam istilah dasarnya. *Stemming* juga dapat mengurangi jumlah indikator dalam sebuah cerita, sehingga kata-kata yang memiliki awal atau akhir dapat kembali ke bentuk aslinya[18].

### 3.4 Pemberian Label

Menandai data dengan tujuan menyederhanakan proses analisis dan pemrosesan disebut pelabelan data. Pemberian label menggunakan *Google Colab* dan *Library Textblob* yang dapat menggunakan data dari komentar *channel YouTube* untuk memberikan label. *TextBlob* adalah *library Python* yang bekerja dengan data tekstual dan hanya bisa mengenali bahasa Inggris. Hasil *TextBlob* digunakan dalam proses pembelajaran bahasa alami seperti ekstraksi frasa atau istilah benda, analisis sentimen, serta penjabaran[19].

### 3.5 TF-IDF

Dengan mengubah informasi sastra menjadi informasi matematika, sebuah siklus yang dikenal sebagai penilaian TF-IDF (*Term Frekuensi-Invers Dokumen Frekuensi*) memberikan bobot pada setiap kata. Untuk memilih seberapa krusial suatu kata pada sebuah dokumen, digunakan ukuran faktual yang disebut TF-IDF. Pengulangan suatu istilah dalam setiap laporan, yang menunjukkan seberapa besar kata tersebut dalam setiap catatan, dikenal sebagai TF. DF merupakan pengulangan kata dalam dokumen yang menunjukkan betapa berbedanya kata tersebut. Tingkat keparahan sebuah kata akan lebih jelas jika sering muncul dalam laporan dan kurang jelas jika sering muncul dalam catatan[20]. Nilai berapa kali sebuah kata muncul dalam sebuah dokumen disebut *term frekuensi* (TF). Rumus dapat digunakan untuk menentukan nilai TF.

$$TF = \frac{\text{(jumlah kemunculan kata dalam dokumen)}}{\text{(jumlah kata dalam dokumen)}} \dots\dots\dots(1)$$

Nilai TF ini tidak menyampaikan informasi perihal pentingnya kata pada catatan. Konjungsi dan istilah-istilah umum lainnya, yang seringkali ada dalam sebuah dokumen, memiliki nilai TF yang tinggi, meskipun dokumen tersebut kurang penting. Oleh karena itu, diperlukan metode yang berbeda, yaitu *Inverse Document Frekuensi* (IDF). IDF memberi bobot pada kata-kata yang hanya muncul di sebagian kecil dokumen.

$$IDF = \log \frac{N}{n} \dots\dots\dots(2)$$

Dimana N yaitu jumlah arsip pada koleksi file, serta n menyatakan jumlah laporan yang berisi kata tersebut. Setelah nilai TF dan IDF didapat, maka nilai TF-IDF dapat ditentukan dengan menaikkan nilai TF sebesar nilai IDF. Kata-kata dengan nilai TF-IDF tinggi dianggap lebih signifikan dan memberikan kontribusi lebih besar terhadap pemilihan topik dokumen atau koleksi[21].

### 3.6 Klasifikasi Naïve Bayes

Metode pengelompokan probabilistik langsung yang dikenal sebagai Naïve Bayes menggunakan frekuensi informasi dan campuran nilai untuk menghitung probabilitas. Fakta bahwa metode ini hanya memerlukan sejumlah kecil data penyiapan untuk mengetahui ambang batas penilaian yang diharapkan untuk siklus klasifikasi inilah yang membuatnya berguna[22].  
Persamaan Metode *Naïve Bayes*:

$$P(H|K) = \frac{P(X|H)P(H)}{P(X)} \dots\dots\dots(3)$$

$P(X)$ 

Di mana:

H = Prediksi X = Aktual

$P(H)$  = Peluang Prediksi  $P(X)$  = Peluang Aktual

$P(X|H)$  = Peluang kelas Aktual untuk kelas Prediksi

### 3.7 Visualisasi

Proses penyajian data olahan dalam bentuk gambar, diagram, atau animasi dikenal dengan istilah *visualisasi*. Model yang dibuat sebelumnya untuk membuat prediksi berdasarkan data pengujian menampilkan jumlah positif dan negatif dalam bentuk diagram[23]. Merepresentasikan visual dari kata-kata yang muncul dalam data set, menggunakan gambar *masking* untuk memberikan bentuk khusus pada *Word Cloud*.

### 3.8 Evaluasi

Hasil analisis sentimen dievaluasi dan diukur dalam evaluasi. Baik kinerja model analisis sentimen maupun keakuratan dan konsistensi serangkaian pengukuran hasil analisis dievaluasi selama proses evaluasi. Performa model pembelajaran mesin dievaluasi menggunakan matriks[24]. Dalam klasifikasi, evaluasi hasil model dapat dipresentasikan dalam bentuk *confusion matrix* dengan bantuan *Classification report* yang menampilkan berbagai metrik evaluasi kinerja model klasifikasi, mencakup *matrix-matrix* seperti *presisi*, *recall*, *F1-score*, dan *support*. Mengukur performa model *Naïve Bayes* pada dataset untuk memberikan gambaran sejauh mana model dapat melakukan pengelompokan secara akurat.

## 4. Hasil Dan Pembahasan

### 4.1 Hasil Pengumpulan Data

Pengumpulan data dengan teknik *crowling* atau proses otomatis pengumpulan data, melalui penggunaan *tools Google Colab* yang disediakan oleh *google* dengan bahasa pemrograman *python*. Pada *syntax* dibawah untuk bisa mengambil data komentar *youtube* dibutuhkan *Youtube Data API* agar bisa mendapatkan *API Key* yang bisa didapatkan pada situs *web https://console.developers.google.com* agar bisa mendapatkan data komentar pada *youtube*.

```
# isikan dengan api key
api_key = 'A*****7gkE'
# contoh url video = https://www.youtube.com/watch?v=u5pI1fxseII
video_id = "u5pI1fxseII" #isikan dengan kode atau ID video
```

Data yang akan diambil berasal dari kanal *youtube* Kompas TV pada *link* berikut <https://www.youtube.com/watch?v=u5pI1fxseII> pada judul "Per 1 Januari 2024 Fotokopi KTP Sudah Tak Berlaku sebagai Syarat Berkas". Dan berhasil didapatkan data dengan rentan waktu dari 29 Desember 2023 sampai 30 Januari 2024, dengan jumlah data yang didapatkan yaitu 7166, dengan 4 kolom yaitu *publishedAt* berisikan tanggal komentar, *authorDisplayName* berisikan *user* yang memberi komentar, *textDisplay* isi dari komentar tersebut, dan *likeCount* berisikan *user* yang menyukai komentar tersebut, untuk mempermudah proses berikutnya dilakukan pengambilan 1 kolom data yang dibutuhkan yaitu kolom *textDisplay* berisikan komentar, sebagai contoh berikut dalam gambar 3:



Gambar 3. Pengumpulan data

Selanjutnya dilakukan pembersihan data yang dimulai dengan pengecekan data duplikat dan *missing value*, dan terdeteksi 89 data duplikat serta 0 *missing value* yang akan dihapus. Menunjukkan hasil pembersihan data yang sebelumnya sebesar 7166 baris menjadi 7077 baris.

## 4.2 Hasil Text Preprocessing

*Text preprocessing* merupakan tahapan penting pada pengolahan teks yang meliputi beberapa langkah untuk membersihkan serta mengganti teks mentah menjadi format yang lebih mudah untuk diproses. Berikut langkah-langkah pada *text preprocessing*.

### 4.2.1 Cleansing Data

Penghapusan kesalahan, ketidak sesuaian dan ketidak akuratan dalam suatu set data. Mulai dari menghilangkan URL, menghilangkan angka, menghilangkan karakter tanda baca, menghilangkan emoji, menghilangkan spasi berlebihan, menghilangkan karakter *non-Alphanumeric*, terlihat pada *syntax* dibawah ini.

```
def clean_text(text):
text = re.sub(r"@[A-Za-z0-9_]+", "", text) # Menghapus Mention
text=re.sub(r"#\w+", "", text) # Menghapus Hashtag
text=re.sub(r"https?:\/\/\S+", "", text) # Menghapus URL
text=re.sub(r"^[A-Za-z0 9]", "", text) # Menghapus Karakter Non Alphanumeric
text=re.sub(r"\s+", "", text).strip() # Menghapus Spasi Berlebih
text=re.sub(r"[0-9]+", "", text) # Menghapus Angka
```

Dari hasil *cleansing* data didapatkan kata pada Tabel 1.

Tabel 1. Hasil cleansing data

Sebelum	Sesudah
Tambah tahun ,malah tambah ribet. Ditengah tengah harga pangan meningkat, pajak meningkat.. aturan juga meningkat. Lihat dulu rakyat ini...	Tambah tahun malah tambah ribetbrDitengah tengah harga pangan meningkat pajak meningkat aturan juga meningkatbrLihat dulu rakyat ini
Sosialisasi dulu yg bener ke warga masyarakat, biar seluruh masyarakat benar <sup>2</sup> faham, pemerintah jgn terburu <sup>2</sup> ....	Sosialisasi dulu yg bener ke warga masyarakat biar seluruh masyarakat benar faham pemerintah jgn terburu

### 4.2.2 Case folding

Proses transformasi data huruf-huruf pada teks dirubah menjadi bentuk kecil atau *lowercase*. *Syntax* dibawah menunjukkan proses mendefinisikan sebuah fungsi bernama '*case\_folding*' yang menerima parameter '*text*' parameter tersebutlah yang akan diubah menjadi huruf kecil. Lalu '*text.lower*' yang akan mengembalikan teks semuanya dalam huruf keci,

```
def case_folding(text):
return text.lower() # Mengubah teks menjadi lowercase
```

Bertujuan untuk memastikan bahwa perbandingan antara teks tidak terpengaruh oleh perbedaan antara huruf besar dan huruf kecil, yang diuraikan Tabel 2 di bawah ini:

Tabel 2. Hasil casfolding

Sebelum	Sesudah
Tambah tahun malah tambah ribetbrDitengah tengah harga pangan meningkat pajak meningkat aturan juga meningkatbrLihat dulu rakyat ini	tambah tahun malah tambah ribetbrditengah tengah harga pangan meningkat pajak meningkat aturan juga meningkatbrlihat dulu rakyat ini
Sosialisasi dulu yg bener ke warga masyarakat biar seluruh masyarakat benar faham pemerintah jgn terburu	sosialisasi dulu yg bener ke warga masyarakat biar seluruh masyarakat benar faham pemerintah jgn terburu

### 4.2.3 Word Normalization

Tahapan *normalization* ialah proses dimana kata singkatan, salah ejaan atau kata yang kurang tepat penulisannya didalam teks diperbaiki menjadi kata yang sesuai. *Normalization* ini menggunakan sebuah file referensi yang dibuat dengan berisikan kata-kata singkatan dan kata-kata dalam dataset yang telah diperbaiki.

```
normalize = pd.read_csv('/content/drive/MyDrive/TA/normalisasi.csv')
Melakukan normalisasi
def text_normalize(text):
    text = ' '.join([str(normalize[normalize['singkat'] ==
word]['hasil'].values[0])
    # Tekx diatas jika kata ditemukan di kolom 'singkat', ambil kata hasil
normalisasi dari kolom 'hasil' dan ubah ke string
    if (normalize['singkat'] == word).any() # Memeriksa apakah kata ada di kolom
'singkat'
    else word for word in text.split()) # tidak ditemukan, gunakan kata asli
    text = str.lower(text) # Mengubah karakter dalam teks menjadi huruf kecil
    return text # Mengembalikan teks yang telah dinormalisasi
```

Dari *syntax normalization* diatas melakukan proses membaca file yang bernama "*normalisasi.csv*" dimana file referensi tersebut yang berisikan kata-kata. Selanjutnya memisahkan teks membagi teks input menjadi daftar kata dengan metode '*text.split()*'. Menormalkan kata untuk setiap kata tersebut dalam kolom singkat pada dataframe *normalize*. Pada hasil proses ini kata-kata singkatan atau kata yang salah ejaan akan diganti dengan kata yang sesuai pada file referensi yang telah dibuat, untuk contoh dapat dilihat pada Tabel 3 dibawah ini.

Tabel 3. Hasil normalization

Sebelum	Sesudah
tambah tahun malah tambah ribetbrditengah tengah harga pangan meningkat pajak meningkat aturan juga meningkatbrlihat dulu rakyat ini	tambah tahun bahkan tambah ribet ditengah tengah harga pangan meningkat pajak meningkat aturan juga meningkat lihat dulu rakyat ini
sosialisasi dulu yg bener ke warga masyarakat biar seluruh masyarakat benar faham pemerintah jgn terburu	sosialisasi dulu yang benar ke warga masyarakat biar seluruh masyarakat benar paham pemerintah jangan terburu

### 4.2.4 Tokenisasi

Memecah teks menjadi token-token (kata-kata) memisahkan kumpulan kata yang berdiri sendiri digunakan sebagai token atau potongan kata tunggal dalam kalimat, paragraf, atau halaman didalamnya. Untuk *tokenisasi* ini dapat dilihat pada *syntax* dibawah dimana sebuah fungsi bernama '*tokenize\_text*' menerima parameter '*text*' yang akan ditokenisasi. '*Word\_tokenize(text)*' ialah fungsi yang memisahkan teks input menjadi kata-kata (token). Hasil dari *tokenisasi* diuraikan pada Tabel 4.

```
# Menggunakan fungsi word_tokenize untuk membagi teks menjadi token (kata)
def tokenize_text(text):
    return word_tokenize(text)
```

Tabel 4. Hasil tokenisasi

Sebelum	Sesudah
tambah tahun bahkan tambah ribet ditengah tengah harga pangan meningkat pajak meningkat aturan juga meningkat lihat dulu rakyat ini	[tambah,tahun,bahkan,tambah,ribet,ditengah, tengah,harga,pangan, meningkat,pajak, meningkat,aturan,juga,meningkat,lihat,dulu, rakyat,ini]
sosialisasi dulu yang benar ke warga masyarakat biar seluruh masyarakat benar paham pemerintah jangan terburu	[sosialisasi,dulu,yang,benar,ke,warga,masyarak at,biar,seluruh,masyarakat,benar,paham,pemeri ntah,jangan, terburu]

#### 4.2.5 Stopword

Berikutnya penghapusan tahap dimana kata-kata yang sering muncul atau kurang menunjukkan relevansi dengan teks dihilangkan.

```
def remove_stopwords(text):
    list_stopwords = stopwords.words('indonesian')
    # Mengambil daftar stopwords standar dari NLTK untuk bahasa Indonesia
    txt_stopword = pd.read_csv("/content/drive/MyDrive/TA/add_stopwords.txt",
names = ["stopwords"],header=None)
    # Membaca file tambahan stopwords dari CSV tanpa header
    list_ \stopwords. extend (txt_stopword["stopwords"][0].split(' '))
    # Menambahkan file ke dalam daftar stopwords
    return [word for word in text if word not in list_stopwords]
    # Mengembalikan daftar kata yang bukan stopwords
```

Dari *syntax stopword* diatas dengan fungsi '*stopwords.words*' menggunakan *library* NLTK untuk memanggil *stopwords* dalam bahasa indonesia. Dengan tambahan file yang dipanggil yang berisikan *stopword* tambahan kata-kata, *stopword* tambahan akan ditambahkan kedalam '*list\_stopwords*' yang sudah ada dengan menggunakan metode '*extend*', dan menghasilkan Tabel 5 berikut ini.

Tabel 5. Hasil stopword

Sebelum	Sesudah
[tambah,tahun,bahkan,tambah,ribet, ditengah,tengah,harga,pangan,meningkat,p ajak,meningkat,aturan,juga,meningkat,lihat, dulu, rakyat,ini]	[ribet,ditengah,harga,pangan,meningkat, pajak,meningkat,aturan,meningkat,lihat]
[sosialisasi,dulu,yang, benar,ke,warga, masyarakat,biar,seluruh,masyarakat,benar, paham,pemerintah, jangan,terburu]	[sosialisasi,warga,biar, paham,pemerintah, terburu]

#### 4.2.6 Stemming

*Library Sastrawi* merupakan fitur *stemming* bahasa Indonesia yang digunakan untuk membantu *stemming* menghilangkan imbuhan pada kata untuk mendapatkan kata dasar.

```
# Membuat objek factory untuk stemmer
factory = StemmerFactory() # Membuat instance dari StemmerFactory
# Membuat objek stemmer dari factory
stemmer = factory.create_stemmer() # Menggunakan factory
# Melakukan stemming pada setiap token dalam daftar
def stemming(tokens):
    stemmed_tokens = [stemmer.stem(token) for token in tokens]
    # Menggunakan metode stem pada setiap token dalam daftar
    return stemmed_tokens
    # Mengembalikan daftar token yang sudah di-stem
```



*Syntax stemming* yang diatas menjelaskan dari mulai *mengimport 'stemmerFactory'* dari *library sastrawi*, *'factory = StemmerFactory()'* untuk membuat objek *'StemmerFactory'* digunakan menghasilkan *stemmer*, lalu untuk melakukan proses digunakan *'stemmer = factory.create\_stemmer()'*. Kemudian fungsi *'stemming'* menerima parameter *'token'* menerapkan metode *'stem'* dari objek *'stemmer'* pada setiap kata token, *'stem'* mengubah kata menjadi bentuk dasarnya. Yang dihasilkan pada Tabel 6.

Tabel 6. Hasil stemming

Sebelum	Sesudah
[ribet,ditengah,harga,pangan,meningkat,pajak ,meningkat,aturan, meningkat,lihat]	[ribet,tengah,harga,pangan,tingkat,pajak,tin gkat,atur, tingkat,lihat]
[sosialisasi,warga,biar, paham,pemerintah, terburu]	[sosialisasi,warga,biar, paham,perintah,buru]

### 4.3 Hasil Pemberian Label

Hasil pada pemberian label menggunakan *google colab* bahasa pemrograman *python* dengan bantuan *library Textblob*. Proses pelabelan data dilakukan dengan menganalisa teks untuk menentukan polaritas positif, negatif dan netral dikarenakan jumlah dataset yang banyak maka dilakukam proses tersebut agar mempermudah pengerjaan. Sebelum data diberi label, data diubah menjadi bahasa inggris menggunakan *library GoogleTranslator* karna *textblob* hanya mengenali bahasa inggris, lalu *mengimport library textblob* untuk menganalisis sentimen, kemudian mengambil data pada teks *english*, dengan contoh *syntax translate* dibawah.

```
translator = GoogleTranslator()#Membuat objek GoogleTranslator
def convert_eng(tweet):
    if pd.isna(tweet): # Memeriksa apakah tweet adalah NaN (Not a Number)
        return "" # Jika tweet adalah NaN, mengembalikan string kosong
    else:
        return translator.translate(tweet, lang_tgt='en')
        # Jika tweet bukan NaN, menerjemahkan tweet ke bahasa Inggris
```

Lalu diberi polaritas dengan bantuan *library Textblob* untuk menghitung jumlah variabel teks positif, negatif, dan netral, kemudian *looping* jika polaritas lebih besar dari 0 maka positif bila sama dengan 0 maka netral jika kurang dari 0 maka negatif. Dengan hasil pelabelan sebagai berikut ditampilkan dalam Gambar 2.

clean_data	english_tl	sentimen
pilih program biar gampang gonta ganti data gelap	choose a program to make it easy to replace da...	Positif
ribet	complicated	Negatif
tanbah ribet iya android bodoh	It's more complicated, stupid Android	Negatif
bikin paspor sogok salah korupsi hidup pegawai	making passports, bribery, corruption of emplo...	Netral
enak iya kuasa atur	it's the power of order	Netral
...	...	...
wajib loh fotocopy	photocopy is mandatory	Netral
wow maju	wow forward	Positif
nomor pokok mantap	steady master number	Positif
golong putih	white	Netral

Gambar 2. Pemberian label

### 4.4 Hasil TF-IDF

Dari hasil TF-IDF pada Gambar 3, merupakan hasil penggunaan TF-IDF untuk menghitung skor pembobotan dari kolom kata, dan kolom tf-idf memunculkan hasil kata yang sudah melewati pembobotan kata, sedangkan kolom frekuensi menunjukkan kemunculan kata dalam dokumen, dan kata ribet memiliki nilai pembobotan terbesar.

Kata	tf-idf	Frekuensi
ribet	195.328899	1
susah	98.485122	1
sulit	90.441125	2
kasih	78.283902	1
fotocopy	75.142908	1
foto	73.151605	1
mudah	67.429619	1
aman	63.601587	4
laku	61.777967	1
rumit	60.366839	3

Gambar 3. Hasil nilai TF-IDF

Proses ini dimulai *mengimport* 'TfidfVectorizer' dari 'sklearn' dan membersihkan teks dari nilai *NaN*. Kemudian TF-IDF *Vectorizer* digunakan untuk menghitung nilai TF-IDF dari teks, yang diubah menjadi *array* dan memasukkannya ke dalam DataFrame. Nilai TF-IDF untuk setiap kata dijumlahkan, disusun, dan frekuensi kata ditambahkan ke DataFrame akhir. Hasilnya adalah daftar kata yang disusun berdasarkan nilai TF-IDF tertinggi, yang terlihat pada *syntax* dibawah.

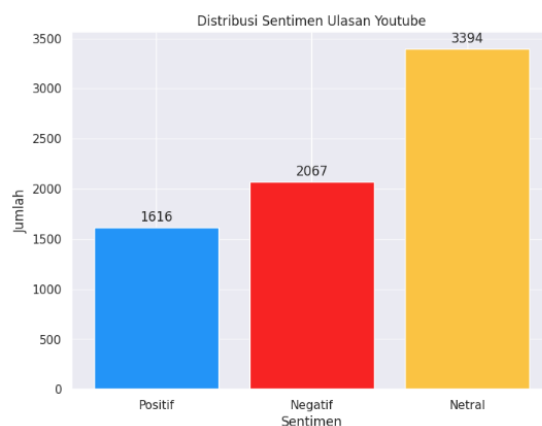
```
# Inisialisasi TF-IDF Vectorizer
tf_idf = TfidfVectorizer(ngram_range=(1, 1))
# Fit dan transform data teks
tf_idf_matrix = tf_idf.fit_transform(X)
# Ubah TF-IDF matrix ke array
X_tf_idf_array = tf_idf_matrix.toarray()
# Buat DataFrame dengan array TF-IDF dan nama kolom sesuai kata-kata
data_tf_idf = pd.DataFrame(X_tf_idf_array,
columns=tf_idf.get_feature_names_out())
# Jumlahkan nilai TF-IDF untuk setiap fitur
tf_idf_sums = data_tf_idf.sum(axis=0)
# Buat DataFrame dengan kata dan nilai TF-IDF
sorted_tf_idf = pd.DataFrame({
    'Kata': tf_idf_sums.index,
    'tf-idf': tf_idf_sums.values
}).sort_values(by='tf-idf', ascending=False).reset_index(drop=True)
# Tambahkan kolom frekuensi kata (jumlah kemunculan kata)
word_counts = (X_tf_idf_array > 0).sum(axis=0)
sorted_tf_idf['Frekuensi'] = word_counts # Sinkronkan frekuensi dengan kata
```

#### 4.5 Hasil Klasifikasi Naïve Bayes

20% data uji dan 80% data latih, dengan fitur "*x\_latih\_dicontohkan*" dan "*y\_latih\_dicontohkan*" menjadi tag stempel. Parameter '*test\_size=0,2*' menentukan total data pengujian, sedangkan 80% untuk pelatihan dengan parameter '*random\_state=42*', terlihat pada *syntax* klasifikasi *naïve bayes* dibawah ini.

```
# Perbandingan 80% untuk data latih, 20% untuk data uji
X_train, X_test, y_train, y_test = train_test_split(X_train_resampled,
y_train_resampled, test_size = 0.2, random_state = 42)
```

Gambar 4 mengilustrasikan hasil klasifikasi. Memberikan informasi bahwa dari hasil klasifikasi data menggunakan parameter yang menentukan ukuran data uji dalam presentase keseluruhan dataset, dilakukan terhadap 7077 data ulasan yang dipisahkan menjadi 3 polaritas yakni sentimen negatif, positif, dan netral. Menunjukkan hasil pada sentimen positif 1616 ulasan, 2067 memberikan ulasan negatif, dan sisanya 3394 ulasan netral.



Gambar 4. Hasil klasifikasi

#### 4.6 Hasil Visualisasi

Setelah Klasifikasi *Naïve Bayes* dilakukan, pada tahapan visualisasi data dengan *word cloud*, akan menciptakan istilah yang sering ada pada *text* dan kata penting dari data opini dalam sentimen positif, netral, dan negatif. Dengan proses semua *text* dalam kolom '*clean\_data*' digunakan menjadi satu *string*, kemudian menggunakan *library wordcloud*, setelah itu menampilkan gambar menggunakan '*matplotlib*', dengan *syntax* visualisasi dibawah ini.

```
all_words = ' '.join([reviews for reviews in data_negatif['clean_data']])
wordcloud = WordCloud(colormap='Reds', width=1000, height=1000,
                      mode='RGBA',
                      background_color='white').generate(all_words)
```

Pada gambar 5 dibawah memberikan hasil *word cloud* dari data opini, menunjukkan visualisasi negatif dengan kata yang banyak yaitu ribet, susah, sulit, yang dapat diartikan bahwa hal-hal tersebut yang dikeluhkan masyarakat.



Gambar 5. Visualisasi negatif

Dari Gambar 6 dibawah yang bersentimen positif didominasi kata mudah, aman, dan bagus. Bahwa dalam pembelian gas 3 Kg menggunakan KTP, berangan-angan dengan kemudahan bila dibandingkan dengan sebelumnya.



Gambar 6. Visualisasi positif

Sedangkan pada Gambar 7 pada visualisasi netral kata yang muncul yakni *fotocopy*, uang.



Gambar 7. Visualisasi netral

#### 4.7 Hasil Evaluasi

Tujuan evaluasi adalah untuk menilai kinerja suatu sistem dalam penelitian ini, *Classification report* digunakan untuk menilai keakuratan model klasifikasi. *precision*, *recall*, dan *f1-score* adalah tiga metode evaluasi umum yang digunakan untuk mengevaluasi keakuratan model klasifikasi. Proses pertama '*accuracy\_score(y\_test,y\_pred)*' digunakan untuk menghitung akurasi model, membandingkan label '*y\_test*' dengan label prediksi '*y\_pred*', selanjutnya '*classification\_report(y\_test, y\_pred, target\_names= [positif,netral,negatif])*' menghasilkan klasifikasi *precision,recall,f1-score*, ditampilkan pada *syntax* dibawah ini.

```
accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report
(y_test, y_pred, target_names = ['Positif', 'Netral', 'Negatif'])
```

Gambar 8 dibawah memberikan hasil gambaran mencakup metrik evaluasi untuk setiap kelas, parameter.

Laporan Klasifikasi :				
	precision	recall	f1-score	support
Positif	0.79	0.93	0.85	675
Netral	0.94	0.63	0.75	651
Negatif	0.84	0.95	0.89	711
accuracy			0.84	2037
macro avg	0.85	0.83	0.83	2037
weighted avg	0.85	0.84	0.83	2037

Gambar 8. Evaluasi classification report

Dari Gambar 8 perbandingan data *testing* dan *training* 80-20 menunjukkan hasil *Classification report* mendapatkan pengujian mendapatkan *precision* sebesar 0,84 pada data sentimen negatif, sentimen netral 0,94, serta 0,79 disentimen positif. *Recall* sebesar 0,95 pada sentimen negatif, sentimen netral 0,63 dan 0,93 disentimen positif, serta *f1-score* sebesar 0,89 pada data sentimen negatif, sentimen netral 0,75 dan 0.85 pada sentimen positif. Pada akurasi mendapatkan 0.84.

```
[[625 14 36]
 [146 408 97]
 [ 20 14 677]]

True Positives (TP) = 625
True Negatives (TN) = 408
False Positives (FP) = 14
False Negatives (FN) = 146
```

Gambar 9. Confusion matriks

Dari Gambar 9 dari total jumlah data yang diuji dan *confusion matriks* untuk mengevaluasi kinerja model klasifikasi tersebut yakni dengan *true* negatif 408 data yang diperdiksikan dengan benar sebagai negatif, serta *true* positif 625 data yang benar sebagai positif, *false* negatif 146 data yang seharusnya positif tetapi salah diprediksi sebagai negatif, serta *false* positif 14 data yang seharusnya negatif, tetapi salah diprediksi sebagai positif.

Berdasarkan penjelasan dari penelitian Analisis Sentimen Kebijakan Pembelian Gas 3 Kg Menggunakan KTP dengan Algoritma Naive Bayes, didapatkan hasil bahwa dengan membagi data testing menjadi 80% data training dan 20% data testing, mendapatkan akurasi sebesar 84%. Data hasil klasifikasi Naive Bayes menunjukkan bahwa data yang paling banyak adalah data netral, namun karena netral merupakan nilai yang bias, maka tidak dapat ditentukan sebagai pembuatan kebijakan pembelian gas 3 kg menggunakan KTP. Hasilnya, sentimennya adalah negatif atau penolakan terhadap kebijakan tersebut. Penelitian Winda tentang Klasifikasi Sentimen Publik terhadap Kebijakan Kartu Prakerja Indonesia yang dilakukan pada tahun 2020 juga menghasilkan penolakan terhadap kebijakan tersebut[7]. Hasil penelitian menggunakan Naive Bayes cukup akurat sebagai mana penelitian oleh Rahmat tahun 2024 pada Sentimen Aplikasi Identitas Kependudukan Digital (IKD).

## 5. Simpulan

Dari hasil penelitian diatas bahwa kanal Kompas TV menggunakan metode *crawling comment youtube* untuk pengumpulan data. Mekanisme *text preprocessing* semacam *cleaning, case folding, normalisasi, tokenisasi, stopword*, serta *stemming* dilakukan oleh 7166 data yang berhasil didapatkan. Nilai sentimen positif, negatif, dan netral untuk data prapemrosesan teks kemudian ditentukan dengan menggunakan *textblob* untuk memberi label. Prosedur pembobotan kata dengan menggunakan *tf-idf* adalah langkah selanjutnya, diikuti dengan klasifikasi Naive Bayes. Akurasi metode klasifikasi Naive Bayes adalah 84%, *f1-score* 83%, presisi 85%, dan *recall* 84%. Menunjukkan bahwa pada model ini dapat mengidentifikasi sentimen dalam komentar dengan cukup akurat. Pihak-pihak terkait dapat menggunakan hasil analisis ini untuk menentukan kebijakan, memahami respon masyarakat, dan mengambil tindakan yang tepat. Selain itu, penelitian ini memberikan peluang untuk pengembangan lebih lanjut, seperti menggunakan algoritma klasifikasi tambahan untuk mendapatkan hasil yang lebih baik dan memproses lebih banyak data pada tahap prapemrosesan teks.

## Daftar Referensi

- [1] L. T. Tysara, "Alasan Beli Gas LPG 3 Kg Wajib Pakai KTP Mulai 1 Januari 2024, Simak Cara Belinya," *liputan6.com*, 2024. <https://www.liputan6.com/hot/read/5494923/alasan-beli-gas-lpg-3-kg-wajib-pakai-ktp-mulai-1-januari-2024-simak-cara-belinya> (accessed Jan. 05, 2024).
- [2] V. K. Ningsih and S. Syalikha, "Implementasi Subsidi Listrik untuk Mendorong Pencapaian SDGs Tujuan 7," *J. Econ. Assets, Eval.*, vol. 1, no. 4, pp. 1–12, 2024.
- [3] T. Nurdiana, "Catat Ini: Pembelian Gas Melon Subsidi Dibatasi Mulai 1 Januari 2024," *insight.kontan*, 2023. <https://insight.kontan.co.id/news/catat-ini-pembelian-gas-melon-subsidi-dibatasi-mulai-1-januari-2024> (accessed Jan. 10, 2024).
- [4] S. Alim and A. F. Dharma, "Youtube Sebagai Ruang Publik Alternatif Bagi Anak Muda," *J. Ilmu Komunikasi*, vol. 4, no. 1, pp. 1–13, 2021.
- [5] F. V. Sari, "Analisis Sentimen Pelanggan Toko Online JD . ID Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi," *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [6] H. D. Prasetyo, T. Pramiyati, and I. N. Isnainiyah, "Sentimen Analisis Pengguna Twitter Terhadap Kebijakan Merdeka Belajar Menggunakan Algoritma Naive Bayes," *SENAMIKA*, no. April, pp. 559–568, 2021.
- [7] W. P. Anggraini and M. S. Utami, "Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Pekerja Di Indonesia," *Fakt. Exacta*, vol. 13, no. 4, p. 255, 2021, doi: 10.30998/faktor\_exacta.v13i4.7964.
- [8] R. Hidayat, R. Nur Rahman, M. Reifin Perdana, and Arbansyah, "Analisis Sentimen Aplikasi Identitas Kependudukan Digital (IKD) Menggunakan Metode Naive Bayes," *J. Sist. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 129–140, 2024, [Online]. Available: <https://doi.org/10.59581/jusiik-widyakarya.v2i1.2320>
- [9] K. A. Lubis and A. Yudertha, "Analisis Sentimen Opini Masyarakat Terhadap Pindahnya \

- Ibu Kota Indonesia Dengan Menggunakan Klasifikasi Naïve Bayes,” *TEKNOINFO*, vol. 18, pp. 226–238, 2024.
- [10] N. Susanti, E. Ronando, N. A. Basyarach, D. Harini, Sulistyawati, and W. Widiasih, “Analysis of the effect of the MBKM Internship Program and Certified Independent Study (MSIB) on university performance Universitas 17 Agustus 1945 Surabaya,” *Tech. Soc. Sci. J.*, vol. 6, no. December, pp. 101–105, 2020.
- [11] B. Hakim, “Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning Data Text Pre-Processing Sentiment Analysis In Data Mining Using Machine Learning School of Computer Science and Technology , Harbin Institute of Technology,” vol. 4, no. 2, pp. 16–22, 2021.
- [12] I. Ahmad, S. Samsugi, and Y. Irawan, “Implementasi Data Mining Sebagai Pengolahan Data,” *J. Teknoinfo*, vol. 16, no. 1, p. 46, 2022, [Online]. Available: <http://portaldata.org/index.php/portaldata/article/view/107>
- [13] I. Gunawan, “Implementasi Algoritma Sentencepiece untuk Meningkatkan Performa Naive Bayes Classifier pada Klasifikasi Artikel Berita,” *J. Tek. Inform.*, pp. 20–26, 2023, [Online]. Available: <https://kc.umn.ac.id/id/eprint/16831>
- [14] M. U. Albab, Y. Karuniawati, and M. N. Fawaiq, “Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic,” *J. Transform.*, vol. 20, no. 2, pp. 1–10, 2023, [Online]. Available: <https://journals.usm.ac.id/index.php/transformatika/page1>
- [15] M. Faiq, A. Putro, and E. B. Setiawan, “Analisis Sentimen Terhadap Kebijakan Pemerintah dengan Feature Expansion Metode GloVe pada Media sosial Twitter,” *e-Proceeding Eng.*, vol. 9, no. 1, pp. 54–66, 2022.
- [16] R. Nurhidayat and K. E. Dewi, “Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek,” *Komputa J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, pp. 91–100, 2023, doi: 10.34010/komputa.v12i1.9458.
- [17] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, “Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine,” *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/teknov10i2.419.
- [18] R. Adyatma Subagja, Y. Widiastiwi, and N. Chamidah, “Klasifikasi Ulasan Aplikasi Jenius pada Google Play Store Menggunakan Algoritma Naive Bayes,” *Inform. J. Ilmu Komput.*, vol. 17, no. 3, p. 197, 2021, doi: 10.52958/iftk.v17i3.3652.
- [19] R. Parlika, S. I. Pradika, A. M. Hakim, and K. R. N M, “Analisis Sentimen Twitter Terhadap Bitcoin Dan Cryptocurrency Berbasis Python Textblob,” *J. Ilm. Teknol. Inf. dan Robot.*, vol. 2, no. 2, pp. 33–37, 2020, doi: 10.33005/jifti.v2i2.22.
- [20] R. Wati, S. Ernawati, and H. Rachmi, “Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH,” *J. Manaj. Inform.*, vol. 13, no. 1, pp. 84–93, 2023, doi: 10.34010/jamika.v13i1.9424.
- [21] A. Ryansyah and S. Andayani, “Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen,” *J. Sist. Teknol. Inf. Komun.*, vol. 1, no. 1, pp. 58–62, 2022, [Online]. Available: <http://pdfbox.apache.org/>
- [22] A. M. Siregar, S. Faisal, and B. Widiharto, “Analisis Sentimen Masyarakat Terhadap Universitas Buana Perjuangan Karawang Dengan Algoritme SVM dan Naive Bayes,” *Pros. Konf. Nas. Penelit. Dan Pengabd. Univ. Buana Perjuangan Karawang*, vol. 3, no. 1, pp. 25–36, 2023, [Online]. Available: <https://journal.ubpkarawang.ac.id/index.php/ProsidingKNPP/article/view/4894>
- [23] R. Al Ghivary, N. Wulandari, N. Srikandi, D. A. Publik, and U. M. Jakarta, “Peran Visualisasi Data Untuk Menunjang Analisa Data The Role Of Data Visualisation To Support Population,” *J. Adm. Publik*, vol. 1, no. 1, pp. 57–62, 2023.
- [24] M. Hidayatullah *et al.*, “Sentiment Analysis of Police Performance On Twitter Users Using Naïve Bayes Method,” *RISTEC Res. Inf. Syst. Technol.*, vol. 2, no. 2, pp. 29–40, 2021, doi: 10.31980/ristec.v2i2.1945.