

PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN KELANGKAAN MINYAK GORENG PADA MEDIA SOSIAL TWITTER

Muhammad Nasrul Fadillah¹, Rini Nuraini Sukmana²

^{1,2} Teknik Informatika, Universitas Sangga Buana

¹ e-mail korespondensi: nasrulfadillah@gmail.com

ABSTRACT

Twitter is one of the most popular social media today. Recently, what has become a trending topic on Twitter is the issue of cooking oil scarcity, where the Indonesian people are currently facing a cooking oil crisis. Cooking oil is one of the basic needs in society. So that there are both positive and negative opinions on the issue. To find out the opinions on Twitter, what must be done is to process the tweet data to be classified into positive and negative opinions. One way of analyzing that can be applied is Sentiment Analysis. In namely the Naïve Bayes Classifier to get maximum accuracy results and to support the application in classifying tweet data on Twitter social media on the issue of scarcity of cooking oil.

Keywords: Sentiment Analysis, Cooking Oil, Naïve Bayes Classifier, Twitter

ABSTRAK

Twitter merupakan salah sebagian dari sosial media yang banyak digemari hingga saat ini. Pada akhir – akhir ini yang menjadi trending topic di Twitter yaitu mengenai isu kelangkaan minyak goreng, dimana masyarakat Indonesia saat ini sedang dihadapi krisis minyak goreng. Minyak goreng merupakan salah sebagian dari kebutuhan pokok di masyarakat. Sehingga bermunculan opini baik positif maupun negatif terhadap isu tersebut. Untuk mengetahui opini yang ada di Twitter maka yang harus dilakukan adalah mengolah data tweet untuk diklasifikasikan menjadi opini positif dan negatif. Salah satu cara dalam menganalisis yang dapat diaplikasikan yaitu Analisis Sentimen. Dalam melangsungkan analisis sentimen diperlukan suatu algoritma yang mumpuni yaitu Naïve Bayes Classifier untuk mendapatkan hasil akurasi yang maksimal serta untuk menunjang aplikasi tersebut dalam mengklasifikasikan data tweet pada media sosial Twitter pada isu kelangkaan minyak goreng.

Kata Kunci: Analisis Sentimen, Minyak Goreng, Naïve Bayes Classifier, Twitter

PENDAHULUAN

Seiring dengan perkembangan teknologi di tengah – tengah transisi menuju era teknologi digital yang begitu pesat, membuat berbagai macam aktivitas manusia bergantung kepada kemajuan teknologi. Pada era sekarang, dimana internet dijadikan acuan dalam menunjang setiap aktivitas yang dilakukan. Contohnya pada penggunaan media sosial yang digunakan oleh masyarakat umum yang merupakan bagian dari internet. Dalam hal ini,

penggunaan media sosial di masyarakat digunakan untuk mengekspresikan opini pribadi, mengungkapkan perasaan, dan juga hal lain yang menjadi perhatian mereka [1].

Twitter merupakan salah sebagian dari media sosial yang banyak digemari hingga saat ini. Twitter adalah layanan mikroblog yang memungkinkan penggunanya untuk membuat postingan yang disebut dengan tweet dengan batas maksimal hingga 140 karakter teks [2]. Melalui tweet, pengguna media sosial Twitter

bisa saling berinteraksi dan berbagi pendapat tentang apa yang akan disampaikan pada tweet tersebut. Tweet yang dibagikan di Twitter biasanya merupakan bagian dari topik yang sedang hangat dibicarakan dan terkadang menjadi pembahasan populer di Twitter [3].

Seperti pada akhir – akhir ini yang menjadi trending topic di Twitter yaitu mengenai isu kelangkaan minyak goreng, dimana masyarakat Indonesia saat ini sedang dihadapi krisis minyak goreng. Minyak goreng merupakan salah sebagian dari kebutuhan pokok di masyarakat, sehingga bermunculan opini – opini baik positif maupun negatif terhadap isu tersebut. Untuk mengetahui opini yang ada di Twitter maka yang harus dilakukan adalah mengolah data tweet untuk diklasifikasikan menjadi opini yang positif dan negatif [4].

Salah satu cara dalam menganalisis yang dapat diaplikasikan yaitu Analisis Sentimen. “Analisis Sentimen bertujuan untuk mendapatkan sebuah uraian tentang kualitas sebuah layanan, apakah layanan tersebut cenderung mendapatkan nilai positif, negatif maupun netral” [5].

Berdasarkan latar belakang yang telah diuraikan di atas, dalam melangsungkan analisis sentimen, diperlukan suatu algoritma yang mumpuni untuk mendapatkan hasil akurasi yang maksimal. Maka digunakanlah algoritma Naïve Bayes Classifier (NBC) untuk menunjang aplikasi tersebut dalam mengklasifikasikan data *tweet* pada media

sosial Twitter pada isu kelangkaan minyak goreng.

METODE

Tahapan Text Mining

Analisis sentimen yang digunakan pada penelitian ini memanfaatkan metode *Text Mining* dengan menggunakan *keyword* ‘minyak goreng’ dalam melakukan crawling data yang diambil sebanyak 600 data tweet, yang kemudian akan dilakukan pengklasifikasian terhadap data yang dikumpulkan [6]. Adapun tahapannya sebagai berikut [7] :

1) *Text Preprocessing*

Pada langkah ini, kita memiliki data yang telah dikumpulkan dan selanjutnya data tersebut akan melalui beberapa tahap. Terdapat beberapa tahap proses dalam *Text Preprocessing*, yaitu :

a. *Cleaning*

Cleaning merupakan langkah untuk membersihkan dokumen dari kata atau karakter yang tidak digunakan untuk mengurangi gangguan pada proses pengklasifikasian. Kata yang dihilangkan seperti karakter *Hashtag* (#), *Username* (@nama), URL ([HTTP://.com](http://.com)), dan *emoticon*.

b. *Case Folding*

Case Folding merupakan langkah penyamaan bentuk huruf dengan merubah setiap huruf menjadi huruf kecil, serta menghilangkan tanda baca

juga angka, Huruf yang digunakan antara huruf a sampai dengan z

c. *Tokenzing*

Tokenzing merupakan langkah pemotongan string berdasarkan setiap kata yang menyusunnya, dapat diartikan juga sebuah langkah untuk memisahkan dokumen atau kalimat menjadi sebuah kata dengan melakukan analisis terhadap gabungan kata dengan memecah setiap kata tersebut.

d. *Filtering*

Filtering memiliki tujuan untuk menghilangkan atau membuang stopwords atau kata – kata yang tidak memiliki makna.

Naïve Bayes Classifier

Naïve Bayes dimanfaatkan untuk memprediksi data dengan menggunakan probabilitas sederhana yang bersumber pada teori Bayes dengan independensi atau ketidakterikatan. Maka dari itu, disimpulkan bahwa Naïve Bayes menggunakan model independen dalam melakukan klasifikasi data [8].

Pada model Naïve Bayes, independensi atau ketidakterikatan terhadap fitur merupakan sebuah sifat pada sebuah dataset tidak berhubungan dengan ada atau tidaknya fitur lain dalam dataset yang sama [9]. Prediksi Bayes bersumber dari teori Bayes dengan rumus sebagai berikut :

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \dots \dots \dots (1)$$

Pembahasan :

1. P(H|E) Probabilitas akhir bersyarat, merupakan suatu hipotesis H terjadi jika diberikan bukti E terjadi.
2. P(E|H) Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H.
3. P(H) adalah probabilitas awal (Priori) hipotesis H terjadi tanpa memandang bukti apapun.
4. P(E) Probabilitas awal (Priori) bukti E terjadi tanpa memandang hipotesis/bukti lain.

Confusion Matrix

Confusion Matrix merupakan metode yang digunakan untuk menghitung performa dari sebuah model. Setiap baris pada matrix, merepresentasikan kelas aktual dari data, dan setiap kolom merepresentasikan kelas prediksi dari data [10].

Tabel 1 : Confusion Matrix

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	True Negatif (TN)	False Positif (FP)
Aktual Positif	False Negatif (FN)	True Positif (TP)

1. *True Positif* = Jumlah banyak data aktual kelas positif, dan model memprediksi positif.

2. *True Negatif* = Jumlah banyak data aktual kelas positif, dan model memprediksi negatif
3. *False Positif* = Jumlah banyak data aktual kelas negatif, dan model memprediksi positif.
4. *False Negatif* = Jumlah banyak data aktual kelas positif, dan model memprediksi negatif.

Berdasarkan 4 data di atas, dapat diperoleh perhitungan untuk menghitung performa dari sebuah model, yaitu :

1. Accuracy = Jumlah rasio prediksi benar terhadap seluruh data.

$$\frac{TP + TN}{TP + TN + FP + FN} \dots \dots \dots (2)$$

2. Precision = Saat model memprediksi positif, seberapa sering prediksi itu benar

$$\frac{TP}{FP + TP} \dots \dots \dots (3)$$

3. Recall = Ketika kelas aktual positif, seberapa sering model memprediksi positif.

$$\frac{TP}{FN + TP} \dots \dots \dots (4)$$

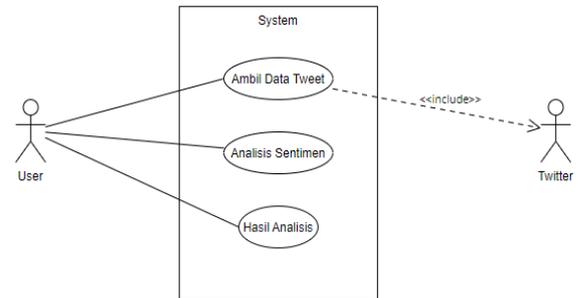
HASIL DAN PEMBAHASAN

Sistem yang hendak dibangun terhadap penelitian ini merupakan sistem untuk menilai sentimen masyarakat dengan menerapkan metode Naïve Bayes Classifier untuk pengklasifikasian data *tweet* yang akan digunakan. *Tweet* yang digunakan bersumber dari Twitter yang diunggah oleh masyarakat

untuk menilai sentimen pada kasus kelangkaan minyak goreng.

Sistem yang Diusulkan

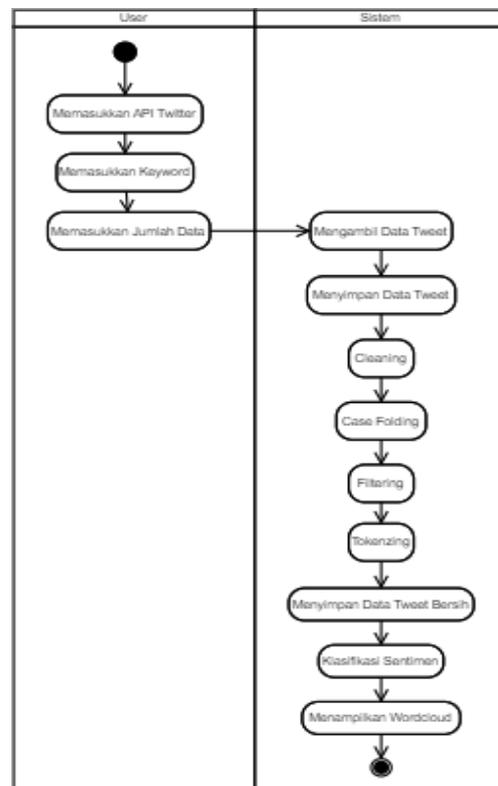
a. Use Case Diagram



Gambar 1 : Use Case Diagram

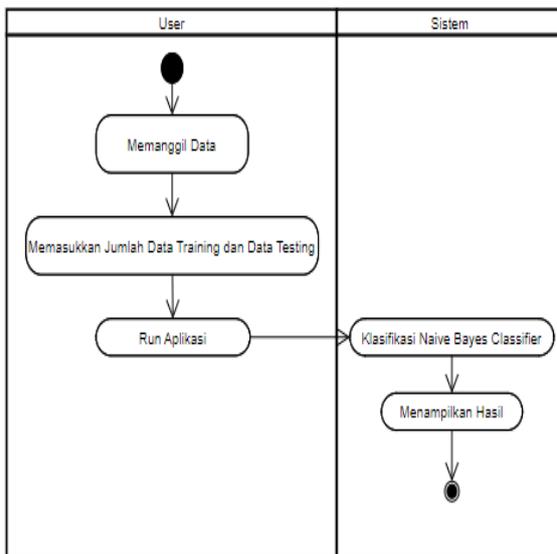
Pada model *Use Case Diagram* di atas, dapat kita lihat bahwa user dapat menjalankan seluruh *Activity* seperti mengambil data *tweet*, Analisis Sentimen serta *view* Hasil Analisis.

b. Activity Diagram



Gambar 2 : Activity Diagram User

Pada model *Activity Diagram* di atas dapat dilihat bahwa user melakukan aktivitas seperti memasukkan API Twitter, lalu memasukkan *keyword*, dan memasukkan jumlah data *tweet* yang akan diambil. Selanjutnya sistem yang akan mengolah perintah dari user seperti mengambil data *tweet*, menyimpan data *tweet*, melakukan *cleaning data*, menyimpan data yang sudah dibersihkan, melakukan klasifikasi sentimen, dan menampilkan *wordcloud*.



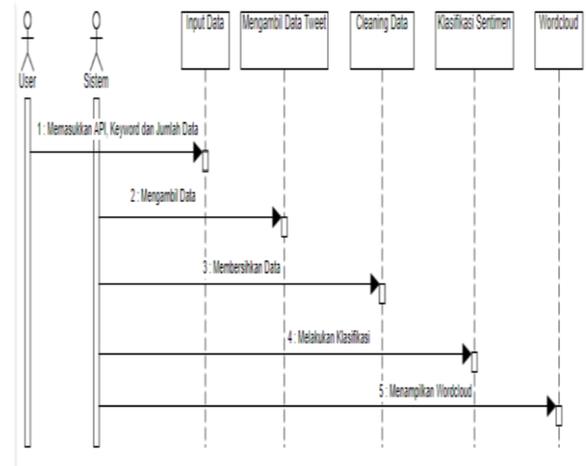
Gambar 3 : Activity Diagram NBC

Pada model *Activity Diagram* di atas dapat dilihat bahwa user memanggil data untuk dilakukan klasifikasi *Naïve Bayes*, lalu memasukkan jumlah data *training* dan *testing*, lalu run aplikasi, kemudian sistem akan mengolah untuk di klasifikasikan, dan menampilkan hasil.

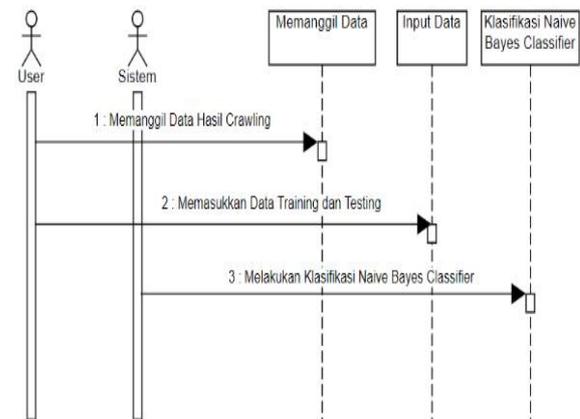
c. *Sequence Diagram*

Pada model *Sequence Diagram* di atas dapat dilihat bahwa peran user hanya menginputkan data seperti memasukkan kode API,

memasukkan *keyword*, memasukkan jumlah data. Kemudian sistem akan menjalankan perintah yang dimasukkan *user* dengan tahapan pengambilan data, *cleaning data*, klasifikasi sentimen dan *wordcloud*.



Gambar 4 : Sequence Diagram User



Gambar 5 : Sequence Diagram NBC

Pada model *Sequence Diagram* di atas dapat dilihat bahwa peran user hanya memanggil data, memasukkan data *training* dan *testing*, kemudian sistem yang akan menjalankan perintah untuk dilakukan *Klasifikasi Naïve Bayes Classifier*.

Implementasi

Pada tahap ini akan dilakukan implementasi, yaitu tahap untuk menampilkan hasil program Analisis Sentimen Twitter Kelangkaan Minyak Goreng yang dibuat dalam bahasa pemrograman R. Adapun hasil dari implementasi sistem sebagai berikut :



Gambar 6 : Tampilan Data Twitter 1

Tampilan data Twitter merupakan tampilan awal setelah dilakukannya *Crawling* Data, dimana setelah *Crawling* Data berhasil dilakukan, sistem akan menampilkan hasil dari langkah tersebut dimana disajikan berupa data dan juga setiap kata – kata yang ditulis didalam *tweet*. Data yang diambil sebanyak 600 data *Tweet*.

	text
1	g tingkatkan ekonomi pemerintah jamin harga minyak gore...
2	penggunaan peduli lindungi atau ktp sbg syarat beli minya...
3	pemerintah jamin ketersediaan minyak goreng curah dan ke...
4	harga minyak goreng murah g tingkatkan ekonomi
5	dukung kebijakan stabilisasi harga minyak goreng g tingkat...
6	petani sawit dukung kebijakan stabilisasi harga minyak gore...
7	g tingkatkan ekonomi indonesia dan bangladesh yang masi...
8	pak luhut dan erick thohir harusnya buat bumh baru khusu...
9	menteri perdagangan zulhas janji stabilkan harga minyak go...
10	beli minyak goreng pakai pedullindungi diduga akalakalan ...

Gambar 7 : Tampilan Data Twitter 2

Setelah *Crawling* Data berhasil lakukan, langkah selanjutnya adalah *Cleaning* Data dengan 4 tahapan seperti *Cleaning*, *Case Folding*, *Filtering*, dan *Tokenzing*.

Tabel 2 : Proses Cleaning

Sebelum	Sesudah
Dukung Kebijakan	Dukung
Stabilisasi Harga Minyak	Kebijakan
Goreng G20 Tingkatkan	Stabilisasi
Ekonomi	Harga Minyak
https://t.co/7v1AOaYarl	Goreng G20
	Tingkatkan
	Ekonomi

Tabel 3 : Proses Case Folding

Sebelum	Sesudah
Minyak goreng	Minyak goreng
langka mahal	langka mahal
mungkin biar semua	mungkin biar semua
orang pakai Air	orang pakai air fryer
fryer ya?	ya

Filtering

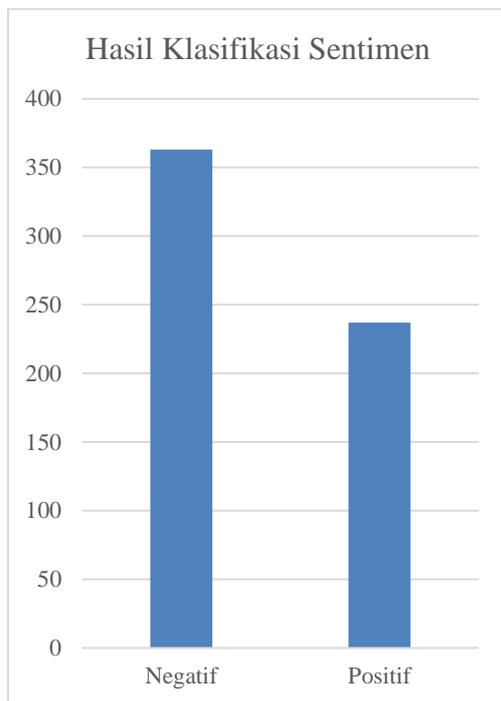
Membersihkan kata yang tidak memiliki makna. Contoh kata yang digunakan adalah “kalo”, “gak”, “org”, “bpk”, “haha”, “wkwk”.

Tabel 4 : Proses Tokenzing

Sebelum	Sesudah
Minyak goreng	“minyak”,
langka mahal	“goreng”, “langka”,
mungkin biar semua	“mahal”,
	“mungkin”, “biar”,

Sebelum	Sesudah
orang pakai air fryer ya	“semua”, “orang”, “pakai”, “air”, “fryer”, “ya”

Setelah tahapan tersebut dilakukan kemudian data akan disimpan dan setelahnya akan ditampilkan seperti pada gambar di atas.



Gambar 8 : Visualisasi Data Grafik

klasifikasi	Count of klasifikasi
Negatif	363
Positif	237

Gambar 9 : Visualisasi Data Angka

Visualisasi data merupakan tampilan setelah dilakukannya klasifikasi sentimen, dimana setelah data digolongkan ke dalam bentuk Positif dan Negatif, maka data akan diolah untuk dibuat dalam bentuk chart seperti pada gambar diatas. Terlihat dimana sentimen negatif lebih dominan dengan hasil sebanyak 363, dan sentimen positif sebanyak 237.



Gambar 10 : Wordcloud

word	freq
minyak	548
goreng	536
beli	332
rakyat	211
sawit	125
pedulilindungi	121
pakai	117
murah	112
kartu	110
sembako	109

Gambar 11 : Wordcloud Frequent

Wordcloud merupakan tampilan dimana kata – kata yang sering digunakan pada setiap tweet akan muncul dalam bentuk gambar seperti berikut, dimana – kata – kata dengan ukuran yang besar merupakan kata yang sering digunakan dan untuk kata – kata dengan ukuran yang kecil merupakan kata yang frekuensinya sedikit digunakan.

```
> conf.mat
Confusion Matrix and Statistics

          Reference
Prediction Negatif Positif
Negatif      88      18
Positif        2      33

      Accuracy : 0.8582
      95% CI   : (0.7895, 0.9112)
No Information Rate : 0.6383
P-Value [Acc > NIR] : 5.05e-09

      Kappa : 0.6704

McNemar's Test P-Value : 0.0007962

      Sensitivity : 0.9778
      Specificity : 0.6471
      Pos Pred Value : 0.8302
      Neg Pred Value : 0.9429
      Prevalence : 0.6383
      Detection Rate : 0.6241
      Detection Prevalence : 0.7518
      Balanced Accuracy : 0.8124

      'Positive' Class : Negatif
```

Gambar 12 : Hasil Uji Akurasi

Pada tahap ini, dilakukan pengklasifikasian Naïve Bayes Classifier untuk menghitung *Accuracy*, *Precision* dan *Recall* menggunakan Data *Training* sebanyak 459 Data dan Data *Testing* sebanyak 141 Data.

$$Accuracy = \frac{TP + TN}{Total} = \frac{121}{141} = 85,8\%$$

$$Precision = \frac{TP}{FP + TP} = \frac{33}{51} = 64,7\%$$

$$Recall = \frac{TP}{FN + TP} = \frac{33}{35} = 94,2\%$$

Dan didapat hasilnya seperti pada hitungan diatas. Hasil yang didapatkan adalah *Accuracy* sebesar 85,8%, *Precision* sebesar 64,7% dan *Recall* sebesar 94,2%.

SIMPULAN

Setelah dilakukan berbagai macam rangkaian pengujian pada sistem, dapat ditarik kesimpulan bahwa :

1. Metode Naïve Bayes Classifier dapat mengklasifikasikan data berupa teks, terutama teks yang berasal dari Twitter berupa *tweet*.
2. Hasil klasifikasi sentimen dari 600 tweet, sentimen negatif lebih dominan dengan jumlah sentimen sebanyak 363, dan sentimen positif sebanyak 237. Sehingga sentimen masyarakat terhadap kelangkaan minyak goreng adalah Negatif
3. Hasil uji akurasi menghasilkan nilai *Accuracy* sebesar 85,8%, *Precision* sebesar 64,7%, dan *Recall* sebesar 94,2%.

DAFTAR PUSTAKA

- [1] F. V. Sari, "ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD . ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI," vol. 10, no. 2, pp. 681–686, 2019.
- [2] T. T. Widowati *et al.*, "ANALISIS SENTIMEN TWITTER TERHADAP TOKOH PUBLIK DENGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR," vol. 11, no. 2, 2020.
- [3] J. Teknologi and I. Jtsi, "Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan," vol. 2, no. 3, pp. 32–37, 2021.
- [4] L. Ardiani, H. Sujaini, and T. Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 2, p. 183, 2020, doi: 10.26418/justin.v8i2.36776.
- [5] N. Haqqizar and T. N. Larasyanti, "Analisis Sentimen Terhadap Layanan Provider Telekomunikasi Telkomsel Di Twitter Dengan Metode Naïve Bayes," *Pros. TAU SNAR-TEK 2019 Semin. Nas. Rekayasa dan Teknol.*, vol. 10, no. 2, pp. 1–15, 2019.
- [6] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, and R. Damaiyanti, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *Sistemasi*, vol. 10, no. 1, p. 222, 2021, doi: 10.32520/stmsi.v10i1.1179.
- [7] I. Fahrur Rozi, A. Taufika Firdausi, and K. Islamiyah, "Analisis Sentimen Pada Twitter Mengenai Pasca Bencana Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram," *J. Inform. Polinema*, vol. 6, no. 2, pp. 33–39,

- 2020, doi: 10.33795/jip.v6i2.316.
- [8] A. F. Cahyanti, "Penentuan Model Terbaik pada Metode Naive Bayes Classifier dalam Menentukan Status Gizi Balita dengan Mempertimbangkan Independensi Parameter," vol. 4, no. 1, pp. 28–35, 2015.
- [9] A. Aziz, F. Fauziah, and I. Fitri, "Analisis Sentimen Terhadap Kebijakan Pemerintah Tentang Larangan Mudik Hari Raya Idulfitri di Indonesia Tahun 2021 Menggunakan Metode Naïve Bayes," *J-SAKTI (Jurnal Sains Komput. dan Inform.,* vol. 5, no. 2, pp. 842–851, 2021, [Online]. Available: <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/381>.
- [10] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.,* vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.