

ANALISIS SENTIMEN PENGGUNA *TWITTER* TERHADAP ROKOK ELEKTRIK (*VAPE*) DI INDONESIA MENGGUNAKAN METODE *NAÏVE BAYES*

Dio Rizki Aditya¹⁾, Endang Supriyati²⁾, Tri Listyorini³⁾

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muria Kudus
Jl. Lkr. Utara, Kayuapu, Gondangmanis, Kec. Bae, Kabupaten Kudus, Jawa Tengah 59327
e-mail: diorizki25@email.com¹⁾, endang.supriyati@umk.ac.id²⁾, trilistyorini@umk.ac.id³⁾

ABSTRAK

Media sosial semacam Twitter ialah media komunikasi yang banyak diminati oleh penduduk Indonesia dari berbagai golongan. Banyak dari masyarakat yang mengutarakan pendapat atau opini sesuatu hal, dari hal baik, maupun hal buruk atau mencela di twitter. Karena hal ini, menjadikan rokok elektrik atau yang biasa disebut vape sebagai objek analisis, yang sudah banyak pengguna rokok elektrik atau vape di masyarakat Indonesia dari golongan remaja sampai dewasa. Menjadikan pendapat atau opini pengguna di twitter berupa tweet-tweet yang kelak diperlukan untuk analisis sentimen, dimana kita dapat menentukan sentimen dari setiap tweet, yang berupa sentimen positif, negatif, dan netral. Pengambilan informasi tweet di twitter kita memerlukan twitter API yang sudah disediakan oleh twitter developer. Menggunakan program yang dibuat memakai bahasa pemrograman python dan memakai model klasifikasi Naïve Bayes. Dari sentimen analisis yang dilakukan memakai model pengklasifikasian Naïve Bayes mendapatkan tingkat akurasi 77.5%, jumlah polarity negatif 11.7%, jumlah polarity netral 77.3%, dan jumlah positif 11%.

Kata Kunci: Analisis Sentimen, Naïve Bayes, Python, Rokok Elektrik (vape), Twitter,

ABSTRACT

Social media such as Twitter is a communication medium that is in great demand by Indonesian people from various groups. Many of the people who express opinions or opinions on something, from good things to bad things or criticize on twitter. Because of this, making e-cigarettes or commonly called vapes as an object of analysis, which has many users of e-cigarettes or vapes in Indonesian society from teenagers to adults. Making user opinions or opinions on Twitter in the form of tweets which will later be used for sentiment analysis, where we can determine the sentiment of each tweet, in the form of positive, negative, and neutral sentiments. Retrieval of tweet information on twitter we use the twitter API that has been provided by the twitter developer. Using a program created using the python programming language and using the Naïve Bayes classification model. From the sentiment analysis carried out using the Naïve Bayes classification model, with an accuracy rate of 77.5%, the number of negative polarity 11.7%, the number of neutral polarity 77.3%, and the number of positive 11%.

Keywords: E-cigarettes (vape), Naïve Bayes, Python, Sentiment Analysis, Twitter

I. PENDAHULUAN

Media sosial ialah media yang dipakai pengguna dalam bentuk mempresentasikan dirinya, berinteraksi, bekerja sama, berbagi informasi, maupun berinteraksi dengan pengguna lainnya. Salah satu platform media sosial yang populer khususnya di Indonesia yakni *Twitter*. *Twitter* ialah salah satu media sosial yang didirikan oleh Jack Dorsey yang umumnya berfungsi untuk mengirimkan pesan yang disebut dengan kicauan atau (*tweet*). Pengguna *twitter* umumnya diakses para pengguna di Indonesia mencapai 59% dan menduduki peringkat ke-5 media sosial yang sering digunakan pada tahun 2020 [1].

Analisis sentimen termasuk ke dalam salah satu bidang *Natural Language Processing* (NLP) dan merupakan suatu proses yang digunakan untuk membantu mengidentifikasi isi dari dataset yang berupa opini atau pandangan (sentimen) berbentuk teks terhadap suatu isu atau kejadian bersifat positif, negatif atau netral. Analisis sentimen merupakan bidang penelitian yang lumayan populer dan dianggap mampu memberikan keuntungan dalam berbagai aspek [2].

Salah satu teknik pembelajaran dari *text mining* untuk analisis sentimen adalah *Naïve Bayes Classifier*. Metode *Naïve Bayes* ini salah satu metode terbaik dalam pengklasifikasian. *Naïve Bayes* menggunakan statistik sederhana berdasarkan *teorema bayes* yang mengasumsikan keberadaan atau ketiadaan dari suatu fitur tertentu, dari suatu kelas yang tidak berhubungan dengan fitur lainnya [3].

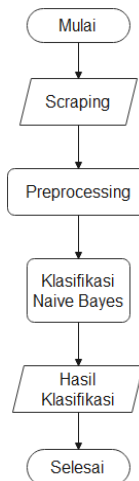
Analisis sentimen dalam pembahasan ini memakai bahasa pemrograman *python*, dan memakai klasifikasi algoritma *Naïve Bayes*. Dengan tahapan-tahapan utama, yakni *scraping*, *pre-processing*, klasifikasi *Naïve Bayes*.

II. METODE PENELITIAN

Sebuah dokumen dapat diklasifikasi ke dalam kategori tertentu berdasarkan kata-kata yang terkandung dalam dokumen tersebut. Kumpulan kata pada dokumen dapat digunakan untuk menentukan kategorinya karena kata-kata tersebut dinilai memiliki makna tertentu. Salah satu algoritma yang digunakan untuk klasifikasi teks adalah *Naive Bayes* [4].

A. Perancangan Alur Secara Umum

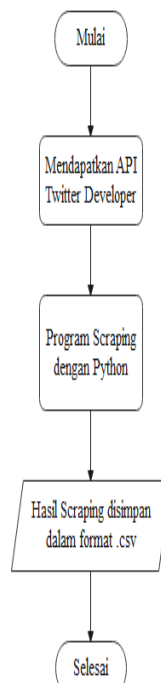
Alur secara umum pada program bisa diperhatikan di Gambar dibawah:



Gambar 1. Alur Secara Umum

Pada Gambar 1 menggambarkan alur secara umum dari program, yang diawali dengan *scraping* pengumpulan data *tweet* dari *twitter*. Selanjutnya data hasil *scraping* diolah ke proses *preprocessing* dimana dari informasi mentah menjadi informasi terstruktur. Setelah data selesai diolah, terakhir proses klasifikasi memakai metode *Naive Bayes*.

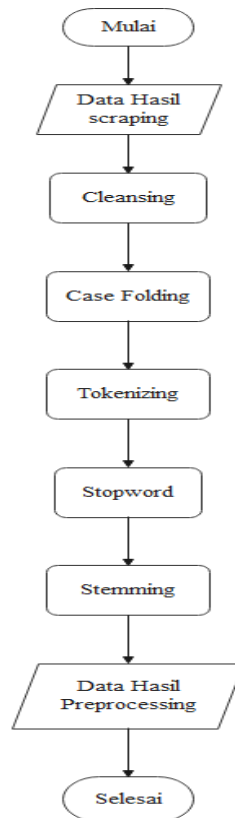
1) Pada proses *scraping*, jalan alur bisa diperhatikan di Gambar bawah:



Gambar 2. Alur Scraping

Pada Gambar 2 menggambarkan proses alur dari *Scraping*. *Web scraping* merupakan sebuah proses untuk mengestrak data dari internet [5]. Langkah pengumpulan data dengan melaksanakan scraping melalui API yang disediakan oleh *Twitter* dengan memanfaatkan *library tweepy* dengan bahasa pemrograman *python* [6]. Hasil dari scraping disimpan pada file format CSV, yang nantinya siap diolah lebih lanjut.

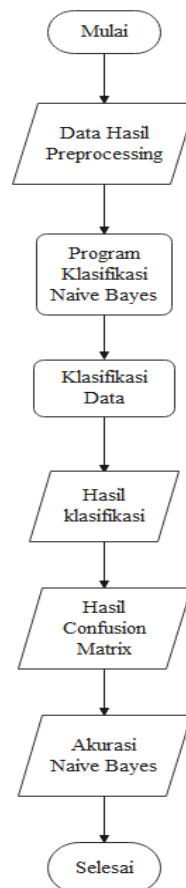
2) Pada proses *preprocessing*, jalan alur bisa diperhatikan di Gambar bawah:



Gambar 3. Alur Preprocessing

Pada Gambar 3 menggambarkan alur proses *preprocessing*. *Preprocessing* adalah mempersiapkan dokumen teks yang tidak terstruktur menjadi data terstruktur yang siap digunakan untuk proses selanjutnya [7]. Tahap *Cleansing*, yaitu proses membersihkan *tweets* dari kata yang tidak diperlukan untuk mengurangi noise [8]. Tahap kedua *case folding*, untuk menyeragamkan bentuk huruf menjadi huruf kecil. Hal ini dilakukan untuk mempermudah pencarian [9]. Tahap ketiga *tokenizing* merupakan proses untuk memisahkan data teks menjadi beberapa *token*. *Tokenizing* secara garis besar memecah sekumpulan karakter dalam suatu teks menjadi suatu kata [10]. Tahap keempat *stopword*, didefinisikan sebagai term yang tidak berhubungan (*irrelevant*) dengan subyek utama dari database meskipun kata tersebut sering kali hadir di dalam dokumen [8]. Tahap terakhir *stemming*, *stemming* adalah proses mencari kata dasar dari sebuah kata imbuhan, dengan menghilangkan imbuhan-imbuhan pada kata dalam dokumen [11].

3) Pada proses klasifikasi *Naïve Bayes*, jalan alur bisa diperhatikan Gambar dibawah:



Gambar 4. Alur Klasifikasi

Pada Gambar 4 menggambarkan alur proses klasifikasi. Algoritma ini menggunakan perhitungan probabilitas dan statistik dalam pengoperasiannya. Metode klasifikasi *Naïve Bayes* digunakan untuk mengambil keputusan dengan melakukan prediksi suatu kasus berdasarkan hasil dari klasifikasi yang telah diperoleh. Pada penelitian ini, metode *Naïve Bayes* digunakan untuk menentukan sentimen dari dokumen tweet [4]. *Naïve Bayes Classifier* dikenal sebagai pengklasifikasi bayesian sederhana dan telah menjadi model probabilistik yang penting dan telah berhasil dalam praktiknya. Walaupun memiliki asumsi independensi yang kuat, *Naïve Bayes Classifier* telah terbukti efektif dalam klasifikasi dalam bentuk teks, diagnosa medis dan manajemen kinerja computer [12]. Pada persamaan (1) ialah rumus *Naïve Bayes*.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

Di mana:

- A : Hipotesis data merupakan suatu clas spesifik.
- B : Data dengan class yang masih belu diketahui.
- $P(A|B)$: Probabilitas hipotesis berdasarkan kondisi.
- $P(A)$: Probabilitas hipotesis.
- $P(B|A)$: Probabilitas berdasarkan kondisi pada hipotesis.
- $P(B)$: Probabilitas B.

Pada persamaan (1) menunjukkan persamaan algoritma *Naïve Bayes*. Merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling tepat, pengklasifikasi bayesian memiliki akurasi dan kecepatan yang lebih tinggi, terutama bila diterapkan pada dataset besar [13].

4) Evaluasi

Confusion matrix yang digunakan untuk melakukan evaluasi klasifikasi metode *naïve bayes*. Dengan menggunakan *confusion matrix* diharapkan dapat mengetahui seberapa akurat metode yang dievaluasi[14]. Langkah yang selalu dipakai ialah menilai akurasi, *precision*, *recall*, *f1-score*. Berikut ialah model tabel *confusion matrix* bisa diperhatikan di Tabel 1.

TABEL 1.
CONFUSION MATRIX

Nilai prediksi	Nilai Nyata	
	True Positif (TP)	False Negatif (FN)
	False Positif (FP)	True Negatif (TN)

Akurasi merupakan presentasi dari sebuah prediksi model yang dibuat. Berikut akurasi dihitung melalui persamaan (2).

$$\text{akurasi} = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (2)$$

Precision ialah tingkat akurasi antara data yang diterima pemakai. Berikut *precision* dihitung melalui persamaan (3)

$$\text{precision} = \frac{TP}{TP + FP} \quad (3)$$

Recall ialah kualitas seberapa lengkap hasil relavan yang ditampilkan oleh sistem prediksi kelas. Berikut *recall* dihitung melalui persamaan (4).

$$\text{recall} = \frac{TP}{TP + FN} \quad (4)$$

F1-score ialah perhitungan salah satu perhitungan evaluasi dalam informasi temu kembali yang menggabungkan *recall* dan *precision*. Berikut *f1-score* dihitung melalui persamaan (5).

$$f1 - \text{score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (5)$$

Dimana TP ialah hasil prediksi positif dan nilai aktual positif, FP ialah hasil prediksi positif dan nilai aktual negatif, FN ialah hasil prediksi negatif dan aktual positif, dan TN ialah prediksi negatif dan nilai aktual negatif [15].

III. HASIL DAN PEMBAHASAN

A. Scraping

Menggunakan data yang sudah didapat dengan *scraping* dari *twitter* dalam analisis sentimen. Data yang diambil dengan kata kunci vape, data yang diambil sebanyak 1000 data *tweet*. Dengan menerapkan bahasa pemrograman python. Setelah itu data disimpan pada dokumen dengan format .csv untuk selanjutnya diolah kembali. Hasil data *scraping* bisa diperhatikan di Tabel 2.

TABEL 2.
CUPLIKAN DATA SCRAPING

Keyword	Date	Username	Tweet
vape	8/23/2021 6:43	dayahdotdot	wujud lagi ke laki yg tak isap benda yg haram mcm rokok , vape dan seangkatan dgnnya
vape	8/23/2021 5:38	esprressoo	njing gw kira apaan ternyata vape
vape	8/23/2021 0:07	adrioonyaa	Alhamdulillah,Puji Tuhan udah bise masok toko vape
vape	8/22/2021 16:15	pelakujanjiphp	@rofantremonti @juventusfcid Mantap om udh ngebul vape hajar surya lagi,,hehe
vape	8/22/2021 15:35	sebvlack	kangen vape aa, gua suka icip walaupun akhirnya gua batuk sampe berair matanya

B. Preprocessing

Dalam tahap *preprocessing*, data hasil *scraping* yang disimpan dalam format CSV selanjutnya diolah kembali. Dimana data yang tidak terstruktur menjadi lebih baik, dengan menghapus atribut yang tidak mempengaruhi dalam proses analisis sentimen. Hasil dari *text preprocessing* bisa diperhatikan di Tabel 3.

TABEL 3.
CUPLIKAN HASIL TEXT PREPROCESSING

Hasil Scraping	Hasil Text Preprocessing
wujud lagi ke laki yg tak isap benda yg haram mcm rokok , vape dan seangkatan dgnnya	wujud laki isap benda haram rokok vape seangkatan
njing gw kira apaan ternyata vape	njing kira apaan ternyata vape
Alhamdulillah,Puji Tuhan udah bise masok toko vape	alhamdulillah puji tuhan masok toko vape
@rofantremonti @juventusfcid Mantap om udh ngebul vape hajar surya lagi,,hehe	mantap om ngebul vape hajar surya
kangen vape aa, gua suka icip walaupun akhirnya gua batuk sampe berair matanya	kangen vape suka icip walaupun akhirnya batuk sampe berair matanya

C. Labelling

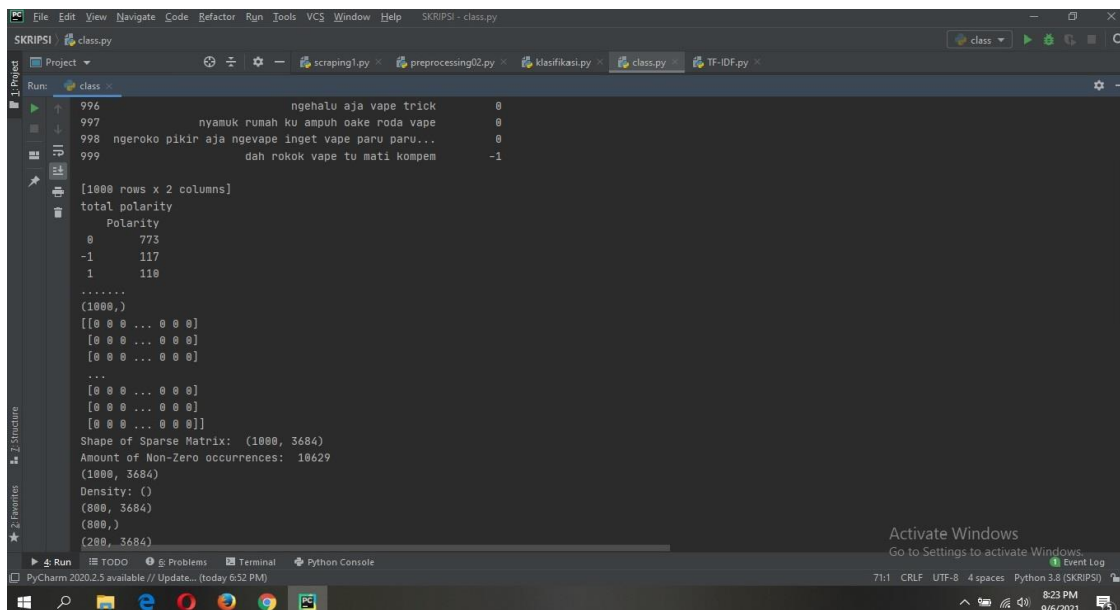
Pada tahap *labelling* data *tweet* yang sudah dilakukan proses *preprocessing* diberikan *labelling* secara manual. *Labelling* ini digunakan untuk klasifikasi *class* pada setiap data *tweet*. *Class* yang digunakan yaitu, *class* positif, *class* negatif, dan *class* netral. *Class* positif diberi tanda “1”, *class* negatif diberi tanda “-1”, dan *class* netral diberi tanda “0”. Berikut ini ialah informasi yang diterima dari hasil *text preprocessing* yang diberikan *class labelling*. Hasil *labelling* bisa diperhatikan di Tabel 4.

TABEL 4.
CUPLIKAN DATA LABELLING

Hasil <i>Text Preprocessing</i>	Label
wujud laki isap benda haram rokok vape seangkatan	-1
njing kira apaan ternyata vape	-1
alhamdulillah puji tuhan masuk toko vape	1
mantap om ngebul vape hajar surya	0
kangen vape suka icip walaupun akhirnya batuk sampe berair matanya	1

D. Klasifikasi Naïve Bayes

Pada tahap klasifikasi yang dilakukan dalam analisis sentimen menggunakan perbandingan 80:20 antara data *training* dan data *testing*. Berikut adalah hasil dilakukannya uji coba dengan menerapkan informasi *tweet* yang ditarik dari *twitter* sekitar 1000 data *tweet* yang sudah melalui tahap *preprocessing* dan *labelling* secara manual. Berikut hasil dari klasifikasi bisa diperhatikan Gambar dibawah.



```

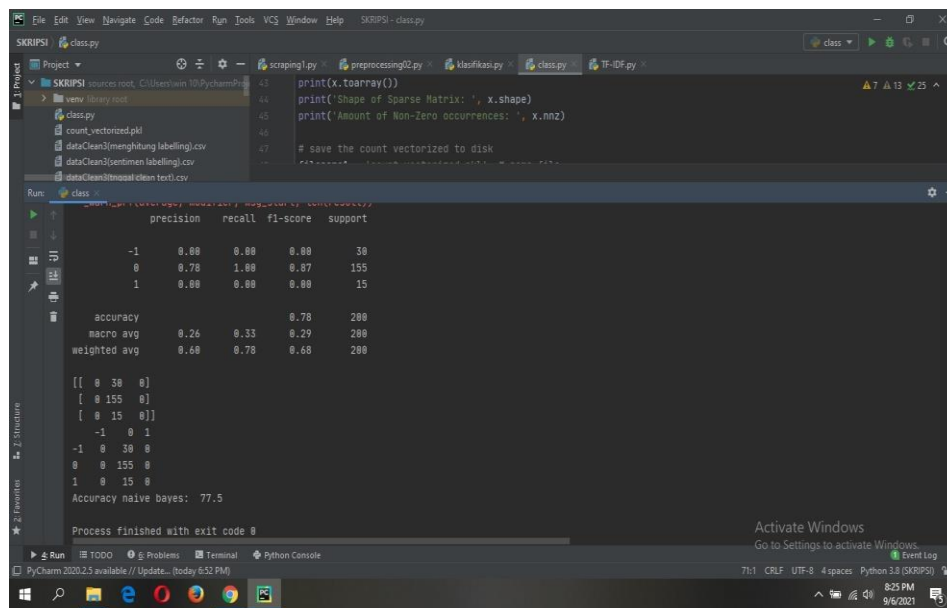
996 ngehalu aja vape trick 0
997 nyamuk rumah ku ampuh oake roda vape 0
998 ngeroko pikir aja ngevape inget vape paru paru... 0
999 dah rokok vape tu mati kompen -1

[1000 rows x 2 columns]
total polarity
Polarity
0 773
-1 117
1 110
.....
(1000,)
[[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]]
Shape of Sparse Matrix: (1000, 3684)
Amount of Non-Zero occurrences: 10629
(1000, 3684)
Density: ()
(800, 3684)
(800,)
(200, 3684)

```

Gambar 5. Hasil Polaritas

Pada Gambar 5 diatas menunjukkan polaritas yang ada dalam data. Merupakan polaritas dari setiap *class*, yakni netral, negatif, positif. Pada polaritas *class* netral memiliki 773 data, *class* negatif 117 data, *class* positif 110 data.



Gambar 6. Hasil Klasifikasi dan Confusion Matrix

Pada Gambar 6 diatas menunjukkan hasil klasifikasi dan *confusion matrix* dari sebuah data. Hasil klasifikasi yang memiliki tingkat akurasi 77.5% dan *confusion matrix* yang menunjukkan data *matrix* dari setiap class.

IV. KESIMPULAN

Berlandaskan hasil dan pembahasan diatas, dapat dirangkumkan bahwa analisis sentimen terhadap rokok elektrik (vape) menggunakan algoritma *Naïve Bayes* dengan informasi yang ditarik dari *tweet-tweet* pengguna *twitter* sebagai target pembahasan. Dari analisis sentimen informasi yang ditarik di *twitter* sekitar 1000 data dengan objek rokok elektrik (vape) memiliki tingkat akurasi 77.5%. Terdapat polaritas sentimen negatif sebanyak 11.7%, sentimen netral sebanyak 77.3%, dan sentimen positif sebanyak 11%.

DAFTAR PUSTAKA

- [1] T. Krisdiyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 7, no. 1, pp. 32–37, 2021, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/coreit/article/view/12945>.
- [2] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [3] T. A. M, Y. Alkhalifi, N. A. Mayangky, and W. Gata, "Analisis Sentimen Opini Publik Mengenai Larangan Mudik pada Twitter Menggunakan Naïve Bayes," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 2, pp. 85–88, 2020.
- [4] N. M. A. J. Astari, Dewa Gede Hendra Divayana, and Gede Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naïve Bayes Classifier," *J. Sist. dan Inform.*, vol. 15, no. 1, pp. 22–29, 2020, doi: 10.30864/jsi.v15i1.332.
- [5] A. Priadana and A. W. Murdiyanto, "Pemantauan Tren Hashtag pada Instagram Menggunakan Web Scraping," *J. Pekommas*, vol. 5, no. 1, pp. 23–30, 2020, doi: 10.30818/jpkm.2020.2050103.
- [6] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [7] J. Ipmawati, Kusriani, and E. Taufiq Luthfi, "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen," *Indones. J. Netw. Secur.*, vol. 6, no. 1, pp. 28–36, 2017.
- [8] F. Syadid, "Analisis Sentimen Komentar Netizen Terhadap Calon Presiden Indonesia 2019 Dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (Tf- Idf) Dan Metode Multi Layer Perceptron (Mlp) Neural Network," *Skrripsi Univ. Islam Negeri Syarif Hidayatullah Jakarta*, pp. 1–89, 2019.
- [9] A. Nugroho, "Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier Dengan Ekstrasi Fitur N-Gram," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 2, no. 2, pp. 200–209, 2018, doi: 10.30645/j-sakti.v2i2.83.
- [10] M. W. A. Putra, Susanti, Erlin, and Herwin, "Analisis Sentimen Dompot Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 72–86, 2020, doi: 10.25299/itjrd.2020.vol5(1).5159.
- [11] E. Supriyati and M. Iqbal, "Pengukuran Similarity Tema Pada Juz 30 Al Qur'an Menggunakan Teks Klasifikasi," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 361–370, 2018, doi: 10.24176/simet.v9i1.1955.
- [12] A. A. Rahman and Y. I. Kurniawan, "Aplikasi Klasifikasi Penerima Kartu Indonesia Sehat Menggunakan Algoritma Naïve Bayes Classifier," *jtm J. Teknol. Dan Manaj. Inform.*, vol. 4, no. 1, 2018.
- [13] S. Juanita, "Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 4, no. 3, pp. 552–558, 2020, doi: 10.30865/mib.v4i3.2140.
- [14] E. Febri Dharmawan, E. D. Wahyuni, and A. A. Arifiyanti, "Klasifikasi Opini : Pengguna Smartphone Pada Twitter Di Indonesia," *J. Inform. dan Sist. Inf.*, vol. 1, no. 1, pp. 121–126, 2020.
- [15] K. Ramadhan and K. M. L., "Analisis Sentimen Terhadap Toko Online Menggunakan Naïve Bayes pada Media Sosial Twitter," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8141–8151, 2018.