

## ANALISIS PENGARUH FILM “ICE COLD” KASUS KOPI SIANIDA TERHADAP SENTIMEN PENGGUNA YOUTUBE DENGAN SVM DAN RANDOM FOREST

Bellatrix Salungweni<sup>1)</sup>, Winsy Weku<sup>2)</sup>, Eliasta Ketaren<sup>3)</sup>

Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Sam Ratulangi

Manado, Indonesia

email: bellatrixsalungweni106@student.unsrat.ac.id<sup>1)</sup>, winsy\_weku@unsrat.ac.id<sup>2)</sup>,

eliastaketaren@unsrat.ac.id<sup>3)</sup>

### Abstrak

Kemajuan teknologi saat ini memungkinkan pertukaran informasi dengan mudah melalui media sosial, yang telah berkembang pesat sejak tahun 1970-an. Di Indonesia, dengan populasi 276,4 juta jiwa, terdapat 167 juta pengguna aktif media sosial, termasuk 139 juta pengguna YouTube, platform berbagi video terpopuler. Fitur YouTube memungkinkan interaksi pengguna melalui komentar dan melihat topik *trending*, seperti kasus kematian Wayan Mirna Salihin, yang kembali menjadi perhatian setelah penayangan film dokumenter “Ice Cold” di Netflix pada 23 September 2023. Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna YouTube dengan jumlah 4000 data komentar terkait kasus Kopi Sianida dengan konten yang berfokus kepada Jessica Wongso sebelum dan sesudah penayangan film menggunakan *lexicon based* dan dilanjutkan dengan klasifikasi menggunakan dua algoritma yaitu *Support Vector Machine* (SVM) dan *Random Forest*. Dengan metode *Lexicon-based* menggunakan kamus InSet, penelitian ini mengelompokkan sentimen menjadi dua kategori: positif dan negatif. Hasil penelitian menunjukkan bahwa penayangan film “Ice Cold” dapat meningkatkan simpati pengguna Youtube terhadap Jessica Wongso, dibuktikan dengan peningkatan sentimen positif dari data sebelum penayangan ke sesudah penayangan film. Selain itu, evaluasi performa model algoritma menunjukkan bahwa SVM memiliki akurasi terbaik dalam klasifikasi sentimen, baik pada data sebelum maupun sesudah penayangan film “Ice Cold”.

**Kata Kunci:** Analisis sentimen, Jessica Wongso, *Lexicon Based*, SVM, *Random Forest*

### 1. Pendahuluan

Kemajuan teknologi pada masa kini sangat pesat, pertukaran informasi dapat dilakukan dengan sangat mudah melalui media sosial. Media sosial diawali pada tahun 1970-an, dan terus berkembang sampai sekarang menjadi salah satu sarana komunikasi utama. Indonesia sendiri dengan total populasi 276,4 Juta jiwa memiliki 167 Juta pengguna aktif media sosial dimana YouTube menjadi salah satu platform media sosial yang banyak digunakan masyarakat Indonesia.

Pengguna juga dapat melihat topik-topik yang sedang banyak dibahas dengan melihat fitur *trending* di YouTube. Salah satu topik yang menjadi tren di YouTube ialah kasus Kopi sianida yaitu kasus kematian Wayan Mirna Salihin yang terjadi pada Januari 2016 dan kuasa hukum memutuskan Jessica Kumala Wongso, sebagai tersangka pembunuhan dan dijatuhi hukuman selama 20 tahun penjara. Kasus ini, khususnya Jessica, kembali menjadi *trending topic* setelah penayangan film dokumenter dari Netflix dengan judul “Ice Cold” pada 23 September 2023, dan menimbulkan berbagai sentimen dari masyarakat.

Analisis sentimen dalam penelitian ini akan berfokus dalam menganalisa sentimen dalam bentuk teks komentar pengguna YouTube yang akan dikelompokkan ke dalam 2 kategori yaitu sentimen positif atau sentimen negatif. Analisis sentimen berguna untuk melihat pandangan, penilaian, dan emosi masyarakat. Penelitian ini akan menggunakan 2 algoritma dalam analisis sentimen yaitu metode *Support Vector Machine* (SVM) dan *Random Forest*, dan penelitian ini akan melihat nilai sentimen (positif dan negatif) pengguna YouTube sebelum dan sesudah penayangan film “Ice Cold”. Penelitian ini menggunakan metode *Lexicon based* dengan kamus InSet (Indonesia Sentimen) untuk pelabelan dataset dan menggunakan bahasa pemrograman python.

### 2. Landasan Teori

#### Kasus Kopi Sianida

Kasus ini bermula saat Jessica, Hani, dan Mirna yang merupakan teman lama bertemu di *caffé Oliver*, Mirna dinyatakan meninggal dunia setelah minum kopi vietnam yang sebelumnya dipesan oleh Jessica. Setelah 10 bulan melewati banyak proses hukum, pada tanggal 27 Oktober 2016 Pengadilan Negeri Jakarta Pusat menyatakan Jessica sebagai tersangka pembunuhan dari kasus kematian Mirna Salihin. Pada 23 September 2023 Netflix

menayangkan film dokumenter dengan judul “Ice Cold”, dalam film dokumenter tersebut mengulas kembali kasus kematian Mirna dan masyarakat pun memberikan banyak sentimen terhadap proses hukum kasus kopi sianida ini

### YouTube

Youtube merupakan platform interaksi dengan video tersebar didunia, menurut riset we are social Youtube merupakan platform sosial dengan pengguna aktif terbanyak kedua di dunia yaitu pengguna Youtube mencapai 2,514 miliar pengguna aktif di seluruh dunia tercatat pada januari 2023. Youtube juga menjadi platform ter-*update* dikarenakan banyak sekali stasiun TV yang mengunggah berita di Youtube, berbagai isu-isu dan permasalahan dibahas dan masyarakat menggunakan fitur komentar Youtube untuk memberikan sentimen secara langsung di video yang di tonton.

### Pengumpulan Data

Pengumpulan data merupakan proses menyerap data dari berbagai sumber seperti gudang data dan basis data, sumber *online* (Komentar, unggahan status, jumlah suka, dan lainnya), dan sistem informasi berupa data penduduk, data pasien dan lainnya. Dalam penelitian ini data diambil dari komentar video Youtube dengan topik Jessica Wongso, masing-masing video pada periode sebelum penayangan dan sesudah penayangan. Berikut merupakan tabel informasi video sumber data komentar pengguna Youtube.

**Tabel 1.** Daftar video komentar sebelum penayangan film “Ice Cold”

No	Judul video	Nama channel	Jumlah komentar (terhitung 02/06/2024)	Jumlah tayangan	Tanggal unggah
1	Episode Akhir Sidang Jessica Wongso	KompasTV	2.6 ribu	2 Juta	27/10/2016
2	Hakim Desak Jessica Jelaskan Siapa yang Taruh Sianida	KompasTV	867	1.3 Juta	19/09/2016
3	Aktivitas Jessica saat Kopi Datang Hingga Dipindahkan	KompasTV	2.9 Ribu	3.6 Juta	5/10/2016
4	Jessica Berdebat dengan Saudara Kembar Mirna	KompasTV	1.1 Ribu	2.1 Juta	13/07/2016

Berdasarkan Tabel 1 data komentar sebelum penayangan diambil dari video YouTube yang diunggah pada tahun 2016, dimana proses hukum atas kasus kopi sianida sedang dijalankan dan video dipilih berdasarkan jumlah penayangan terbanyak dengan fokus pembahasan pada konten adalah Jessica Wongso.

**Tabel 2.** Daftar video komentar sebelum penayangan film “Ice Cold”

No	Judul video	Nama channel	Jumlah komentar (terhitung 02/06/2024)	Jumlah tayangan	Tanggal unggah
1	[FULL] #JusticeForJessica Bergema, Akankah Kasus Kopi Sianida Kembali Dibuka?   Catatan Demokrasi	tvOneNews	11 Ribu	2.4 Juta	10/10/2023
2	[FULL] Kasus Kopi Sianida Jessica, Kenapa Heboh Lagi?   Fakta tvOne	tvOneNews	5.1 Ribu	1.2 Juta	09/10/2023
3	Aktivitas Jessica saat Kopi Datang Hingga Dipindahkan	Metro TV	1.9 Ribu	587 Ribu	11/10/2023

Berdasarkan Tabel 2 data komentar sesudah penayangan diambil dari video YouTube yang diunggah setelah penayangan film “Ice Cold” yang ditayangkan pada tanggal 23 September 2023.

### Analisis Sentimen

Analisis Sentimen adalah merupakan gabungan antara *text mining* dan *natural language processing*, dimana proses ini dilakukan untuk memahami, mengekstrak, dan mengolah data tekstual secara otomatis guna mendapatkan suatu informasi sentimen yang terkandung dalam suatu kalimat [1]. Analisis sentimen banyak digunakan untuk menjadi bahan evaluasi dan ide di berbagai bidang, terutama dalam menyikapi suatu kejadian atau peristiwa yang menimbulkan banyak opini publik [2].

### Text Mining

Analisis sentimen merupakan bagian dari *text mining*, dimana *text mining* merupakan sebuah teknologi untuk memproses dan menganalisa data teks untuk menemukan suatu informasi, *text mining* adalah bagian dari *Artificial Intelligence* (AI) yang menggunakan *Natural Language Processing* (NLP) dalam memproses informasi dari kumpulan data teks [3]. Informasi yang didapatkan melalui *text mining* dapat dimaksimalkan dengan melalui tahap

*pre-processing*, *pre-processing* sendiri bertujuan untuk meningkatkan kualitas dari data teks sebelum dianalisis lebih lanjut [4], berikut merupakan langkah-langkah yang digunakan dalam tahap *pre-processing* data teks [5]:

1) *Cleaning*

Dalam proses *cleaning* akan dilakukan pembersihan pada data, seperti penghapusan URL/Link, Simbol, *Hastag*, *Mention*, dan mengubah semua isi teks kedalam huruf kecil.

2) *Tokenize*

Tahap *tokenize* merupakan tahap memecah data atau kalimat kedalam bentuk potongan kata/token.

3) *Filtering*

Proses penyaringan kata-kata *slang*, singkatan dan membuang *stopword* atau kata penghubung (kata, dan, adalah, untuk, dan lainnya).

4) *Stemming*

Mengubah kata menjadi kata dasar dengan menghilangkan imbuhan pada kata dan menyatukan kembali data dalam bentuk token ke bentuk komentar semula.

### **Lexicon Based**

*Lexicon based* merupakan salah satu metode dalam *Natural Language Processing* (NLP). Penelitian ini menggunakan kamus InSet (*Indonesian Sentiment*) yaitu kamus kata yang disusun oleh Fajri Koto dan Gemal Y [6]. Label sentimen ditentukan dari *polarity compound* masing-masing komentar dengan menggunakan modul *Vader Lexicon*, dimana jika *polarity compound* bernilai  $\geq 0$  maka data komentar akan diberi label “Positif” dan jika *polarity compound* bernilai  $< 0$  maka data komentar akan diberi label “Negatif”

$$\text{Polarity Compound} = \frac{x}{\sqrt{x+a}} \quad (2.1)$$

### **Term Frequency–Inverse Document Frequency (TF-IDF)**

TF-IDF digunakan untuk pembobotan kata yang direpresentasikan sebagai suatu vektor[7]. Kegunaan dari pembobotan TF-IDF untuk menentukan keterkaitan antara kata-kata dengan dokumen atau kalimat, dengan cara memberikan bobot atau nilai pada setiap kata [8].

$$\frac{TP+TN}{TP+TN+FN+FP} \quad (2.2)$$

Selanjutnya dilakukan normalisasi pada nilai bobot mendapatkan hasil pembobotan berskala 0-1 dilakukan normalisasi menggunakan L2 norm [9], dengan rumus berikut.

$$V_{norm} = \frac{v}{\sqrt{v_1^2+v_2^2+v_3^2+\dots+v_n^2}} \quad (2.3)$$

### **Support Vector Machine (SVM)**

SVM merupakan salah satu algoritma dalam *machine learning* yang sering digunakan untuk klasifikasi dan regresi, dan SVM termasuk dalam *supervised learning* yaitu pembelajaran terawasi. SVM bekerja dengan memplot nilai-nilai data pada ruang n-dimensi (n = jumlah fitur), proses klasifikasi dilakukan dengan menemukan *hyperplane* yang paling baik memisahkan dua kelas [10].

### **Random Forest (RF)**

*Random Forest* adalah dalam satu metode klasifikasi dengan basis sekumpulan *decision tree* atau bisa disebut dengan metode *bootstrap* yang kemudian dijadikan data latih [11], hasil akhir dari klasifikasi *random forest* adalah hasil rata-rata dari kumpulan *decision tree*, oleh karena itu algoritma *random forest* masuk ke dalam metode *Ensamble Learning* yaitu memperkuat hasil klasifikasi dengan menggabungkan sejumlah *decision tree* untuk meningkatkan kinerja dan hasil klasifikasi.

### **K-Fold Cross Validation**

*K-Fold Cross Validation* adalah salah satu dari jenis teknik validasi yang berfungsi untuk menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K dari *k-fold* [12].

### **Confusion Matrix**

*Confusion matrix* merupakan metode untuk mengukur ketepatan hasil klasifikasi, berdasarkan kelas yang telah ditentukan dengan algoritma yang di pakai, dalam *confusion matrix* ada 4 nilai utama yaitu dapat dilihat pada Tabel 4.

**Tabel 3.** *Confussion Matrix*

Nilai prediksi	Nilai aktual	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Ada 4 poin utama yang diukur dalam sebuah model klasifikasi yaitu *accuracy*, *recall*, *precision*, dan *F1-Score*. Berikut merupakan rumus perhitungan *Accuracy*, *recall*, *Precision*, dan *F1-Score*:

$$a) \text{ Accuracy: } \frac{TP+TN}{TP+TN+FN+FP} \quad (2.4)$$

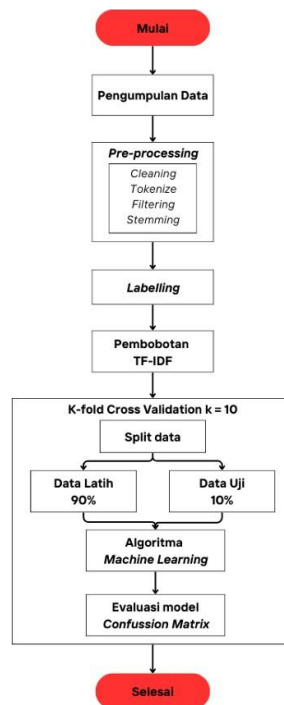
$$b) \text{ Recall: } \text{rec} = \frac{TP}{TP+FN} \quad (2.5)$$

$$c) \text{ Precision: } \text{pre} = \frac{TP}{TP+FP} \quad (2.6)$$

$$d) \text{ F1-Score: } \text{F1-score} = \frac{(2*Recall*Precision)}{Recall+Precision} \quad (2.7)$$

### 3. Metode Penelitian

Penelitian ini menggunakan bahasa pemrograman python menggunakan aplikasi Google colab dengan alur penelitian pada gambar berikut:



**Gambar 1.** Diagram alir penelitian

#### Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini menggunakan metode *crawling*, yaitu dengan menggunakan aplikasi ‘Octoperse’ untuk *crawling* data komentar YouTube. Diawali dengan memasukkan tautan video pada Tabel 1 dan Tabel 2 ke dalam aplikasi Octoperse, selanjutnya akan dilakukan *crawling* data, setelah data terkumpul data duplikat akan dihapus, dan disatukan ke dalam file CSV masing-masing kategori, yaitu file komentar sebelum penyayangan dan sesudah penyayangan.

#### Pre-processing

Melakukan 4 tahap dalam *pre-processing* yaitu *cleaning*, *tokenize*, *filtering*, dan *stemming*. Untuk membersihkan data dari karakter selain huruf dan spasi antar kata, dan mengubah semua huruf ke dalam huruf kecil.

### Labelling

Setelah tahap *pre-processing*, data yang telah dibersihkan akan masuk ke dalam tahap *labelling*. Pada tahap ini akan menggunakan *lexicon based* dengan menghitung bobot setiap komentar sesuai dengan kamus *lexicon*.

### Pembobotan TF-IDF

Proses pembobotan dilakukan setelah proses labelling selesai, selanjutnya akan diterapkan pembobotan pada data menggunakan metode TF-IDF untuk mengoptimalkan proses klasifikasi.

### Klasifikasi

Proses selanjutnya dilanjutkan dengan pembagian data latih dan data uji menggunakan *K-Fold Cross Validation* dengan nilai  $k=10$  dan proses klasifikasi menggunakan 2 algoritma yaitu algoritma *Support Vector Machine* dan *Random Forest*. Pada proses ini akan dilakukan proses pelatihan model algoritma menggunakan data latih yang ada, setelah proses training selesai model akan diuji menggunakan data uji untuk melihat ketepatan algoritma dalam klasifikasi data menggunakan *confussion matrix* untuk evaluasi nilai *accuracy*, *precision*, *recall* dan *f1-score*.

## 4. Hasil Penelitian

### Pre-Processing

Data berjumlah 2000 komentar pada masing-masing kategori yaitu data sebelum penayangan dan data sesudah penayangan, berikut merupakan hasil penerapan tahap pre-processing pada data sampel:

Tabel 4. Hasil pre-processing

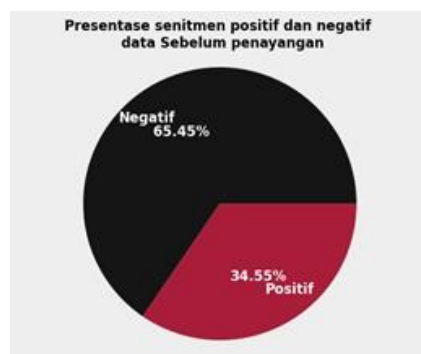
Sebelum <i>pre-processing</i>	Sesudah <i>pre-processing</i>
Kasian banget yh udah 7 tahun di penjara tapi kasusnya masih ragu GX bisa bayangin deh pasti GX enak banget	kasihan banget ya penjara kasus ragu bayang deh enak banget

### Labelling

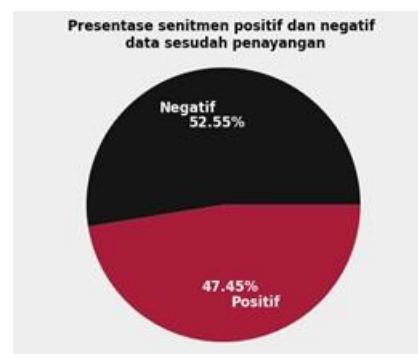
Berikut merupakan hasil pelabelan menggunakan *lexicon based*:

Tabel 5. Hasil labelling

Data Komentar	Label "Positif"	Label "Negatif"	Total
Sebelum	691	1309	2000
Sesudah	949	1051	2000
Selisih	+ 258	- 258	



Gambar 2. Persentase sentimen data sebelum penayangan



Gambar 3. Persentase sentimen data sebelum penayangan

Dari hasil *labelling* kedua data komentar tersebut, diketahui sentimen "Positif" sebelum ke sesudah penayangan memperoleh kenaikan sebesar 12.9%.

### TF-IDF

Sebelum melakukan proses klasifikasi, data komentar yang telah melewati tahap *pre-processing* dan *labelling* akan dilakukan pembobotan menggunakan fitur TF-IDF. Berikut contoh pembobotan menggunakan fitur TF-IDF menggunakan Persamaan (2.3) pada data sampel:

Sample data yang sudah melewati tahap pre-processing:

“tenang cuek sesal asli psikopat”

**Tabel 6.** Hasil TF-IDF

Kata / Term	TF	df	D/df	IDF	TF-IDF (TF * log(D/df))	v-norm
tenang	0.2	44	2000/44	4.795	0.959	0.36213
cuek	0.2	2	2000/2	6.503	1.501	0.56666
sesal	0.2	10	2000/10	5.204	1.241	0.46853
asli	0.2	13	2000/13	4.962	1.192	0.45032
psikopat	0.2	49	2000/49	3.689	0.938	0.35418
$\sqrt{V1^2 + V2^2 + V3^2 + \dots Vn^2}$					<b>2.648</b>	

### Klasifikasi

Setelah dilakukan pembobotan pada data, selanjutnya dilakukan proses klasifikasi menggunakan dua algoritma yaitu SVM dan *Random Forest* dengan teknik validasi menggunakan *k-fold cross validation* dengan nilai  $k=10$ , berikut hasil evaluasi dari masing masing model algoritma:

**Tabel 7.** Nilai akurasi masing-masing model algoritma *machine learning*

Data	Algoritma	Akurasi terbaik (%)
Sebelum	SVM	<b>87%</b>
	RF	<b>86%</b>
Sesudah	SVM	<b>85%</b>
	RF	<b>83%</b>

### 5. Kesimpulan

Penayangan film “*Ice Cold*” yang berisi dokumentasi kasus Kopi Sianida meningkatkan simpatik pengguna Youtube khususnya kepada Jessica Wongso, dengan peningkatan sentimen positif dari data sebelum penayangan ke sesudah penayangan film “*Ice Cold*” dengan teknik *labelling* menggunakan *lexicon based*. Klasifikasi sentimen pengguna Youtube tersebut menggunakan algoritma SVM dan Random Forest, ditemukan bahwa SVM memiliki akurasi terbaik berdasarkan teknik validasi *K-fold cross validation* dengan nilai  $k=10$ .

### 6. Daftar Pustaka

- [1] E. G. Ruus, L. A. Latumakulita, and J. D. Prang, “Analisis Sentimen di Media Online menggunakan metode Naive Bayes,” *d’CARTESIAN J. Mat. dan Apl.*, no. September, pp. 18–22, 2022, [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/decartesian/article/view/36624%0Ahttps://ejournal.unsrat.ac.id/index.php/decartesian/article/viewFile/36624/39057>
- [2] A. P. Natasuwarna, “Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring,” *Techno.Com*, vol. 19, no. 4, pp. 437–448, 2020, doi: 10.33633/tc.v19i4.4044.
- [3] H. B. Tambunan and T. W. D. Hapsari, “Analisis Opini Pengguna Aplikasi New PLN Mobile Menggunakan Text Mining,” *Petir*, vol. 15, no. 1, pp. 121–134, 2021, doi: 10.33322/petir.v15i1.1352.
- [4] M. J. Aufa and A. Qoiriah, “Analisis Sentimen Pengguna Platform Belajar Online Coursera menggunakan Random Forest dengan Metode Ekstraksi Fitur Word2vec,” *J. Informatics Comput. Sci.*, vol. 04, pp. 244–255, 2023, doi: 10.26740/jinacs.v4n02.p244-255.
- [5] S. I. Nurhafida and F. Sembiring, “Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM),” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 317–327, 2022.
- [6] F. Koto and G. Y. Rahmaningtyas, “Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs,” *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, no. December, pp. 391–394, 2017, doi: 10.1109/IALP.2017.8300625.
- [7] S. R. Yustihan, P. P. Adikara, and Indriati, “Analisis Sentimen berbasis Aspek terhadap Data Ulasan Rumah Makan menggunakan Metode Support Vector Machine (SVM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, pp. 1017–1023, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] E. Hokijulandy, H. Napitupulu, and Firdaniza, “Application of SVM and Chi-Square Feature Selection for Sentiment Analysis of Indonesia’s National Health Insurance Mobile Application,” *Mathematics*, vol. 11, no. 17, 2023, doi: 10.3390/math11173765.
- [9] M. Fikri and R. Sarno, “A comparative study of sentiment analysis using SVM and Senti Word Net,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 3, pp. 902–909, 2019, doi: 10.11591/ijeecs.v13.i3.pp902-909.

- [10] S. S. Aljameel *et al.*, “A sentiment analysis approach to predict an individual’s awareness of the precautionary procedures to prevent covid-19 outbreaks in Saudi Arabia,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, pp. 1–12, 2021, doi: 10.3390/ijerph18010218.
- [11] A. ELHAN, M. K. D. HARDHIENATA, H. YENI, S. WIJAYA HARTONO, and J. ADISANTOSO, “Analisis Sentimen Pengguna Twitter terhadap Vaksinasi COVID-19 di Indonesia menggunakan Algoritme Random Forest dan BERT Sentiment Analysis of Twitter Users on COVID-19 Vaccines in Indonesia using Random Forest and BERT Algorithms,” *J. Ilmu Komput. Agri-informatika*, vol. 9, no. 2, pp. 199–211, 2022, [Online]. Available: <https://jurnal.ipb.ac.id/index.php/jika/article/view/44459>
- [12] D. Cahyanti, A. Rahmayani, and S. A. Husniar, “Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.