

ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP TAYANGAN #MATANAJWAMENANTITERAWAN DENGAN METODE NAÏVE BAYES CLASSIFIER

Mathoril Hudha¹, Endang Supriyati², Tri Listyorini³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Muria Kudus
Email: mathoril.hudha@gmail.com, endang.supriyati@umk.ac.id, trilistyorini@umk.ac.id

(Naskah masuk: 03 Agustus 2021, diterima untuk diterbitkan: 27 Januari 2022)

Abstrak

Tayangan #MataNajwaMenantiTerawan menuai pro dan kontra setelah Najwa Shihab dalam acara Mata Najwa mewawancarai kursi kosong yang direpresentasikan sebagai Menteri Kesehatan Terawan Agus Putranto. Wawancara dengan kursi kosong biasa dilakukan di luar negeri yang memiliki histori pers yang lama. Seperti yang dilakukan Piers Morgan di CNN dan Kay Burley di acara Kay Burley Show di Sky News. Media sosial mengalami peningkatan penggunaan sehingga analisis sentimen perlu dilakukan untuk mengetahui opini publik terhadap isu yang sedang terjadi. Analisis sentimen dalam penelitian ini dilakukan untuk mengetahui bagaimana penilaian publik terhadap tayangan #MataNajwaMenantiTerawan di *YouTube*. Data dalam penelitian ini diambil dengan cara *scrapping* dari komentar video *YouTube* #MataNajwaMenantiTerawan. Dataset mentah hasil *scrapping* dilakukan *preprocessing* untuk memudahkan proses klasifikasi. Algoritma *Naive Bayes* digunakan untuk mengklasifikasikan komentar ke kategori sentimen positif, sentimen negatif dan netral. Algoritma *Naive Bayes* menghasilkan akurasi sebesar 90.36%. Sentimen netral lebih mendominasi dengan jumlah 1232 data netral, 90 data negatif dan 78 data positif.

Kata kunci: Analisis Sentimen, Naive Bayes, Text Preprocessing, Scrapping, #MataNajwaMenantiTerawan

SENTIMENT ANALYSIS OF USER COMMENTS ON YOUTUBE VIDEO #MATANAJWAMENANTITERAWAN WITH NAIVE BAYES CLASSIFIER METHOD

Abstract

The show #MataNajwaMenantiTerawan reaped pros and cons after Najwa Shihab in the Mata Najwa program interviewed an empty chair represented as Minister of Health Terawan Agus Putranto. Interviews with empty chair are usually conducted abroad, which have a long press history. Like Piers Morgan on CNN and Kay Burley on the Kay Burley Show on Sky News. Social media has increased its use so that sentiment analysis needs to be done to find out public opinion on the issues that are happening. Sentiment analysis in this study was conducted to find out how the public evaluates the #MataNajwaMenantiTerawan show on *YouTube*. The data in this study were taken by *scrapping* from comments on the *YouTube* video #MataNajwaMenantiTerawan. The raw dataset from *scrapping* is *preprocessed* to facilitate the classification process. *Naive Bayes* algorithm is used to classify comments into categories of positive sentiment, negative sentiment and neutral. The *Naive Bayes* algorithm produces an accuracy of 90.36%. Neutral sentiment dominates with 1232 neutral data, 90 negative data and 78 positive data.

Keywords: Analisis Sentimen, Naive Bayes, Text Preprocessing, Scrapping, #MataNajwaMenantiTerawan

1. PENDAHULUAN

Pers merupakan subsistem dari sistem sosial. Bagi bangsa Barat perkembangan pers dan teknologinya mengikuti perkembangan masyarakatnya yang berjalan secara linear. Ketika mesin cetak ditemukan pada 1450 sampai ditemukan lagi kamera/film 400 tahun kemudian, tumbuhlah budaya membaca dan Pendidikan yang meluas dan mapan. 50 tahun setelah teknologi visual, pers

masuk ke media audio dengan ditemukannya pemancar radio. Masa-masa berikutnya teknologi pers dan penyiaran terus diperpendek dengan ditemukannya media berbasis audio-visual yang dikenal dengan tabung gambar atau televisi 30 tahun berikutnya. Disusul 10 tahun berikutnya masyarakat barat sampai kepada teknologi satelit komunikasi [1].

Era *Internet of Things* (IoT) menjadikan perkembangan teknologi tersebut dapat dinikmati melalui gadget. Apple, Windows dan Android menyediakan berbagai *platform* untuk menikmati konten-konten digital yang dapat diakses dengan aplikasi dan kuota data internet. Konten visual seperti berita teks atau info grafis dapat dinikmati melalui *web browser*, konten audio dapat didengarkan melalui *Spotify* atau konten audio-visual yang dapat ditonton melalui *YouTube*.

Nielsen Cross-Platform pada tahun 2017 melakukan survei terhadap pengguna internet. Survei tersebut menunjukkan terjadinya peningkatan akses internet oleh netizen di banyak tempat, Di antaranya adalah (51%) di kafe atau restoran dan (53%) kendaraan umum. Acara konser pun mengalami peningkatan (24%) dalam jumlah akses media digital dibanding tahun 2015. Rumah dan tempat kerja juga mengalami peningkatan akses internet. Berbagai portal berita seperti *jawapos.com*, *detik.com*, *kompas.com*, *tribun.com* dan *tempo.co* merupakan contoh media yang menyesuaikan dengan berkembangnya teknologi. Semula media tersebut hanya menyajikan produk cetak berupa koran, majalah atau buku. Tayangan televisi pun menyesuaikan perkembangan teknologi seiring masuknya internet. Salah satu layanan penyiaran populer berbasis audio-visual adalah *YouTube*. Pengguna *YouTube* mencapai 95% dari pengguna internet. Konten yang tersedia di *YouTube* memiliki 80 bahasa yang berbeda. Lebih dari 2 miliar pengguna baru mengunjungi *YouTube* dalam setiap bulan dan videonya ditonton lebih dari satu miliar jam setiap hari [2]. Selain sebagai media penyiaran, *YouTube* juga media sosial yang memungkinkan komunikasi antar penggunanya. *YouTube* adalah salah satu platform yang banyak dimanfaatkan TV untuk memublish siarannya. Seperti Kick Andy di *Metrotv*, Indonesia *Lawyers club* di *TvOne*, dan *Aiman* di *Kompas TV* yang juga mempublish tayangannya di *YouTube*.

Masyarakat membutuhkan informasi dari sumber terpercaya untuk menghadapi pandemi. Najwa Shihab melalui program Mata Najwa di *Trans TV* berusaha mencari kejelasan informasi dari Menteri Kesehatan Terawan Agus Putranto. Wawancara tersebut juga di *publish* melalui kanal *YouTube* Najwa Shihab setelah acara tersebut ditayangkan di TV secara live. Najwa Shihab dalam acara tersebut mewawancarai kursi kosong yang direpresentasikan sebagai Menteri Kesehatan saat Pandemi COVID-19 mewabah. Wawancara tersebut dilakukan Najwa Shihab karena Menteri Kesehatan sangat jarang memberikan keterangan mengenai pandemi COVID-19.

Wawancara dengan kursi kosong biasa dilakukan di luar negeri yang memiliki histori pers yang lama. Seperti yang dilakukan Piers Morgan di *CNN* dan Kay Burley di acara *Kay Burley Show* di *Sky News*. Namun Hal tersebut belum pernah

dilakukan di Indonesia sehingga menuai pro dan kontra yang memicu penonton tayangan #MataNajwaMenantiTerawan mengungkapkan emosinya.

Emosi yang diluapkan di kolom komentar tayangan #MataNajwaMenantiTerawan inilah yang menjadi objek penelitian untuk melakukan analisis sentimen dengan metode *Naïve Bayes Classifier*.

Tayangan #MataNajwaMenantiTerawan perlu dilakukan analisis sentimen karena wawancara dengan kursi kosong adalah produk baru dalam jurnalistik Indonesia, sehingga mengundang public untuk beropini. Tayangan tersebut memperoleh 58.000 komentar hingga bulan Agustus 2021. Banyaknya komentar dalam tayangan tersebut mengandung opini atau sentiment yang bermacam-macam. Opini public terhadap tayangan tersebut dapat diketahui dengan analisis sentiment [3].

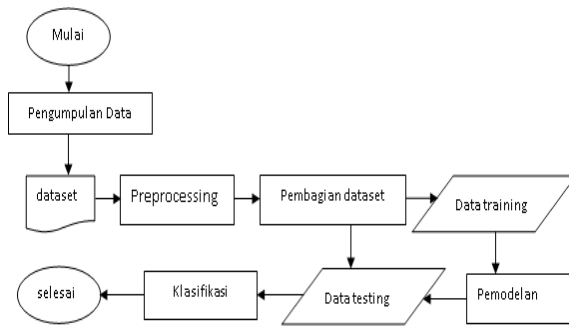
Opini atau sentiment dalam penelitian ini dikelompokkan dengan algoritma *Naïve Bayes Classifier*. *Naïve Bayes Classifier* adalah model sederhana untuk klasifikasi [4]. Model ini merupakan bentuk sederhana dari *Bayesian Network*, dimana semua atribut independen diberi nilai kelas variabel [5]. Waktu pemrosesan *Naïve Bayes Classifier* yang cepat, mudah diimplementasikan dengan strukturnya yang cukup sederhana dan tingkat efektifitas yang tinggi [6].

Penelitian terkait analisis sentimen menggunakan naive bayes pernah dilakukan oleh [7-8] namun pada kasus objek wisata. Untuk itu pada penelitian akan dilakukan analisis sentimen menggunakan naive bayes pada tayangan #matanajwamenantiterawan. Hasil dari penelitian ini adalah klasifikasi komentar kedalam sentiment positif, sentiment negative dan sentiment netral. Hasil pengujian algoritma yang digunakan diperoleh nilai akurasi yang menunjukkan seberapa persen ketepatan algoritma *Naïve Bayes Classifier* untuk melakukan klasifikasi sentiment.

2. METODE PENELITIAN

2.1 Diagram Alur Proses

Penelitian dilakukan dengan mengumpulkan komentar *YouTube* tayangan #MataNajwaMenantiTerawan. Komentar tersebut merupakan data yang akan dianalisis dalam penelitian ini. Berikut ini adalah jalannya penelitian yang digambarkan di Gambar 1.



Gambar 1. Diagram Alur Proses

Gambar 1 menjelaskan tentang alur secara umum yang dilakukan dalam penelitian ini. Penelitian ini dimulai dengan pengumpulan data yang berupa teks komentar. Data tersebut dilakukan preprocessing, vektorisasi, pembagian data, pemodelan dan klasifikasi.

2.2 Pengumpulan Data

Penelitian ini mengambil data dengan cara *scrapping* komentar di *YouTube* tayangan #MataNajwaMenantiTerawan. *Scrapping* dilakukan dengan memanfaatkan *library Selenium* dengan bahasa pemrograman Python. *Chrome driver* digunakan untuk membuka google chrome dan mengendalikannya secara otomatis. Data yang telah diambil dari proses *scrapping* disimpan dalam bentuk CSV yang nantinya akan diproses lebih lanjut.

2.3 Preprocessing

Data yang tidak terstruktur dan tidak konsisten perlu dibersihkan melalui proses text preprocessing. Setelah data bersih maka akan dilakukan proses pelabelan secara manual [9]. Pada proses ini digunakan googlesheet dengan add-ons Logic sheet untuk melakukan *Cleansing*, *Case Folding*, *Tokenize*, *Stopword Removal* dan *Stemming*. Berikut ini adalah penjelasan dari tahap *preprocessing*:

Cleansing: Dataset mentah dibersihkan dari *symbol*, emoji, tanda baca, link url, tanda baca dan angka.

Case Folding: Dataset yang memuat huruf besar diubah menjadi huruf kecil.

Tokenize: Kalimat komentar dipecah menjadi kata per kata seperti “dia sedang mewawancarai kursi” menjadi “dia”, “sedang”, “mewawancarai”, “kursi”.

Stopword Removal: Kata hubung atau kata-kata yang tidak penting dalam kalimat, misalnya “di”, “ke”, “dari” atau “yang”.

Stemming: Pada proses ini dilakukan perubahan kata-kata yang memiliki imbuhan seperti “mainan” menjadi “main”.

2.4 Naïve Bayes Classifier

Metode yang simpel dengan hasil yang baik untuk mengklasifikasi teks adalah keunggulan *Naïve Bayes Classifier*. Keunggulan lain melakukan klasifikasi dengan *Naïve Bayes* adalah cepat dan akurasi yang tinggi [10-11]. Berikut ini adalah rumus *Naïve Bayes* dalam persamaan 1 [12].

$$P(A|S) = \frac{P(S|A)*P(A)}{P(S)} \quad (1)$$

Keterangan:

$P(A|S)$: Probabilitas A terhadap kondisi S

$P(S|A)$: Probabilitas S terhadap kondisi A

$P(A)$: Probabilitas A

$P(S)$: Probabilitas S

3. HASIL DAN PEMBAHASAN

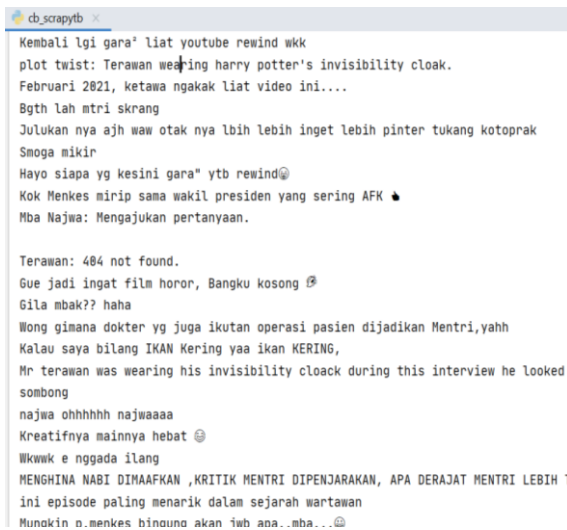
3.1 Scrapping

Data dalam penelitian ini diperoleh dari komentar pengguna *YouTube* dalam tayangan #MataNajwaMenantiTerawan. Data tersebut diambil dengan cara *scrapping* menggunakan Bahasa pemrograman python dengan memanfaatkan *library selenium*, *chrome driver* dan *xpath*. Dalam penelitian ini diambil data sebanyak 1400 komentar dalam waktu 1 menit. Komentar diambil pada tanggal 8 Juli 2021. Hasil *scrapping* dipaparkan di Tabel 1.

Tabel 1. Hasil *Scrapping*

Komentar
krispi banget, ngomong sama mahluk ghoib
Yang kesini gara gara YouTube rewind like 🤔🤔
Menteri taunya cuma duit doang n korupsi
Prestasi nihil
BPJS doesn't cover that burn, bro.
Semoga Indonesia lebih baik dengan adanya orang orang seperti mba nana untuk kritik yang membangun 😊
pertanyaannya bagus loh, jdi penasaran bgt jawabannya, sayang gabisa hadir
Hahahaha... 😂😂😂
Gile yang komen 58 ribu.. luar biasa mba nana
Plot twist : Terawan is just the chair's name
pak?? kemana pak meninggoy

Hasil *Scrapping* juga dapat dilihat dari jalannya program di Gambar 2 sebagai berikut:



Gambar 2. Hasil Scapping

3.2 Text Preprocessing

Sebelum data diproses dengan Algoritma *Naive Bayes*, data harus melewati tahap preprocessing. Data yang tidak konsisten perlu dilakukan *preprocessing* agar menjadikan data siap untuk diproses [13]. *Preprocessing* terdiri dari langkah *Tokenizing*, *Normalisasi*, *Stemming* dan Menghapus *Stopword*. Pada proses ini digunakan *GoogleSheet* dan *ad-ons* untuk membersihkan dan merapikan data. Hasil *text preprocessing* dipaparkan di Tabel 2.

Tabel 2. Hasil Text Preprocessing

Komentar	Hasil Preprocessing
krispi banget, ngomong sama mahluk ghoib	krispi banget ngomong sama mahluk ghoib
Yang kesini gara gara YouTube rewind like 😊😂	yang kesini gara gara YouTube rewind like
Menteri taunya cuma duit doang n korupsi	menteri taunya cuma duit doang n korupsi
Prestasi nihil	prestasi nihil
BPJS doesn't cover that burn, bro.	bpjs doesnt cover that burn bro
Semoga Indonesia lebih baik dengan adanya orang orang seperti mba nana untuk kritik yang membangun 😊	semoga indonesia lebih baik dengan adanya orang orang seperti mba nana untuk kritik yang membangun
pertanyaannya bagus loh, jdi penasaran bgt jawabannya, sayang gabisa hadir	pertanyaannya bagus loh jadi penasaran banget jawabannya sayang gabisa hadir
Hahahaha... 😂😂😂	hahahaha
Gile yang komen 58 ribu.. luar biasa mba nana	gile yang komen ribu luar biasa mba nana
Plot twist : Terawan is just the chair's name	plot twist terawan is just the chairs name
pak?? kemana pak meninggoy	pak kemana pak meninggoy

3.3 Klasifikasi Naive Bayes

Proses *training* dan proses *testing* adalah tahapan dari klasifikasi *Naive Bayes*. Proses *training* merupakan proses yang digunakan sebagai model analisis sentimen terhadap data *testing* atau proses *testing*.

Penelitian ini dilakukan dengan 1400 data komentar yang telah melalui tahap preprocessing dan labelling. Data tersebut dibagi menjadi data training sebanyak 1120 data dan data testing sebanyak 280 data. Proses klasifikasi *Naive Bayes* digunakan data training yang digunakan untuk pemodelan, sedangkan data testing akan digunakan untuk menguji akurasi dari model yang dibentuk [14]. Proses klasifikasi *Naive Bayes* dilakukan dengan Bahasa pemrograman python dengan *library scikit learn*.

Sebelum penentuan class sentiment positif, negative atau netral perlu dilakukan proses labelling. Proses labelling adalah proses memberikan label pada setiap kalimat. Tahap labelling dilakukan dengan melakukan perhitungan frekuensi kata yang muncul untuk diberi label positif atau negatif.

Kalimat yang telah diberi label akan dibagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk melatih data *testing*.

Tahap terakhir adalah tahap evaluasi hasil, dalam tahap ini algoritma *Naive Bayes* di ukur kinerjanya. Tahap ini dilakukan pengukuran *Precision*, *F1-Score*, *Recall* dan akurasi. Berikut ini adalah contoh tabel *confusion matrix* [15].

Tabel 3. Confusion Matrix

Kondisi		Nilai Sebenarnya	
		Relevan	Tidak Relevan
Nilai Prediksi	Tampil	True Positif (TP)	False Negatif (FN)
	Tidak Tampil	False Positif (FP)	True Negatif

Keempat parameter di atas mempunyai rumus sebagai berikut:

1. Recall

Rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Berikut ini di persamaan (2) adalah rumus *recall*.

$$Recall = \frac{TP}{TP+FP} \quad (2)$$

2. Precision

Rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi negative. Berikut ini di persamaan (3) adalah rumus *precision*.

$$Precision = \frac{TP}{TP+FN} \quad (3)$$

3. F1-Score

Perbandingan rata-rata presisi dan recall yang dibobotkan. Berikut ini di persamaan (4) adalah rumus *F1-Score*.

$$F1 - Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (4)$$

4. Akurasi

Perbandingan kelas ya prediksi ya ditambahkan kelas tidak prediksi tidak dibagi kelas positif dan kelas negative. Berikut ini di persamaan (5) adalah rumus akurasi.

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (5)$$

Algoritma *Naive Bayes* dalam penelitian ini menghasilkan akurasi sebesar 90.36%.

Berikut di gambar ini adalah hasil evaluasi algoritma *Naive Bayes* yang dilakukan secara otomatis dengan Bahasa pemrograman python.

	precision	recall	f1-score	support
-1	0.00	0.00	0.00	14
0	0.90	1.00	0.95	253
1	0.00	0.00	0.00	13

Gambar 3. Hasil Evaluasi Program

Berikut ini di Tabel 4 dipaparkan hasil klasifikasi. Hasil klasifikasi menunjukkan data negative data positif dan netral.

Sentimen	Polaritas
Positif	78
Negatif	90
Netral	1232

Hasil klasifikasi juga dapat dilihat dari jalannya program di Gambar 3 sebagai berikut:

```
[1400 rows x 2 columns]
total polarity
      Polarity
0          1232
-1           90
1            78
```

Gambar 4. Hasil Klasifikasi

4. KESIMPULAN

Hasil dan pembahasan mengenai analisis sentimen pengguna YouTube terhadap tayangan #MataNajwaMenantiTerawan dengan metode *Naive Bayes Classifier* diperoleh kesimpulan yaitu metode *Naive Bayes* Mampu Melakukan Klasifikasi komentar secara otomatis dengan bahasa pemrograman python. Selain itu, Analisis sentimen terhadap data komentar YouTube terhadap tayangan #MataNajwaMenantiTerawan dapat dilakukan dengan metode *Naive Bayes* dengan nilai akurasi 90.36%.

5. DAFTAR PUSTAKA

- [1] Z. A. Achmad. 2014. "Perbandingan Sistem Pers," *Lutfansah Mediat.*, [Online]. Available: https://www.researchgate.net/profile/Zainal_Achmad/publication/326436775_Perbandingan_Sistem_Pers_dan_Sistem_Pers_di_Indonesia/link/s5b4d5a86a6fdcc8dae2468ea/Perbandingan-Sistem-Pers-dan-Sistem-Pers-di-Indonesia.pdf.

- [2] D. Abul-Fottouh, M. Y. Song, and A. Gruzd. 2020. "Examining algorithmic biases in YouTube's recommendations of vaccine videos," *Int. J. Med. Inform.*, vol. 140, no. April, p. 104175, doi: 10.1016/j.ijmedinf.2020.104175.
- [3] G. A. Buntoro. 2017. "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," vol. 2, no. 1, pp. 32–41.
- [4] L. L. Dhande and G. K. Patnaik. 2014. "Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier," *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 3, no. 4, pp. 313–320.
- [5] E. Indrayuni. 2019. "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 29–36, doi: 10.31294/jki.v7i1.1.
- [6] S. Taheri and M. Mammadov. 2013. "Learning the naive bayes classifier with optimization models," *Int. J. Appl. Math. Comput. Sci.*, vol. 23, no. 4, pp. 787–795, doi: 10.2478/amcs-2013-0059.
- [7] L. Wilianto., P. T. Hendro., dan F. R. Umbara. (2017). Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Naive Bayes Classifier Studi Kasus Jawa Barat. *Prosiding SNATIF*, pp.439-448
- [8] R. Sari., dan H. Y. Ratih. (2019). Penerapan Algoritma Naive Bayes Untuk Analisis Sentiment Pada Wisata TMII Berbasis Website. *IJSE- Indonesian Journal on Software Engineering*, Vol 5(2), pp. 51-60
- [9] A. N. Yusuf, E. Supriyati, and T. Listyorini. 2020. "Analisis Sentimen Mengenai Layanan Provider Indihome Berdasarkan Pendapat Pelanggan Melalui Media Sosial Twitter dengan Metode Naive Bayes Classifier," vol. 04, pp. 75–78.
- [10] M. Ranjasmara., A. Khairan., F. Tempola., and R. Rosihan. 2020. "Implementasi Metode Naive Bayes Untuk Memprediksi Resiko Penyakit Jantung," *PATRIA ARTHA TECHNOLOGICAL JOURNAL* . Vol. 4 (2), pp. 66–70.
- [11] A. Ali., A. Khairan., F. Tempola., and A. Fuad. 2021. "Application Of Naive Bayes to Predict the Potential of Rain in Ternate City," *E3S Web of Conferences* . vol. 328, pp. 1–5.
- [12] P. Harrington, P., 2012. *Machine Learning In Action*, Manning Publication, New York. USA.
- [13] P. Y. Saputra, D. H. Subhi, and F. Z. A. 2019. Winatama, "Implementasi Sentimen Analisis

- Komentar Channel Video Pelayanan Pemerintah Di YouTube Menggunakan Algoritma Naïve Bayes,” *J. Inform. Polinema*, vol. 5, no. 4, pp. 209–213, doi: 10.33795/jip.v5i4.259.
- [14] A. Indrawati. 2021. “Penerapan Teknik Kombinasi Oversampling Dan Undersampling Untuk Mengatasi Permasalahan Imbalanced Dataset,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 1, pp. 38–43, doi: 10.33387/jiko.v4i1.2561.
- [15] F. Gorunescu., 2011. *Data Mining: Concepts, Models, and Techniques*. Springer: London. UK