

# Analisis Sentimen Keberhasilan Debut Grup K-Pop Pada Platform X Menggunakan Algoritma *Support Vector Machine*

Indri Febriana Putri<sup>1</sup>, Wina Witanti<sup>2</sup>, Edvin Ramadhan<sup>3</sup>

Program Studi Informatika, Fakultas Sains & Informatika

Universitas Jenderal Achmad Yani

Cimahi, Indonesia

e-mail: <sup>1</sup>indrifp24@gmail.com, <sup>2</sup>witanti@gmail.com, <sup>3</sup>edvin.ramadhan@lecture.unjani.ac.id

Correspondence : e-mail: <sup>1</sup>indrifp24@gmail.com

Diajukan: 17 Agustus 2024; Direvisi: 19 Agustus 2024; Diterima: 20 Agustus 2024

## Abstrak

Grup K-pop baru memiliki peran penting dalam industri hiburan Korea yang terus berkembang dan meraih pengakuan global. Evaluasi keberhasilan debutnya sangat dipengaruhi oleh respons dan pandangan penggemar yang tersebar di berbagai platform media sosial. Penelitian ini bertujuan untuk mengungkap isu analisis sentimen terkait debut grup K-pop dan meramalkan keberhasilan debut tersebut. Aspek yang memerlukan perhatian khusus termasuk analisis dan pemrosesan data teks yang besar, kompleksitas dalam memahami variasi bahasa, serta penentuan sentimen yang subjektif. Kumpulan data yang digunakan mencakup komentar dari platform media sosial X yang berkaitan dengan debut grup K-pop. Pendekatan analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM) digunakan untuk mengklasifikasikan komentar menjadi positif dan negatif. Faktor-faktor sentimen yang berperan dalam prediksi, seperti respons penggemar terhadap teaser, lirik, dan video musik debut grup, telah diidentifikasi. Identifikasi faktor-faktor sentimen tersebut yang berperan dalam prediksi keberhasilan debut grup K-pop memberikan wawasan penting dalam mengarahkan perkembangan algoritma dan model untuk menganalisis respons penggemar terhadap konten hiburan. Penelitian ini memberikan kontribusi penting dalam pemahaman tentang dampak sentimen penggemar terhadap keberhasilan debut grup K-pop, serta memberikan panduan bagi praktisi industri hiburan untuk merancang strategi promosi yang lebih responsif terhadap sentimen penggemar. Penelitian ini dapat menjadi landasan bagi pengembangan lebih lanjut dalam analisis sentimen di dunia hiburan, memperkuat penggunaan teknologi dalam mendukung keputusan strategis dalam industri hiburan, serta mendorong pengembangan metode analisis sentimen yang lebih baik dalam konteks yang terus berubah.

**Kata kunci:** analisis, K-Pop, negatif, positif, prediksi, sentimen, SVM.

## Abstract

New K-pop groups play an important role in Korea's growing entertainment industry and are gaining global recognition. The evaluation of their debut success is greatly influenced by fans' responses and views spread across various social media platforms. This research aims to uncover sentiment analysis issues related to K-pop group debuts and forecast the success of their debuts. Aspects that require special attention include the analysis and processing of large text data, the complexity of understanding language variations, and the subjective determination of sentiment. The dataset used includes comments from X social media platform relating to K-pop group debuts. A sentiment analysis approach using *Support Vector Machine* (SVM) algorithm was used to classify the comments into positive and negative. Sentiment factors that play a role in prediction, such as fan responses to the group's debut teaser, lyrics and music video, were identified. The identification of such sentiment factors that play a role in predicting the success of a K-pop group's debut provides important insights in directing the development of algorithms and models for analyzing fan responses to entertainment content. This research makes an important contribution to the understanding of the impact of fan sentiment on the success of K-pop group debuts, and provides guidance for entertainment industry practitioners to design promotional strategies that are more responsive to fan sentiment.

**Keywords** analysis, K-Pop, negative, positive, prediction, sentiment, SVM.

### 1. Pendahuluan

Sebagai makhluk sosial, manusia membutuhkan komunikasi dan sosialisasi [1]. Perkembangan internet telah memperluas pilihan untuk berinteraksi, dengan media sosial sebagai ruang publik baru yang menyediakan informasi terkini, data dan pengetahuan. Salah satunya adalah *Twitter* sebagai media sosial yang paling banyak digunakan saat ini, dengan Indonesia menempati peringkat keempat dalam hal jumlah pengguna [2]. Sejak 22 Juli 2023, *Twitter* resmi berganti nama menjadi X, menawarkan berbagai fitur untuk alat komunikasi [3], [4].

Topik yang sering menjadi perbincangan di X adalah industri hiburan, khususnya K-pop. Industri K-pop telah berkembang pesat dan menjadi fenomena global dalam beberapa tahun terakhir [5]. Debut grup K-pop baru selalu ditunggu-tunggu oleh penggemar dan menarik perhatian publik. Penggemar dapat berbagi pendapat dan perasaan mereka tentang debut grup baru melalui platform X [6]. Analisis sentimen terhadap komentar dan *tweet* penggemar di X dapat memberikan wawasan berharga tentang penerimaan publik terhadap grup K-pop yang baru debut [7]. Informasi ini penting bagi agensi hiburan, promotor, dan peneliti industri hiburan untuk memahami tren dan preferensi penggemar [7]. Dalam konteks ini, penggunaan teknik pembelajaran mesin *Support Vector Machines* (SVM) untuk analisis sentimen sangat penting [8].

Beberapa penelitian terdahulu menggunakan berbagai algoritma pembelajaran mesin untuk menganalisis sentimen tentang K-pop atau topik hiburan serupa, tetapi hasilnya menunjukkan bahwa beberapa algoritma memiliki akurasi yang lebih rendah daripada SVM. Penelitian [9] menggunakan algoritma Naive Bayes untuk menganalisis sentimen penggemar K-pop di *Twitter*, namun tingkat akurasinya hanya mencapai 75%. Studi lain [10] menggunakan algoritma *Ensemble* untuk menganalisis sentimen terkait deteksi ujaran kebencian di *Twitter*, dengan akurasi sebesar 79,8%. Sementara itu, penelitian [11] menggunakan algoritma *K-Nearest Neighbor* (K-NN) untuk analisis sentimen, menghasilkan akurasi sebesar 87%. Penelitian [12] menunjukkan bahwa algoritma *Decision Tree* mencapai akurasi 83,3%. Penelitian [13] menemukan bahwa algoritma *Naive Bayesian Classification* menghasilkan akurasi 72%, sedangkan *Support Vector Machine* menghasilkan akurasi 92%.

Berdasarkan pada latar belakang yang telah diuraikan, meskipun algoritma seperti Naive Bayes, *Ensemble*, K-NN, *Decision Tree*, dan *Naive Bayesian Classification* dapat digunakan untuk analisis sentimen, akurasi yang dihasilkan cenderung rendah. Oleh karena itu, pada penelitian ini digunakan algoritma *Support Vector Machine* (SVM) karena memiliki kemampuan generalisasi untuk mengklasifikasikan suatu pola, tidak termasuk data yang digunakan dalam fase pembelajaran metode tersebut sehingga dapat mencapai akurasi yang tinggi dalam analisis sentimen komentar penggemar K-pop di platform X. Meningkatkan efektivitas model SVM dalam analisis sentimen, kernel linear dipilih sebagai fungsi kernel. Kernel linear bekerja dengan memproyeksikan data ke dalam ruang fitur yang lebih tinggi, memungkinkan pemisahan data dengan hyperplane yang optimal. Penggunaan kernel linear sangat cocok untuk data yang relatif linier atau tidak terlalu kompleks, seperti teks komentar penggemar. Salah satu manfaat utama dari kernel linear adalah kesederhanaannya, yang mengarah pada komputasi yang lebih cepat dan model yang lebih mudah diinterpretasikan. Model dapat dengan cepat dan efisien mengklasifikasikan data tanpa memerlukan transformasi kompleks dengan menggunakan kernel linear.

Tantangan utama dari penggunaan kernel linear adalah keterbatasannya dalam menangani data yang sangat kompleks atau non-linear. Jika pola dalam data tidak dapat dipisahkan secara linier, model dengan kernel linear mungkin tidak memberikan hasil yang optimal. Selain itu, efektivitas kernel linear sangat tergantung pada pemilihan fitur yang relevan. Fitur yang tidak sesuai dapat mempengaruhi akurasi model secara signifikan. Meskipun demikian, dalam konteks analisis sentimen terhadap komentar penggemar K-pop, kernel linear dipilih karena kemampuannya untuk memberikan hasil yang efisien dan akurat pada data yang relatif linier, sesuai dengan tujuan penelitian ini untuk mencapai akurasi tinggi dalam klasifikasi sentimen.

### 2. Metode Penelitian

Alur proses penelitian yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1 - Alur Penelitian

## 2.1. Pengumpulan Data

Pengumpulan data diambil dari *platform* media sosial X dalam bentuk file yang berekstensi xls/xlsx menggunakan metode crawling web lalu setelah itu data diberikan label secara manual apakah data tersebut positif atau negatif pada dokumen. Total data yang terkumpul sebanyak 3.417 data *tweets* (1.033 data *tweets* Grup Babymonster, 1.676 data *tweets* Grup Zerobaseone, dan 708 data *tweets* Grup Kiss Of Life).

## 2.2. Preprocessing Data

*Preprocessing data* adalah tahapan penting dalam analisis sentimen yang bertujuan untuk membersihkan dan mempersiapkan data teks mentah agar siap untuk dianalisis oleh model pembelajaran mesin. Langkah - langkah yang dilakukan pada tahap *preprocessing data* antara lain:

1. *Case Folding* adalah proses untuk menghilangkan perbedaan antara huruf besar dan huruf kecil, proses *case folding* mengubah semua huruf dalam teks menjadi huruf kecil.
2. *Cleaning Data* merupakan proses *cleaning data* dilakukan dengan menghapus karakter yang tidak diinginkan dari teks yaitu menghapus semua karakter selain huruf kecil (a-z) dan spasi.
3. *Tokenization* adalah proses mengumpulkan Data teks yang diubah menjadi bentuk yang dapat dianalisis oleh model dengan memecah teks menjadi kata-kata atau token, sehingga setiap kata dipisahkan dan diidentifikasi secara individual.
4. *Stopwords Removal* adalah proses dihapusnya kata-kata umum yang tidak memiliki arti yang signifikan, seperti “dan”, “atau”, “di”, dan “ini”.
5. *Stemming* adalah proses mengubah kata-kata dalam data *tweet* yang memiliki imbuhan menjadi bentuk kata dasar yang sesuai.

## 2.3. Split Data

Pembelajaran mesin, yang utama adalah untuk membuat model yang dapat menggeneralisasi dengan baik pada data baru yang belum pernah dilihat sebelumnya. Agar model dapat diukur secara objektif, dataset yang tersedia harus dibagi menjadi beberapa subset yang berbeda untuk berbagai tujuan. Hal ini membantu memastikan bahwa penilaian kinerja model tidak bias dan model dapat diaplikasikan secara efektif pada data nyata. Dengan membagi data menjadi data latih dan data uji, model SVM dilatih hanya pada sebagian data dan diuji pada data yang tidak terlihat selama pelatihan. Ini membantu mencegah overfitting, di mana model mungkin belajar pola yang spesifik hanya untuk data latih dan tidak dapat menggeneralisasi dengan baik pada data baru. Setelah membagi dataset, program ini menggunakan perbandingan 80:20 untuk data latih dan data uji, yang artinya, 80% dari dataset digunakan untuk melatih model SVM (*Support Vector Machine*), dan 20% sisanya digunakan untuk menguji kinerja model.

## 2.4. Training Model

Tahapan dalam proses ini, data teks diubah menjadi format numerik yang dapat dipahami oleh model SVM. *Vectorizer TF-IDF (Term Frequency-Inverse Document Frequency)* mengubah teks menjadi vektor fitur yang menunjukkan pentingnya kata-kata dalam dokumen. Tahap TF-IDF, pemrosesan teks dilakukan untuk mengevaluasi seberapa penting suatu kata terhadap kumpulan kata lainnya dalam dokumen. Semakin sering kata muncul dalam dokumen tetapi jarang di seluruh korpus, semakin besar bobot TF-IDF-nya. Metode ini digunakan dalam proses ekstraksi fitur teks di mana nilai TF-IDF yang lebih tinggi menunjukkan bahwa kata tersebut lebih penting atau mewakili dokumen yang relevan daripada kata lain. Untuk melakukan proses ini, fungsi *train\_model* digunakan. Berikut persamaan TF-IDF yang digunakan [14]:

1. *Term Frequency (TF)*:

$$TF(t,d) = \text{Jumlah kemunculan kata } t \text{ dalam dokumen } d \quad (1)$$

Dimana:

*t* adalah kata tertentu (term) yang frekuensinya ingin dihitung.

*d* adalah dokumen tertentu dimana frekuensi kata *t* dihitung.

2. *Inverse Document Frequency (IDF)*:

$$IDF(t) = \log(N/df) \quad (2)$$

Dimana:

$t$  adalah kata tertentu (*term*) yang frekuensinya ingin dihitung.

$N$  adalah jumlah total dokumen.

$df$  adalah jumlah dokumen yang mengandung kata  $t$ .

### 3. TF-IDF:

$$TF\ IDf = TF \times IDF \quad (3)$$

Dimana:

TF adalah Jumlah kemunculan kata  $t$  dalam dokumen yang telah dihitung.

IDF adalah *Inverse Document Frequency* yang telah dihitung.

Setelah data teks diubah menjadi format numerik yang dapat dipahami oleh model SVM, model SVM berusaha menemukan hyperplane ideal yang akan memisahkan kelas-kelas dalam data latih dengan *margin* yang paling besar. Selain itu, untuk masalah klasifikasi teksnya menggunakan kernel "*linear*". Tahap ini memungkinkan model yang telah dilatih juga dikenal sebagai model yang dilatih untuk memprediksi kelas data baru. Berikut persamaan SVM yang digunakan [15]:

$$f(x) = w \cdot x + b \quad (4)$$

Dimana:

$w$  : vektor bobot,

$x$  : vektor fitur input,

$b$  : bias (*intercept*).

## 2.5. Evaluation Model

Proses evaluasi model dalam sistem ini dimulai setelah model SVM dilatih menggunakan data latih. Kemudian, performa model yang telah dilatih dievaluasi menggunakan data uji. Data uji tersebut diubah menjadi vektor fitur menggunakan TF-IDF transformer yang telah dilatih bersamaan dengan model. Setelah data uji diubah, model melakukan prediksi terhadap data tersebut, menghasilkan label prediksi untuk setiap sampel dalam data uji. Hasil prediksi ini kemudian dibandingkan dengan label sebenarnya dari data uji untuk menghitung metrik evaluasi, seperti laporan klasifikasi yang mencakup presisi, *recall*, dan *f1-score* untuk setiap kelas. Selain itu, matriks kebingungan (*confusion matrix*) juga dihitung untuk menunjukkan distribusi prediksi benar dan salah untuk setiap kelas.

## 3. Hasil dan Pembahasan

### 3.1. Pengumpulan Data

Analisis dalam penelitian ini dilakukan dengan data yang diperoleh dari *crawling* dengan kata kunci yang dianalisis yaitu 'babymonster' sebanyak 1.033 data (981 positif & 52 negatif), 'zerobaseone' sebanyak 1.676 data (1.654 positif & 22 negatif), dan 'kissoflife' sebanyak 708 data (685 positif & 23 negatif) dan data ini diberi label secara manual untuk memastikan akurasi dalam proses evaluasi.

### 3.2. Preprocessing Data

Data yang didapat kemudian disiapkan sebelum displit. Ada 5 Langkah yaitu *Case Folding*, *Cleaning*, *Tokenization*, *Stopwords Removal*, dan *Stemming*. Pertama adalah *Case Folding*. *Case Folding* adalah proses untuk menghilangkan perbedaan antara huruf besar dan huruf kecil, proses *case folding* mengubah semua huruf dalam teks menjadi huruf kecil. Kedua *Cleaning Data* yang merupakan proses *cleaning data* dilakukan dengan menghapus karakter yang tidak diinginkan dari teks yaitu menghapus semua karakter selain huruf kecil (a-z) dan spasi. Setelah itu langkah ketiga adalah *Tokenization* yang mana merupakan proses mengumpulkan Data teks yang diubah menjadi bentuk yang dapat dianalisis oleh model dengan memecah teks menjadi kata-kata atau token. Keempat *Stopwords Removal*, adalah proses dihapusnya kata-kata umum yang tidak memiliki arti yang signifikan, seperti "dan", "atau", "di", dan "ini", dan kelima yaitu *Stemming*. *Stemming* adalah proses mengubah kata-kata dalam data *tweet* yang memiliki imbuhan menjadi bentuk kata dasar yang sesuai. Berikut hasil dari *preprocessing data* ditunjukkan pada Tabel 1.

Tabel 1 - Hasil *Preprocessing Data* Grup Babymonster

Teks Awal	Teks Hasil <i>Preprocessing Data</i>
Keluarga Nong Peem pun berangkat menjemput Canny di bandara. Kedua keluarga ini lucu sekali #CHIQUITA #BABYMONSTER <a href="https://t.co/p19JOB1tYj">https://t.co/p19JOB1tYj</a>	keluarga nong peem berangkat jemput canny bandara keluarga lucu chiquita babymonster <a href="httpstcopijobtyj">httpstcopijobtyj</a>

Tabel 2 - Hasil *Preprocessing Data* Grup Zerobaseone

Teks Awal	Teks Hasil <i>Preprocessing Data</i>
zb! Sangat keren bahwa Zerobaseone sangat bangga dengan In Bloom <a href="https://t.co/XUMxDxCvO0">https://t.co/XUMxDxCvO0</a>	zb keren zerobaseone bangga in bloom <a href="httpstcoxumdxcv0">httpstcoxumdxcv0</a>

Tabel 3 - Hasil *Preprocessing Data* Grup Kiss Of Life

Teks Awal	Teks Hasil <i>Preprocessing Data</i>
DIA SANGAT CANTIK #KISSOFLIFE #KIOF #HANEUL #Sky <a href="https://t.co/slfO8t8pbU">https://t.co/slfO8t8pbU</a>	cantik kissoflife kiof haneul sky <a href="httpstcoslft8pbu">httpstcoslft8pbu</a>

### 3.3. Evaluation Model

Setelah model SVM dilatih dengan data latih, proses evaluasi dilakukan menggunakan data uji. Data uji diubah menjadi vektor fitur dengan TF-IDF transformer yang sama. Model kemudian memprediksi label untuk data uji, dan hasil prediksi dibandingkan dengan label sebenarnya untuk menghitung metrik evaluasi seperti presisi, *recall*, dan *f1-score*, serta matriks kebingungan (*confusion matrix*) untuk menunjukkan distribusi prediksi yang benar dan salah. Hasil evaluasi model dijelaskan pada Gambar 2 sampai Gambar 4.

```
Classification Report:
precision recall f1-score support
negatif 0.00 0.00 0.00 10
positif 0.95 1.00 0.98 197

accuracy 0.95 207
macro avg 0.48 0.50 0.49 207
weighted avg 0.91 0.95 0.93 207

Confusion Matrix:
[[ 0 10]
 [ 0 197]]
```

Gambar 2 - Hasil Evaluasi Model Grup Babymonster

```
Classification Report:
precision recall f1-score support
negatif 0.00 0.00 0.00 5
positif 0.99 1.00 0.99 331

accuracy 0.99 336
macro avg 0.49 0.50 0.50 336
weighted avg 0.97 0.99 0.98 336

Confusion Matrix:
[[ 0 5]
 [ 0 331]]
```

Gambar 3 - Hasil Evaluasi Model Grup Zerobaseone

```
Classification Report:
precision recall f1-score support
negatif 1.00 0.25 0.40 4
positif 0.98 1.00 0.99 138

accuracy 0.98 142
macro avg 0.99 0.62 0.69 142
weighted avg 0.98 0.98 0.97 142

Confusion Matrix:
[[ 1 3]
 [ 0 138]]
```

Gambar 4 - Hasil Evaluasi Model Grup Kiss Of Life

Gambar 2 berisi hasil klasifikasi Grup Babymonster yang memiliki akurasi sebesar 95%, Gambar 3 berisi hasil klasifikasi Grup Zerobaseone yang memiliki akurasi sebesar 99%, dan hasil klasifikasi Grup Kiss of Life yang memiliki akurasi sebesar 98%, serta *confusion matrix* yang menunjukkan data matriks dari setiap class setelah dilakukan testing menggunakan model.

## 4. Kesimpulan

### 4.1. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa implementasi metode *Support Vector Machine* (SVM) untuk analisis sentimen keberhasilan debut grup K-pop di *platform X* telah berhasil dilakukan dengan efektif. Proses ini mencakup tahapan *preprocessing data*, pembagian data, pelatihan model, dan evaluasi. Hasil pengujian menunjukkan tingkat akurasi yang sangat tinggi untuk ketiga grup yang diteliti yaitu Babymonster dengan akurasi 95%, Zerobaseone 99%, dan Kiss of Life 98%. Tingkat akurasi yang konsisten tinggi ini membuktikan bahwa metode SVM sangat efektif dan andal dalam mengklasifikasikan sentimen terkait debut grup K-pop di *platform X*, memberikan wawasan berharga tentang persepsi publik terhadap keberhasilan debut grup-grup tersebut.

### 4.2. Saran

Saran penelitian selanjutnya untuk data teks yang dinamis, maka diperlukan eksplorasi algoritma lain seperti *Ensemble* atau LSTM yang digunakan untuk melihat apakah hasil didapat bisa lebih baik dengan teknik yang lebih kompleks.

## Daftar Pustaka

- [1] A. Anggraeni, "Budaya Curhat di Jejaring Sosial," 19 November. Accessed: Jul. 17, 2024. [Online]. Available: <https://www.kompasiana.com/aryanggraeni/552a2f1d6ea8347b32552d61/budaya-curhat-di-jejaring-sosial>
- [2] C. M. Annur, "Jumlah Pengguna *Twitter* Indonesia Duduki Peringkat ke-4 Dunia per Juli 2023," Databoks. Accessed: Jul. 17, 2024. [Online]. Available: [https://databoks.katadata.co.id/datapublish/2023/11/01/jumlah-pengguna-twitter-indonesia-duduki-peringkat-ke-4-dunia-per-juli-2023#:~:text=Posisi itu naik dari laporan,%25 secara kuartalan \(qtq\).](https://databoks.katadata.co.id/datapublish/2023/11/01/jumlah-pengguna-twitter-indonesia-duduki-peringkat-ke-4-dunia-per-juli-2023#:~:text=Posisi%20itu%20naik%20dari%20laporan,%25%20secara%20kuartalan%20(qtq).)
- [3] Bayu Adi Isnanto, "Kenapa *Twitter* Jadi X? Ini Penjelasan dan Perubahan Fiturnya Baca artikel detikinet, 'Kenapa *Twitter* Jadi X? Ini Penjelasan dan Perubahan Fiturnya,'" Inet.Detik.Com. [Online]. Available: <https://inet.detik.com/cyberlife/d-6869515/kenapa-twitter-jadi-x-ini-penjelasan-dan-perubahan-fiturnya>
- [4] T. O. and S. Milstein, *The Twitter Book, 2nd Edition*. O'Reilly Media, Inc., 2011. [Online]. Available: <https://learning.oreilly.com/library/view/the-twitter-book/9781449318949/>
- [5] I. Oh and G. S. Park, "From B2C to B2B: Selling Korean pop music in the age of new social media," *Korea Obs.*, vol. 43, no. 3, pp. 365–397, 2012.
- [6] S. Kim and Y. I. Yoon, "Recommendation system for sharing economy based on multidimensional trust model," *Multimed. Tools Appl.*, vol. 75, no. 23, pp. 15297–15310, 2016, doi: 10.1007/s11042-014-2384-5.
- [7] J. H. Kim, K. J. Kim, B. T. Park, and H. J. Choi, "The Phenomenon and Development of K-Pop: The Relationship between Success Factors of K-Pop and the National Image, Social Network Service Citizenship Behavior, and Tourist Behavioral Intention," *Sustain.*, vol. 14, no. 6, 2022, doi: 10.3390/su14063200.
- [8] I. Pilászy, "Text categorization and support vector machines," *Proc. 6th Int. Symp. Hungarian Res. Comput. Intell.*, vol. 1, pp. 1–10, 2005.
- [9] P. Nurmawati, E. Supriyati, and T. Listyorini, "Analisis Sentimen Terhadap Penggemar K-Pop Di Media Sosial *Twitter* Menggunakan Naive Bayes (Studi Kasus Penggemar Grup Bts)," *J. Inf. Eng. Educ. Technol.*, vol. 4, no. 2, pp. 86–89, 2020, doi: 10.26740/jieet.v4n2.p86-89.
- [10] M. A. Fauzi and A. Yuniarti, "Ensemble method for indonesian *twitter* hate speech detection," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 11, no. 1, pp. 294–299, 2018, doi: 10.11591/ijeecs.v11.i1.pp294-299.
- [11] A. Tanggu Mara, E. Sedyono, and H. Purnomo, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba," *Jointer - J. Informatics Eng.*, vol. 2, no. 01, pp. 24–31, 2021, doi: 10.53682/jointer.v2i01.30.
- [12] M. Syarifuddin, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada *Twitter* Dengan Algoritma Decision Tree, Knn, Dan Naive Bayes," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.
- [13] S. F. Rodiyansyah, "Klasifikasi Posting *Twitter* Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification," vol. 6, no. 1, 2012, doi: 10.1163/ej.9789004182127.i-302.6.
- [14] B. Mustapic, "An Introduction to TF-IDF: What It Is & How to Use It." [Online]. Available: <https://www.semrush.com/blog/tf-idf/>

- 
- [15] A. A. Kasim and M. Sudarsono, “Algoritma *Support Vector Machine* (SVM) untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah di Kecamatan Simpang Raya Sulawesi Tengah,” *Semin. Nas. APTIKOM*, pp. 568–573, 2019.