

PERBANDINGAN METODE *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE* UNTUK ANALISIS SENTIMEN FENOMENA *HALLYU*

1st Cheppi Garda Muhamad
S1 Teknik Informatika
Universitas Telkom Purwokerto
Purwokerto, Indonesia
21102163@ittelkom-pwt.ac.id

2nd Paradise, S.Kom., M.Kom
S1 Teknik Informatika
Universitas Telkom Purwokerto
Purwokerto, Indonesia
paradise@telkomuniversity.ac.id

Abstrak — Pesatnya kemajuan teknologi komunikasi dan informasi, khususnya pada media sosial 'X', telah membuka akses luas bagi masyarakat untuk menerima beragam informasi global, termasuk aspek budaya. Di antara berbagai fenomena budaya yang muncul, Korean Wave atau Hallyu menjadi subjek diskusi yang memicu beragam tanggapan di kalangan pengguna platform tersebut. Studi ini dilaksanakan untuk mengevaluasi pandangan masyarakat terhadap Hallyu menggunakan perbandingan dua pendekatan algoritma, yakni *Naïve Bayes* dan *Support Vector Machine (SVM)*. Pengumpulan data dilakukan melalui proses ekstraksi tweet terkait Hallyu dalam rentang 1 Januari 2018 hingga 31 Desember 2024, menghasilkan 2855 tweet netral, 430 tweet positif, dan 90 tweet negatif. Evaluasi kinerja menunjukkan *Naïve Bayes* memperoleh akurasi 73,4%, sedangkan *SVM* menunjukkan performa lebih unggul dengan akurasi 86,6%, mengindikasikan keunggulan *SVM* dalam klasifikasi sentimen Hallyu.

Kata kunci— analisis sentimen, NB, SVM, hallyu

I. PENDAHULUAN

Perkembangan globalisasi dan teknologi informasi, terutama melalui internet, telah mengubah banyak aspek kehidupan, termasuk penyebaran budaya. Salah satu fenomena yang paling mencolok adalah Hallyu, atau Korean Wave, yang tercermin melalui industri musik Korea Selatan dan tayangan drama televisi dari negara tersebut. Pengaruh Hallyu semakin terasa di seluruh dunia, berkat berkembangnya akses jaringan internet yang memungkinkan penyebaran pop culture secara cepat [1]. Data dari Lowy Institute menunjukkan bahwa Korea Selatan berada di posisi tujuh sebagai negara dengan budaya paling populer di dunia pada tahun 2023 [2]. K-pop dan K-drama menduduki posisi teratas sebagai produk Hallyu yang paling diminati masyarakat, di mana statistik menunjukkan bahwa Indonesia tercatat sebagai salah satu negara yang paling aktif sepanjang tahun 2020-2021 [5].

Fenomena ini juga didorong oleh peran media sosial, khususnya platform "X", yang menjadi saluran utama untuk penyebaran budaya Korea Selatan [4].

Berdasarkan survei dari We Are Social, hingga April 2024, platform "X" masih menjadi salah satu media sosial dengan jumlah pengguna aktif terbanyak di dunia, mencapai lebih dari 611 juta pengguna [6]. Namun, fenomena Hallyu tidak lepas dari pro dan kontra. Beberapa pihak berpendapat bahwa pengaruh budaya Korea dapat memberikan dampak positif bagi anak muda, seperti menjadi inspirasi dan media untuk bersosialisasi, sementara kelompok lain mengkhawatirkan dampak negatifnya terhadap budaya lokal, termasuk fanatisme berlebihan dan pengaruh terhadap perilaku sosial [7]. Data demografi tentang minat masyarakat terhadap K-pop dan K-drama menunjukkan bahwa kedua budaya ini menduduki posisi teratas di antara produk budaya Korea yang paling digemari [3].

Untuk memahami pandangan masyarakat terhadap fenomena ini, penelitian ini bertujuan untuk menganalisis sikap publik terhadap Hallyu menggunakan dua teknik klasifikasi sentimen, yaitu *Naïve Bayes* dan *Support Vector Machine (SVM)*. Penelitian ini menggunakan dataset yang diambil dari platform "X" antara 1 Januari 2018 hingga 31 Desember 2024, yang mencakup tweet dengan kata kunci terkait Hallyu. Metode *Naïve Bayes* dan *SVM* dipilih karena keduanya sering digunakan dalam analisis sentimen, meskipun memiliki perbedaan dalam hal pendekatan dan kinerja. Dengan melakukan perbandingan antara kedua metode tersebut, penelitian yang dilaksanakan ini diharapkan dapat memberikan pemahaman dan wawasan yang lebih dalam mengenai sikap masyarakat terhadap Hallyu serta membandingkan efektivitas kinerja masing-masing metode dalam menganalisis data media sosial [8][9].

II. KAJIAN TEORI

A. Analisis Sentimen

Analisis sentimen adalah sebuah metode dalam pemrosesan bahasa alami atau *Natural Language Processing (NLP)* yang digunakan untuk mengategorikan teks ke dalam kategori positif, negatif,

atau netral [12]. Teknik ini umumnya diterapkan untuk mengevaluasi opini masyarakat terhadap objek tertentu, dengan memanfaatkan data dari media sosial. Banyak perusahaan menggunakan analisis sentimen untuk menilai kepuasan pelanggan terhadap produk mereka. Keunggulan utama dari analisis sentimen adalah kemampuannya untuk memberikan hasil objektif, mengurangi kesalahan manusia, serta menangani volume data yang besar. Selain itu, hasil yang didapatkan bersifat langsung (real-time), yang memungkinkan pengambilan keputusan yang lebih cepat dan lebih tepat [23], [24].

B. Naïve Bayes

Naïve Bayes adalah sebuah metode klasifikasi yang didasarkan pada probabilitas, yang menghitung hasil klasifikasi dengan menjumlahkan frekuensi dan nilai dari dataset [8]. Algoritma *Naïve Bayes* memiliki beberapa varian yang bergantung pada distribusi data, salah satunya adalah *GaussianNB*, yang menggunakan distribusi Gaussian untuk variabel kontinu serta menghitung rata-rata dan deviasi standar pada setiap kelas. Selanjutnya, *MultinomialNB* mengasumsikan fitur mengikuti distribusi multinomial, sering diterapkan pada *Natural Language Processing* (NLP) seperti klasifikasi spam. Varian lainnya, *BernoulliNB*, menggunakan variabel Boolean (benar/salah atau 1/0) dalam proses klasifikasinya.

C. Support Vector Machine

Support Vector Machine (SVM) adalah sebuah algoritma pembelajaran mesin yang diawasi, digunakan untuk tugas klasifikasi dan regresi dengan cara menemukan *hyperplane* terbaik yang memaksimalkan jarak antara kelas-kelas dalam ruang berdimensi tinggi. *Hyperplane* ini berfungsi sebagai pemisah antar kelas dalam data [8]. SVM pertama kali dikembangkan pada dekade 1990-an oleh Vladimir N. Vapnik dan dipublikasikan dalam sebuah artikel dengan judul *Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing* di tahun 1995. Berdasarkan jenis batas keputusan yang digunakan, SVM dapat dibedakan menjadi dua jenis: Linear SVM, yang memisahkan data menggunakan garis lurus, dan Non-linear SVM, yang digunakan ketika data tidak dapat dipisahkan secara linier. Pada situasi non-linear, fungsi kernel diterapkan untuk mengubah data ke dalam ruang fitur yang memiliki dimensi lebih tinggi, yang memungkinkan pemisahan yang lebih tepat.

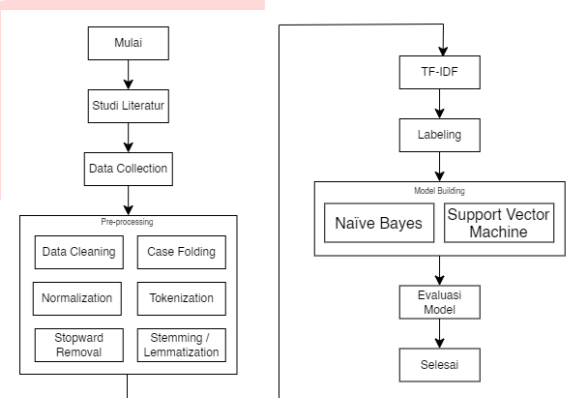
D. Platform "X"

Platform "X" adalah media sosial yang memungkinkan pengguna untuk mengekspresikan opini secara terbuka [20]. Diluncurkan pada 2006 oleh Jack Dorsey dan rekan-rekannya, platform ini berkembang pesat dengan lebih dari 300 juta pengguna. Karakter terbatas 280 dalam setiap postingan dianggap ideal untuk berbagi pemikiran secara singkat. Selain teks, platform ini juga mendukung gambar dan video, meningkatkan daya tariknya bagi pengguna untuk berbagi pengalaman [21]. Sebelumnya bernama

Twitter, platform ini diakuisisi oleh Elon Musk pada 2022 dan berganti nama menjadi "X" pada Juli 2023 [22].

III. METODE

Penelitian ini dilaksanakan dengan tujuan untuk menganalisis sentiment terhadap data yang diperoleh dari platform "X", menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM). Tahapan-tahapan yang dilakukan dalam penelitian ini meliputi studi literatur, *pre-processing data*, ekstraksi fitur, pembangunan model, dan evaluasi model. Setiap tahapan bisa dilihat pada Gambar 1.



GAMBAR 1
(DIAGRAM ALIR PENELITIAN)

1. Studi Literatur

Studi literatur dilakukan untuk mencari referensi ilmiah yang relevan dengan penelitian yang dilakukan. Artikel ilmiah yang digunakan adalah yang terbit dalam jangka waktu lima tahun terakhir dan diperoleh melalui pencarian online menggunakan platform seperti Google Scholar, ScienceDirect, dan IEEE Xplore.

2. Pre-processing Data

Pre-processing data atau pra-pemrosesan data dilakukan untuk mempersiapkan dataset mentah yang diperoleh dari platform "X" agar siap digunakan dalam analisis. Data mentah yang dikumpulkan seringkali mengandung noise dan elemen yang tidak relevan. Tahap ini terdiri dari :

- Pembersihan data** : Proses menghilangkan elemen-elemen yang tidak diperlukan, seperti angka, tanda baca, tautan (link), hashtag, emoticon, dan karakter non-Latin.
- Case Folding** : Mengonversi semua huruf ke dalam bentuk huruf kecil (lowercase) untuk memastikan konsistensi.
- Tokenization** : Memecah kalimat menjadi unit-unit kata (token) guna mempermudah proses analisis.
- Stopword Removal** : Menghilangkan kata-kata yang tidak memiliki makna signifikan dalam konteks analisis.

- e. *Stemming/Lemmatization* : Mengubah kata-kata ke bentuk dasar dengan menghapus imbuhan atau variasi kata.

3. Feature Extraction

Feature Extraction dilakukan dengan metode TF-IDF untuk mengubah teks menjadi numerik. Pembobotan dilakukan menggunakan bantuan library scikit-learn yang akan menghitung frekuensi kata dan invers frekuensi dokumen untuk setiap kata dalam teks.

4. Pelabelan

Pelabelan dilakukan secara otomatis dengan menggunakan bantuan library transformer yang dapat memberikan label pada tweet dari platform "X". Library ini membagi data tweet menjadi tiga kelas yaitu netral, positif, dan negatif.

5. Pembuatan Model

Dalam penelitian ini, dua model utama yang diterapkan adalah *Naïve Bayes* dan *Support Vector Machine* (SVM). Untuk konfigurasi, *Naïve Bayes* menggunakan pendekatan Multinomial, sementara SVM diimplementasikan dengan kernel RBF sebagai dasar pemrosesan datanya.

6. Evaluasi Model

Untuk mengevaluasi kinerja model dalam analisis sentimen, *Confusion Matrix* digunakan sebagai alat ukur utama. Metode ini memungkinkan peneliti menghitung tingkat akurasi model dengan membandingkan hasil prediksi dan data aktual.

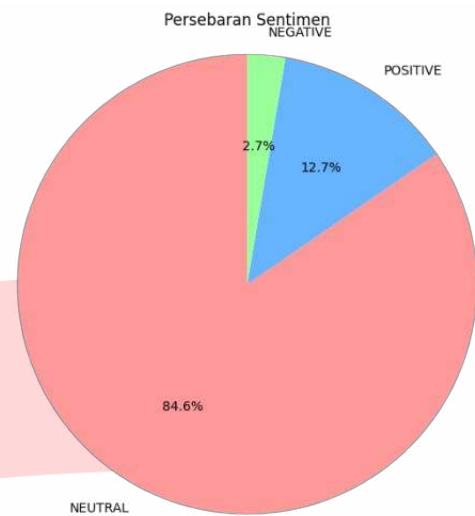
$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (1)$$

IV. HASIL DAN PEMBAHASAN

A. Hasil

Dataset penelitian diperoleh melalui teknik crawling pada platform "X" dalam rentang waktu 1 Januari 2018–31 Desember 2024. Data yang dikumpulkan melibatkan tweet dengan kata kunci terkait dengan fenomena Hallyu, yang menghasilkan total 3.375 tweet sebelum dilakukan proses pre-processing. Setelah tahap pre-

processing dan pelabelan, diperoleh dataset final yang terdiri dari 2.855 data berlabel netral (84,6%), 430 data berlabel positif (12,7%), dan 90 data berlabel negatif (2,7%). Persebaran data setiap kelas dapat dilihat pada Gambar 2.



GAMBAR 2
(PERSEBARAN SENTIMEN)

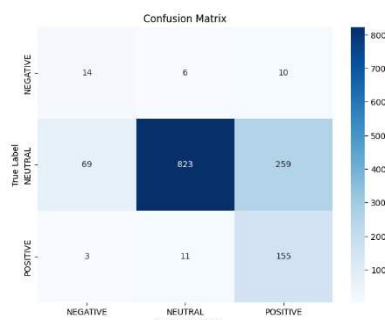
Setelah percobaan dilakukan dengan membagi dataset untuk pelatihan dan pengujian dengan proporsi 60:40, diperoleh tingkat akurasi yang bervariasi. Model *Multinomial Naïve Bayes* menghasilkan akurasi sebesar 0,734, sementara model *Support Vector Machine* dengan kernel RBF memperoleh akurasi sebesar 0,866.

B. Pembahasan

Setiap metode yang diterapkan dalam penelitian ini memberikan tingkat akurasi yang berbeda, meskipun menggunakan dataset yang identik dan rasio pembagian data untuk pelatihan dan pengujian yang serupa. Kinerja *Naïve Bayes* dalam menganalisis sentimen dapat dilihat pada Gambar 3. Berdasarkan data yang ada pada Gambar 3 maka kinerja yang didapat oleh model *Naïve Bayes* dapat dilihat pada Tabel 1.

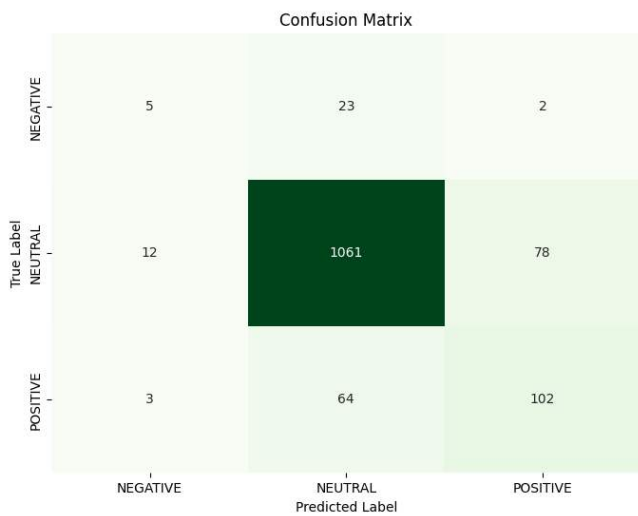
TABEL 1
(CLASSIFICATION REPORT NB)

Kelas	Akurasi	Precision	Recall	f1-score
Negatif	0,734	0,16	0,47	0,24
Positif		0,37	0,92	0,52
Netral		0,98	0,72	0,83



GAMBAR 3
(CONFUSION MATRIX NB)

Sementara itu, kinerja model *Support Vector Machine* dalam menganalisis sentimen dapat dilihat pada Gambar 4.



GAMBAR 4 (CONFUSION MATRIX SVM)

Berdasarkan informasi yang terdapat pada Gambar 4, kinerja yang diperoleh dari model *Support Vector Machine* dapat dilihat pada Tabel 2.

TABEL 2 (CLASSIFICATION REPORT SVM)

Kelas	Akurasi	Precision	Recall	f1-score
Negatif	0,866	0,25	0,17	0,20
Positif		0,56	0,60	0,58
Netral		0,92	0,92	0,92

Pada penelitian ini didapatkan hasil bahwa metode *Support Vector Machine* mendapat tingkat akurasi yang lebih tinggi dibandingkan metode *Naïve Bayes* dalam melakukan analisis sentimen.

V. KESIMPULAN

Berdasarkan data yang didapat dan analisis yang telah dilaksanakan, dapat disimpulkan bahwa pandangan masyarakat terhadap fenomena Hallyu di platform 'X' cenderung netral. Hal ini terlihat dari jumlah data berlabel netral yang mencapai sekitar 2855 entri atau sekitar 84,6% dari total data, sementara data dengan sentimen positif berjumlah 430 entri atau sekitar 12,7%, dan data dengan sentimen negatif hanya sebanyak 90 entri atau sekitar 2,7%. Selain itu, kedua metode yang diterapkan dalam penelitian ini, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM), berhasil mengklasifikasikan sentimen dengan tingkat akurasi yang cukup tinggi. Namun, metode SVM menunjukkan hasil yang lebih unggul dengan mencapai akurasi sekitar 86,6%, sementara *Naïve Bayes* memperoleh akurasi sebesar 73,4%. Oleh karena itu, dapat disimpulkan bahwa SVM lebih efisien dalam menganalisis sentimen masyarakat mengenai fenomena Hallyu di platform 'X' pada penelitian ini.

REFERENSI

- [1] L. Anggraini Putri, 'Dampak Korea Wave Terhadap Prilaku Remaja'.
- [2] lowyinstitute.org, 'Cultural Influence Ability to shape international public opinion through cultural appeal and interaction', lowyinstitute.org. Accessed: May 06, 2024. [Online]. Available: <https://power.lowyinstitute.org/data/cultural-influence/>
- [3] Adi Ahdiat, 'K-Pop vs K-Drama, Mana yang Penggemarnya Lebih Banyak?', databoks.katadata.co.id. Accessed: May 06, 2024. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2022/07/24/k-pop-vs-k-drama-mana-yang-penggemarnya-lebih-banyak>
- [4] cnnindonesia.com, 'Survei: Fan Hallyu di Dunia Lebih dari 178 Juta Orang pada 2022 Baca artikel CNN Indonesia "Survei: Fan Hallyu di Dunia Lebih dari 178 Juta Orang pada 2022"', cnnindonesia.com. Accessed: May 08, 2024. [Online]. Available: <https://www.cnnindonesia.com/hiburan/20230323150343-241-928582/survei-fan-hallyu-di-dunia-lebih-dari-178-juta-orang-pada-2022>
- [5] F. Javier, 'Ada 7,5 Miliar Twit K-Pop pada Juli 2020-Juni 2021, Terbanyak dari Indonesia', data.tempo.co.
- [6] datareportal.com, 'Global Social Media Statistics', datareportal.com. Accessed: Jun. 13, 2024. [Online]. Available: <https://datareportal.com/social-media-users>
- [7] K. Amaliantami Putri, M. Hadi Purnomo, and K. Z. Wave dalam Fanatisme dan Konstruksi Gaya Hidup Generasi, 'Korean Wave dalam Fanatisme dan Konstruksi Gaya Hidup Generasi Z', 2019.
- [8] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, 'Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter'.
- [9] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, 'A Comparative Study of Support Vector Machine and Naïve Bayes Classifier for Sentiment Analysis on Amazon Product Reviews', in *2020 International Conference on Contemporary Computing and Applications, IC3A 2020*, Institute of Electrical and Electronics Engineers Inc., Feb. 2020, pp. 217–220. doi: 10.1109/IC3A48958.2020.233300.
- [10] Z. Li, R. Li, and G. Jin, 'Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary', *IEEE Access*, vol. 8, pp. 75073–75084, 2020, doi: 10.1109/ACCESS.2020.2986582.
- [11] K. H. Manguri, R. N. Ramadhan, and P. R. Mohammed Amin, 'Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks', *Kurdistan Journal of Applied Research*, pp. 54–65, May 2020, doi: 10.24017/covid.8.
- [12] F. P. P. Subandi, F. Romadlon, I. Nurisusilawati, and A. Chindyana, 'Sentiment Analysis of Indonesian Interest in Korean Food Based on Naïve Bayes Algorithm', *Jurnal Sosioteknologi*, vol. 21, no. 3, pp. 337–346, Dec. 2022, doi: 10.5614/sostek.itbj.2022.21.3.10.
- [13] C. A. Nurhaliza Agustina, R. Novita, Mustakim, and N. E. Rozanda, 'The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm', in *Procedia*

- Computer Science*, Elsevier B.V., 2024, pp. 156–163. doi: 10.1016/j.procs.2024.02.162.
- [14] I. P. Wardhani, Y. I. Chandra, and F. Yusra, 'Application of the Naïve Bayes Classifier Algorithm to Analyze Sentiment for the Covid-19 Vaccine on Twitter in Jakarta', *International Journal of Innovation in Enterprise System*, vol. 7, no. 01, pp. 1–18, Jan. 2023, doi: 10.25124/ijies.v7i01.171.
- [15] Korean Cultural Center, 'Tentang Korea', Korean Cultural Center. Accessed: May 20, 2024. [Online]. Available: <https://id.korean-culture.org/id/144/korea/46>
- [16] R. Amelia, D. Darmansah, N. S. Prastiwi, and M. E. Purbaya, 'Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter', *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 338, Apr. 2022, doi: 10.30865/jurikom.v9i2.3895.
- [17] L. Toyryla, 'Kdrama – The Complete Guide to Korean Dramas', 90daykorean.com. Accessed: Jun. 03, 2024. [Online]. Available: <https://www.90daykorean.com/kdrama/#what-is-kdrama>
- [18] C. Tri Rahmawati, "The Massive Korean Wave in Indonesia and Its Effects in the Term of Culture".
- [19] J. Geroes, 'What is Kpop? Here is everything you need to know', 90daykorean.com. Accessed: Jun. 03, 2024. [Online]. Available: <https://www.90daykorean.com/what-is-kpop/#what-does-kpop-stand-for>
- [20] A. Amri, 'Implementasi Algoritma Random Forest Untuk Mendeteksi Hate Speech Dan Abusive Language Pada Twitter Bahasa Indonesia', Universitas Islam negeri Sultan Syarif Kasim, Pekanbaru, Riau, 2020.
- [21] Britannica.com, 'X', Britannica.com.
- [22] Max, 'Apa Itu Twitter?: Asal Usul dan Tren', ecwid.com. Accessed: May 23, 2024. [Online]. Available: <https://www.ecwid.com/id/blog/what-is-twitter.html>
- [23] ibm.com, 'What is sentiment analysis?', ibm.com. Accessed: May 26, 2024. [Online]. Available: <https://www.ibm.com/topics/sentiment-analysis>
- [24] sakshiparikh, 'What is Sentiment Analysis?', geeksforgeeks.org. Accessed: May 26, 2024. [Online]. Available: <https://www.geeksforgeeks.org/what-is-sentiment-analysis/>
- [25] F. Amaliah, I. Kadek, and D. Nuryana, 'Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter', *Journal of Informatics and Computer Science*, vol. 03, 2022.
- [26] geeksforgeeks.org, 'Naive Bayes Classifiers', geeksforgeeks.org.
- [27] ibm.com, 'What are Naïve Bayes classifiers?', ibm.com. Accessed: May 27, 2024. [Online]. Available: <https://www.ibm.com/topics/naive-bayes>
- [28] revou.co, 'Apa itu Algoritma Naive Bayes?', revou.co. Accessed: May 28, 2024. [Online]. Available: <https://revou.co/kosakata/algoritma-naive-bayes>
- [29] R. Gandhi, 'Support Vector Machine — Introduction to Machine Learning Algorithms', towardsdatascience.com. Accessed: May 30, 2024. [Online]. Available: <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- [30] geeksforgeeks.org, 'Support Vector Machine (SVM) Algorithm', geeksforgeeks.org.
- [31] ibm.com, 'What are SVMs?', ibm.com. Accessed: May 30, 2024. [Online]. Available: <https://www.ibm.com/topics/support-vector-machine>