

SENTIMENT ANALYSIS OF THE INFLUENCE OF THE KOREAN WAVE IN INDONESIA USING THE NAIVE BAYES METHOD AND SUPPORT VECTOR MACHINE

ANALISIS SENTIMEN PENGARUH KOREAN WAVE DI INDONESIA DENGAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE

Natasha¹, Ryan Randy Suryono²

^{1,2}Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email: natasha@teknokrat.ac.id¹, ryan@teknokrat.ac.id²,

Abstract - This study analyzes public sentiment towards the influence of the Korean wave in Indonesia using the Naive Bayes and Support Vector Machine (SVM) methods. The Korean wave, as a popular cultural phenomenon from South Korea, has had a significant influence on various aspects of Indonesian society. The dataset consists of 6,237 tweets obtained through a crawling process on social media X, with 80% data divided for training and 20% for testing. The pre-processing process includes cleaning, case folding, tokenizing, stopwords, and stemming. Data imbalance in sentiment distribution is overcome by the SMOTE technique. The test results show that the SVM model has the highest accuracy of 88%, outperforming the Naive Bayes model with an accuracy of 81%. Performance evaluation using precision, recall, and F1-score shows that SVM is more consistent in classifying positive and negative sentiments. Data visualization is done using bar charts and word clouds to illustrate the main patterns and themes in discussions related to the Korean wave in Indonesia. However, this study has limitations, such as data is only taken from one social media platform, so the results are less representative of public opinion as a whole. Nevertheless, this study provides new insights into how Indonesian society responds to popular culture phenomena online. These findings can also be utilized by policy makers to support the development of creative industries based on popular culture.

Keywords - Sentiment Analysis, Korean Wave, Naive Bayes, Support Vector Machine, SMOTE.

Abstrak – Penelitian ini menganalisis sentimen masyarakat terhadap pengaruh korean wave di Indonesia menggunakan metode *Naive Bayes* dan *Support Vector Machine (SVM)*. Korean wave, sebagai fenomena budaya populer dari korea Selatan, yang telah memberikan pengaruh yang signifikan di berbagai aspek kehidupan masyarakat Indonesia. *Dataset* terdiri dari 6.237 tweet di peroleh melalui proses *crawling* di media sosial X, dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Proses pra-pemrosesan meliputi *cleaning*, *case folding*, *tokenizing*, *stopword*, dan *stemming*. Ketidakseimbangan data dalam distribusi sentimen diatasi dengan teknik SMOTE. Hasil pengujian menunjukkan bahwa model SVM memiliki akurasi tertinggi sebesar 88%, mengungguli model Naive bayes dengan akurasi 81%. Evaluasi performa menggunakan *precision*, *recall*, dan *F1-score* menunjukkan bahwa SVM lebih konsisten dalam mengklasifikasikan sentimen positif dan negatif. Visualisasi data dilakukan menggunakan diagram batang dan *word cloud* untuk menggambarkan pola dan tema utama dalam diskusi terkait korean wave di Indonesia. Namun, penelitian ini memiliki batasan, seperti data hanya di ambil hanya dari satu platform media sosial, sehingga hasilnya kurang mempresentasikan opini masyarakat secara keseluruhan. Meskipun demikian, penelitian ini memberikan wawasan baru tentang bagaimana masyarakat Indonesia merespon fenomena budaya populer secara online. Temuan ini juga dapat di manfaatkan oleh pembuat kebijakan untuk mendukung pengembangan industri kreatif berbasis budaya populer.

Kata Kunci - Analisis Sentimen, Korean Wave, Naive Bayes, Support Vector Machine, SMOTE

I. PENDAHULUAN

Pengaruh perkembangan zaman serta globalisasi, membuat budaya asing cepat masuk ke Indonesia [1]. Salah satu fenomena budaya asing yang memiliki dampak signifikan adalah Korean wave, yaitu penyebaran budaya pop korea Selatan yang semakin populer melalui musik (K-pop), drama (K-drama). Film, kuliner, mode dan gaya hidup. Korean wave mulai berkembang sejak tahun 1990-an. Keberhasilan drama korea pada tahun 2000-an menjadi Langkah awal ekspansi budaya korea ke berbagai negara, termasuk Indonesia. Selain itu, pemerintah korea Selatan pun secara aktif mendorong perkembangan industri hiburan melalui berbagai inisiatif dan investasi strategi, serta didukung oleh kemajuan teknologi sehingga mempercepat penyebaran pengaruh ini. Di Indonesia, Korean wave telah menjadi fenomena sosial yang sangat mempengaruhi tidak hanya penggemar budaya korea saja, tetapi masyarakat umum melalui berbagai media, terutama media sosial. Dalam hal ini, pentingnya analisis sentimen pada pengaruh Korean wave di Indonesia terletak dalam kemampuannya untuk mengerti dinamika perasaan, pandangan, serta tingkah laku masyarakat pada fenomena tersebut [2]. Media sosial X yang dulunya Twitter berperan besar dalam penyebaran pengaruh korean wave di Indonesia dengan menjadi ruang utama bagi masyarakat untuk membicarakan budaya korea, baik melalui unggahan, retweet, maupun diskusi daring yang melibatkan beragam perspektif dan menjadi sumber yang sangat potensial untuk analisis sentimen opini publik [3]. Sejalan dengan pandangan bahwa media sosial menjadi platform luas untuk penyebaran informasi [4]. Namun, fenomena ini menimbulkan berbagai respon dari masyarakat. Oleh karena itu, analisis sentimen dilakukan untuk mengidentifikasi atribut pada teks yang memuat komentar, sehingga dapat memahami respon yang ada dan dapat di klasifikasikan sebagai respon positif atau negatif [5]. Pemilihan metode yang sesuai juga menjadi faktor perlu untuk memperoleh hasil yang valid. Dua pendekatan yang sering dipakai pada analisis sentimen ialah *Naive bayes* dan *Support Vector Machine (SVM)* [6]. Dengan menggunakan kedua metode ini, kinerja dapat dibandingkan secara langsung, sehingga dapat dievaluasi kelebihan dan kekurangan dari masing-masing pendekatan [7]. *Support Vector Machine (SVM)* ialah algoritma dalam pengajaran mesin yang menggunakan konsep pemisahan data dengan *hyperplane* dan mampu memberikan akurasi yang lebih akurat pada data yang memiliki kompleksitas tinggi, sedangkan *Naive bayes* adalah algoritma berbasis probabilitas, yang dikenal karena kesederhanaan dalam mengklasifikasikan teks, meskipun asumsi independensi yang digunakan terkadang menjadi keterbatasan [8][9].

Berdasarkan paparan latar belakang, penelitian ini bertujuan untuk menjawab pertanyaan utama, seperti apakah algoritma *Support vector machine (SVM)* memiliki kinerja yang lebih baik dari pada algoritma *Naive bayes* terhadap menganalisis sentimen pengaruh korean wave di Indonesia. Penelitian ini juga diharapkan dapat mengidentifikasi model dengan kinerja terbaik serta memberikan pemahaman lebih mendalam tentang efektivitas metode analisis sentimen [10]. Selain itu, hasil penelitian ini dapat memberikan implikasi sosial dan kebudayaan yang penting, seperti korean wave tidak hanya mempengaruhi perilaku konsumsi masyarakat, tetapi juga memperluas wawasan budaya. Namun, kekhawatiran terkait pengaruh budaya asing terhadap identitas budaya lokal juga muncul. Dengan memahami sentimen masyarakat, diharapkan dapat mendukung pengambilan keputusan yang lebih tepat dan memperluas wawasan tentang pengaruh korean wave di Indonesia.

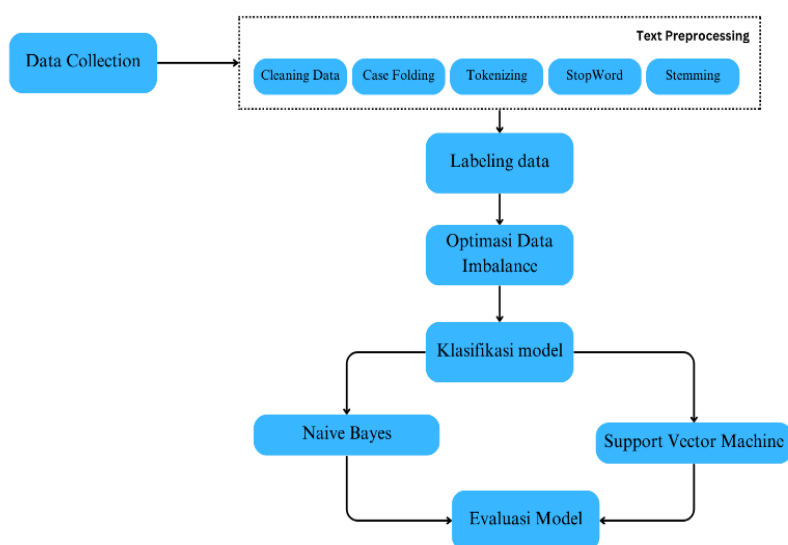
II. SIGNIFIKASI STUDI

A. Studi Literatur

Penelitian tentang analisis sentimen telah banyak dilakukan dalam berbagai konteks, terutama untuk membandingkan kinerja algoritma *Naive bayes* dan *Support Vector Machine*. Adapun analisis ini bertujuan untuk mengisi kesenjangan pengetahuan terkait pengaruh korean wave di Indonesia, dengan fokus pada sentimen masyarakat yang diungkapkan melalui media sosial dan perbandingan dua algoritma yang digunakan. Seperti pada penelitian yang menganalisis sentimen terhadap boy band BTS dengan dataset 1.648 data membuktikan bahwa algoritma SVM memiliki akurasi lebih tinggi 81% jika di bandingkan *Naive bayes* 79% [11]. Penelitian serupa juga dilakukan pada aplikasi satu sehat yang menunjukkan bahwa SVM mempunyai akurasi yang tinggi, sebesar 87,95%, sedangkan *Naive bayes* mencapai 81,65% [12]. Penelitian ini memberikan kontribusi yang unik dengan mengkaji fenomena korean wave, sebuah topik budaya populer yang memiliki pengaruh luas di Indonesia, menggunakan algoritma *Naive bayes* dan *Support Vector Machine (SVM)*, serta optimasi data melalui teknik SMOTE, penelitian ini menawarkan pendekatan metodologis yang komprehensif. Selain itu penelitian memperkaya literatur akademik di bidang analisis sentimen dan pemrosesan bahasa alami, serta memberikan wawasan dalam memahami dampak budaya populer, khususnya korean wave yang mempengaruhi pola konsumsi masyarakat, seperti lonjakan penjualan produk-produk kecantikan korea di Indonesia. Hasil penelitian dapat memberikan pengetahuan penting bagi media dan pemasar dalam menyusun strategi yang berbasis sentimen masyarakat.

B. Metodologi Penelitian

Penelitian ini menerapkan pendekatan metodologi yang sistematis untuk menganalisis sentimen masyarakat terhadap pengaruh korean wave di Indonesia. Seperti pada gambar 1, proses penelitian mencakup beberapa tahapan utama, seperti pengumpulan data (*Data Collection*), pengolahan data awal (*Preprocessing data*), pemberian label pada data (*Labeling data*), optimasi ketidakseimbangan data (*Optimasi data imbalance*) serta penerapan algoritma dan evaluasi hasil. Setiap tahapan disusun untuk memastikan analisis yang akurat dan relevan dengan tujuan penelitian.



Gambar 1. Tahapan penelitian

1. *Data Collection*

Data Collection adalah tahap pengambilan data yang dilaksanakan untuk mendapatkan informasi yang relevan dengan penelitian [13]. Pada penelitian ini data didapatkan dari media sosial X (dahulu Twitter) dengan memakai teknik *crawling* yang memanfaatkan token otentikasi aplikasi X untuk mengakses *Application Programming Interface (API)*. Kata kunci yang digunakan terkait dengan Korean Wave, dan fokus pada pengumpulan data dengan tweet berbahasa Indonesia. Prosesnya dilakukan dengan menggunakan *Library Tweet Harvest*, yang akan secara otomatis mengumpulkan tweet, sehingga memperlancar analisis data. Data yang berhasil dikumpulkan sebanyak 6.237 tweet, yang kemudian disimpan dalam format CSV untuk keperluan analisis selanjutnya.

2. Pra-pemrosesan

Pra-pemrosesan merupakan langkah penting dalam mempersiapkan data mentah agar dapat digunakan secara optimal dalam analisis. Tahapan ini bertujuan untuk membersihkan data serta menyusun data menjadi lebih terstruktur dan relevan untuk proses analisis selanjutnya [14].

a) *Cleaning data*

ialah tahap bertujuan untuk menghapus karakter-karakter yang tidak sesuai dalam analisis. Seperti simbol, angka, tanda baca, URL *hashtag*, dan nama pengguna [15]. Proses ini dapat memastikan data lebih bersih dan fokus pada teks utama untuk analisis yang lebih akurat.

b) *Case Folding*

adalah tahap dimana perubahan pada seluruh teks menjadi huruf kecil atau bisa disebut *lowercase*, untuk memastikan konsistensi dalam data yang di analisis [16]. Langkah ini bertujuan untuk menghilangkan perbedaan antara huruf sehingga dapat mengurangi keragaman data yang tidak relevan pada analisis.

c) *Tokenizing data*

merupakan tahapan pemecahan teks menjadi unit-unit kecil atau biasa disebut juga *token*, umumnya berupa kata atau frasa [17]. Tahapan ini dapat mempermudah analisis teks dengan memproses setiap kata secara individual untuk mengidentifikasi pola konteks dan informasi yang spesifik dari data yang di analisis.

d) *Stopword*

merupakan proses yang bertujuan untuk menghapuskan kata-kata umum yang sering muncul pada data, namun tidak memiliki kontribusi yang berarti untuk analisis, berupa kata “yang”, “atau”, “dan”, dengan kata sejenisnya [18].

e) *Stemming*

adalah tahapan yang bertujuan untuk menyederhanakan data menjadi bentuk dasar [19]. Adapun prosesnya dilakukan dengan menghilangkan imbuhan pada kata baik berupa awalan, akhiran, maupun sisipan, sehingga variasi kata yang berasal dari akar yang sama akan di kelompokkan menjadi satu bentuk.

3. Pelabelan data

Pelabelan data adalah proses menetapkan label atau kategori pada data yang telah di kumpulkan[20]. Proses ini bertujuan untuk membagi data ke dalam kategori spesifik, seperti positif serta negatif, yang berfungsi sebagai dasar dalam pelatihan model pembelajaran mesin. Dengan pelabelan yang akurat, model klasifikasi dapat mempelajari pola-pola sentimen secara lebih efektif, sehingga meningkatkan performa pada data yang akan di uji.

4. *Feature Extraction (TF-IDF)*

Tahapan yang dilakukan untuk mengkonversikan teks menjadi representasi numerik dengan memberikan nilai dalam setiap kata ataupun fitur. *Term frequency (TF)* berfungsi untuk menjumlah frekuensi kemunculan sebuah kata dokumen tertentu, serta *Inverse Document Frequency (IDF)* memiliki tugas untuk menghitung berapa jarang kata tersebut muncul pada keseluruhan data. Kombinasi kedua komponen ini memungkinkan penyesuaian bobot kata berdasarkan pada frekuensi dalam data dan keberadaannya dalam seluruh data. Hasil perhitungan TF-IDF ini lah yang kemudian akan digunakan sebagai fitur utama dalam proses klasifikasi sentimen.

5. *Optimasi data imbalance*

Optimasi data imbalance adalah metode yang digunakan untuk menangani perbedaan jumlah data antara kelas mayoritas serta kelas minoritas pada sebuah *dataset*. Ketidakseimbangan ini muncul ketika salah satu kelas mempunyai jumlah data yang secara signifikan lebih besar atau lebih kecil dari pada kelas lainnya. Dalam penelitian ini, Teknik SMOTE (*Synthetic Minority Over-sampling Technique*) diterapkan untuk dapat mengatasi masalah tersebut, awalnya kelas sentimen negatif hanya memiliki 1849 data, sedangkan sentimen positif berjumlah 4.388 data. Disinilah SMOTE bekerja dengan menghasilkan sampel sintesis untuk kelas minoritas, yaitu sentimen negatif sehingga jumlah meningkat setara kelas mayoritas, yakni 4.388 data. Proses ini dapat memastikan *dataset* yang lebih seimbang, serta meningkatkan kinerja model klasifikasi dan menghasilkan prediksi yang lebih akurat tanpa bias akibat ketidakseimbangan kelas. Namun penggunaan teknik ini membawa tantangan seperti risiko menciptakan data sintesis yang tidak sepenuhnya mencerminkan pola sentimen asli atau kemungkinan overfitting pada model.

6. *Klasifikasi Model*

Klasifikasi model adalah pengelompokan data berdasarkan kelas atau fitur tertentu. Penelitian ini memanfaatkan dua algoritma klasifikasi yakni *Naive bayes* serta *Support Vector Machine (SVM)*. *Naive bayes* ialah algoritma klasifikasi berbasis pendekatan probabilitas menggunakan teorema bayes yang di asumsi bahwa fitur-fitur dalam *dataset* independen. Sedangkan *Support Vector Machine (SVM)* adalah algoritma pembelajaran mesin yang berfungsi dengan menemukan *hyperplane* optimal untuk memisahkan data ke dalam berbagai kelas. Dengan memanfaatkan *kernel*, algoritma ini mampu menangani data yang tidak terpisah secara linear, sehingga cocok untuk melakukan klasifikasi pada *dataset* yang kompleks. Kedua algoritma tersebut dipilih karena keunggulan masing-masing dalam mengelola data teks untuk analisis sentimen. Adapun tujuan dari proses klasifikasi ini adalah untuk mengukur tingkat akurasi kedua model yang kemudian dibandingkan untuk menilai relatif dari masing-masing algoritma.

7. *Evaluasi Model*

Evaluasi model dilaksanakan untuk menghitung performa model klasifikasi yang di gunakan, yaitu *Naive Bayes* serta *Support Vector Machine(SVM)*, pada mengklasifikasi sentimen masyarakat terhadap pengaruh korean wave di Indonesia. Pengukuran performa model ini melibatkan penerapan *confusion matrix* untuk mendapatkan nilai seperti akurasi, presisi, recall dan *F1-score*.

$$Akurasi \frac{TP+TN}{FP+TN+FP+FN} \times 100 \tag{1}$$

$$Presisi \frac{TP}{TP+FP} \tag{2}$$

$$Recall \frac{TP}{TP+FN} \tag{3}$$

$$F1 - Score \frac{2 \times Presisi \times Recall}{Presisi + Recall} \tag{4}$$

III. HASIL DAN PEMBAHASAN

A. Dataset

Dalam penelitian ini, *dataset* didapat dari media sosial X melalui proses *crawling* dengan *Library Tweet Harvest*, dengan kata kunci seputar Korean wave. Proses ini berhasil mengumpulkan data sebanyak 6.237 data tweet, yang kemudian data tersebut akan di simpan dalam format CSV untuk memudahkan tahapan analisis selanjutnya. Proses pengambilan data ini menggunakan google colab dengan memanfaatkan kode autentikasi API dari media sosial X, serta didukung oleh bahasa pemrograman python. Dapat di lihat pada table 1. Hasil dari *dataset* yang telah dikumpulkan.

TABEL 1.
HASIL PENGUMPULAN DATA

No.	Tweet
1	Pengaruh Korean Wave di Indonesia makin besar! tapi sayangnya tidak semua pengaruhnya positif Fanwar toxic fandom hingga lupa budaya lokal jadi masalah!!~!
2	Kafe-kafe berkonsep Korea semakin menjamur di kota-kota besar. Pengaruh Korean Wave di Indonesia juga terasa di dunia kuliner. #koreanwave https://t.co/fiMTDhiZ

B. Pra-pemrosesan data

Data yang di analisis pada penelitian ini berupa teks tidak terstruktur, sehingga di perlukan langkah-langkah pra-pemrosesan untuk mempersiapkan data tersebut, tahapan ini meliputi *cleaning*, *case folding*, *tokenizing data*, *stopword* dan *stemming*, dengan tujuan untuk menghasilkan *dataset* yang bersih serta siap dipakai pada analisis serta meningkatkan kualitas data. Hasil pra-pemrosesan ini juga menunjukkan bahwa data yang telah di proses menjadi lebih terstruktur dan relevan untuk tahapan analisis selanjutnya, dapat di lihat pada table 2, hasil pra-pemrosesan.

TABEL 2.
HASIL PRA-PEMROSESAN

No.	Tahapan	Hasil Pra-pemrosesan
1	Cleaning Data	Pengaruh Korean Wave di Indonesia makin besar tapi sayangnya tidak semua pengaruhnya positif Fanwar toxic fandom hingga lupa budaya lokal jadi masalah
2	Case Folding	pengaruh korean wave di indonesia makin besar tapi sayangnya tidak semua pengaruhnya positif fanwar toxic fandom hingga lupa budaya lokal jadi masalah
3	Tokenizing Data	['pengaruh', 'korean', 'wave', 'di', 'indonesia', 'makin', 'besar', 'tapi', 'sayangnya', 'tidak', 'semua', 'pengaruhnya', 'positif', 'fanwar', 'toxic', 'fandom', 'hingga', 'lupa', 'budaya', 'lokal', 'jadi', 'masalah']

4	StopWord	['pengaruh', 'korean', 'wave', 'indonesia', 'makin', 'besar', 'sayangnya', 'semua', 'pengaruhnya', 'positif', 'fanwar', 'toxic', 'fandom', 'lupa', 'budaya', 'lokal', 'masalah']
5	Stemming	pengaruh korean wave indonesia makin besar sayang semua pengaruh positif fanwar toxic fandom lupa budaya lokal masalah

C. Pelabelan data

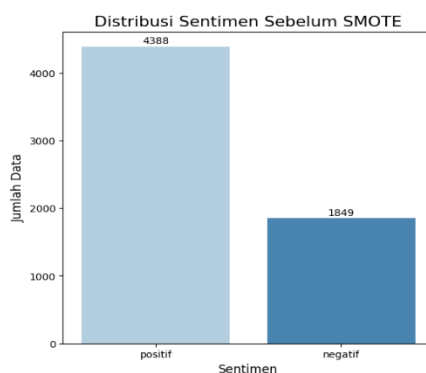
Data yang telah melalui pra-pemrosesan selanjutnya diberi label sentimen, yaitu positif serta negatif. Proses pemberian label ini dilakukan dengan manual atau *human annotation*, untuk memastikan akurasi dan relevansi label terhadap konteks tweet yang dianalisis. Dengan pelabelan manual ini, setiap data dapat diperiksa secara detail untuk menghindari kesalahan penafsiran yang mungkin terjadi pada pelabelan otomatis. Hasil pelabelan bisa dilihat dalam table 3

TABEL 3.
HASIL PELABELAN

No.	Tweet	Pelabelan Data
1	kafe kafe berkonsep korea semakin menjamur di kota kota besar. Pengaruh Korean Wave di Indonesia juga terasa di dunia kuliner.	Positif
2	pengaruh korean wave indonesia makin besar sayang semua pengaruh positif fanwar toxic fandom lupa budaya lokal masalah	Negatif

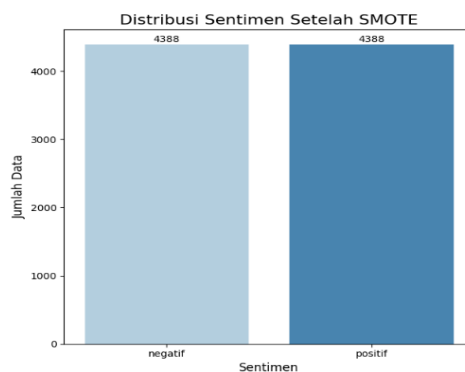
D. Optimasi SMOTE

Setelah dilakukan pelabelan, diperoleh distribusi sentimen dengan 4.388 tweet berlabel positif dan 1.849 tweet berlabel negatif, seperti yang terlihat pada gambar 1. Perbedaan yang signifikan antara jumlah data sentimen positif dan negatif ini menunjukkan adanya ketidakseimbangan data, yang dapat mempengaruhi kinerja dari model dalam mengklasifikasikan sentimen secara akurat. Untuk mengatasi permasalahan ini, dilakukan optimasi dengan menerapkan Teknik SMOTE yang bertujuan untuk menciptakan keseimbangan pada *dataset*.



Gambar 2. Jumlah klasifikasi sentimen sebelum Smote

Setelah penerapan Teknik SMOTE, jumlah data yang bersentimen positif serta negatif di dalam *dataset* berhasil diseimbangkan, masing-masing menjadi 4.388 data, seperti terlihat pada gambar 2. Proses penyeimbangan ini sangat penting untuk mengatasi permasalahan dari kelas data yang tidak seimbang, yang akan mempengaruhi performa model klasifikasi dalam analisis sentimen.



Gambar 3. Jumlah klasifikasi sentimen setelah Smote

D. Tahapan Pengujian

Sebelum masuk ke tahapan pengujian, *dataset* dibagi menjadi dua subset, yakni 80% untuk data latih serta 20% untuk data uji, guna memastikan validitas proses pelatihan dan pengujian model, setelah itu dilakukan penerapan dua algoritma klasifikasi, yakni *Naive Bayes* serta *Support Vector Machine(SVM)*. Kinerja kedua model tersebut akan di evaluasi dengan menggunakan metrik seperti *confusion matrix*, akurasi , presisi, recall serta *F1-score*. Seperti yang dilihat pada tabel 4, hasil pengujian yang menunjukkan perbedaan akurasi antara kedua model tersebut. Meskipun perbedaannya hanya sebesar 7%, SVM berhasil mencapai akurasi tertinggi sebesar 88%, sementara itu, model *Naive bayes* hanya sampai akurasi sebanyak 81%. Hal ini membuktikan bahwa model SVM memiliki performa yang lebih unggul di dibandingkan dengan model *Naive bayes* pada menganalisis sentimen pada *dataset* ini. Keunggulan tersebut juga mencerminkan kemampuan model SVM yang lebih baik dalam mengenali pola data, terutama dalam *dataset* yang telah dioptimalkan dengan dengan teknik SMOTE.

TABEL 4.
HASIL AKURASI

Model	Akurasi
Naive Bayes	81%
Support Vector Machine	88%

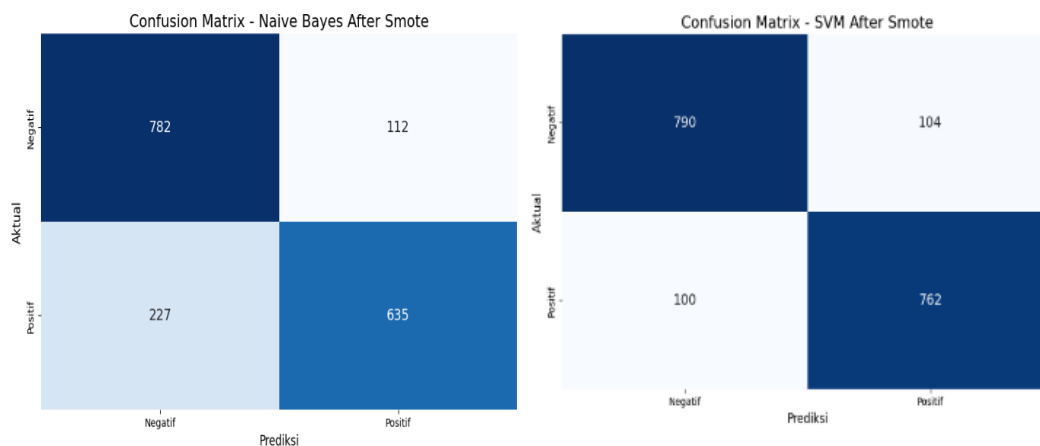
Berdasarkan table 5 hasil klasifikasi yang membandingkan model *Naive bayes* serta *Support Vector Machine(SVM)* setelah penerapan SMOTE, di ketahui juga bahwa model *Naive bayes* mencapai *recall* sebesar 74% untuk ulasan positif dan 87% untuk ulasan negatif. Model ini juga memiliki *precision* masing- masing sebesar 85% untuk ulasan positif dan 78% untuk ulasan negatif, serta *F1-Score* sebesar 79% untuk ulasan positif dan 82% untuk ulasan negatif. Disisi lain, Model SVM menunjukkan performa yang lebih unggul, dengan *recall* konsisten 88% untuk kedua kategori ulasan, *precision* sebesar 88% untuk ulasan positif dan 89% untuk ulasan negatif, serta *F1-score* sebesar 88% ulasan positif dan 89% untuk ulasan negatif.

TABEL 5.
HASIL KLASIFIKASI

Model	Sentimen	Precision	Recall	F1-Score
Naive bayes	Positif	85%	74%	79%
	Negatif	78%	87%	82%
SVM	Positif	88%	88%	88%
	Negatif	89%	88%	89%

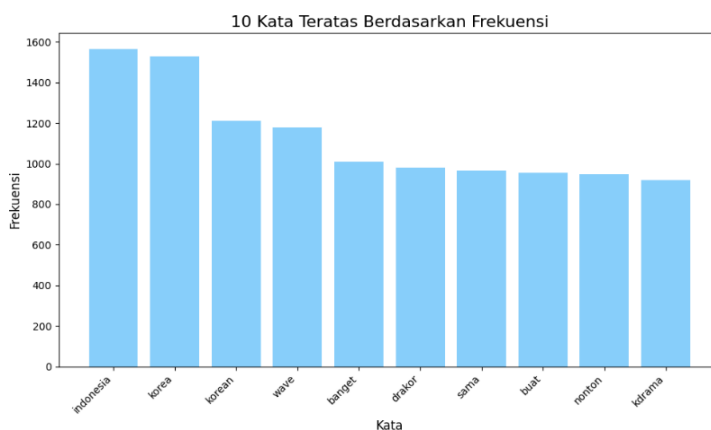
Perbandingan hasil evaluasi model klasifikasi *Naive bayes* dan *Support Vector Machine (SVM)* dengan *confusion matrix* yang dapat dilihat dalam gambar 4. Model *Naive bayes* menghasilkan

matriks dengan nilai *True Negative (TN)* sejumlah 782, *False Positive (FP)* sejumlah 112, *False Negative(FN)* sejumlah 227, dan *True Positive (TP)* sebanyak 635. Hal ini mengindikasikan bahwa model *Naive bayes* memiliki tingkat kesalahan yang cukup tinggi, terutama dalam memprediksi data positif, dengan FN mencapai 227. Sebaliknya Model SVM menghasilkan matrik dengan nilai *True Negative (TN)* sebanyak 790, *False Positive (FP)* sebanyak 104, *False Negative (FN)* sebanyak 100, dan *True positive (TP)* sebanyak 762. Nilai FP dan FN yang lebih rendah pada model SVM menunjukkan bahwa model ini mampu memprediksi kedua kelas(positif dan negatif) dengan tingkatan akurasi yang lebih baik di bandingkan *Naive bayes*.



Gambar 4. (a)confusion matrix naive bayes (b) confusion matrix svm

Tahapan visualisasi



Gambar 5 frekuensi kata tertinggi

Pada tahapan Visualisasi data, penelitian menggunakan diagram batang untuk menampilkan frekuensi kata-kata tertinggi dalam topik analisis. Sepuluh kata dengan frekuensi tertinggi dalam topik pengaruh Korean wave di Indonesia memberikan wawasan yang mendalam tentang fokus dan tema utama yang sering di bahas dalam tweet, seperti yang terlihat pada gambar 5. Kata “Indonesia” mencerminkan konteks tempat serta bagaimana Korean wave diterima dan mempengaruhi masyarakat lokal. Kata “korea” dan “Korean “ menunjukkan asal budaya yang menjadi inti dari perhatian, sementara “wave” mengacu pada fenomena Korean wave itu sendiri. Adapun istilah seperti “drakor” dan “Kdrama” merujuk pada drama korea yang menjadi salah satu produk budaya yang paling populer. Kata “nonton” dan “sama” mencerminkan aktivitas yang dilakukan terkait konsumsi konten, sedangkan “banget” dan “sama” menunjukkan ekspresi emosional atau

REFERENSI

- [1] S. R. Jannah, Z. Khoirunnisa, and A. R. Faristiana, “Pengaruh Korean Wave Dalam Fashion Style Remaja Indonesia,” *J. Ilm. Pendidik. Kebud. Dan Agama*, vol. 1, no. 3, pp. 11–20, 2023, doi: 10.59024/jipa.v1i3.219.
- [2] Asiva Noor Rachmayani, “PEMENUHAN HAK ASASI MANUSIA KEPADA ETNIS ROHINGYA DI MAYANMAR,” 2021.
- [3] 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 Ris. Komputer)*, vol. 9, no. 2, p. 338, 2022, doi: 10.30865/jurikom.v9i2.3895.
- [4] U. P. Mutiarani, I. N. Karimah, and Y. P. Syarafa, “Etika Komunikasi dalam Penggunaan Media Sosial di Kalangan Mahasiswa,” *J. Harmon. Nusa Bangsa*, vol. 1, no. 2, 2024, doi: 10.47256/jhnb.v1i2.301.
- [5] H. Syah and A. Witanti, “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm),” *J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 59–67, 2022, doi: 10.47080/simika.v5i1.1411.
- [6] D. Ananda and R. R. Suryono, “Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 748, 2024, doi: 10.30865/mib.v8i2.7517.
- [7] C. F. Hasri and D. Alita, “Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, doi: 10.33365/jatika.v3i2.2026.
- [8] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [9] A. F. Setyaningsih, D. Septiyani, and S. R. Widiyari, “Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia,” *J. Teknol. Inform. dan Komput.*, vol. 9, no. 1, pp. 224–235, 2023, doi: 10.37012/jtik.v9i1.1409.
- [10] “Eskiyaturrofikoh” and R. R. ’Suryono, “Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (Svm),” *JUPI(Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 3, pp. 1408–1419, 2024, [Online]. Available: <https://www.jurnal.stkipgritulungagung.ac.id/index.php/jipi/article/view/5392>
- [11] Rina Noviana and Isram Rasal, “Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter,” *J. Tek. dan Sci.*, vol. 2, no. 2, pp. 51–60, 2023, doi: 10.56127/jts.v2i2.791.
- [12] F. Matheos Sarimole and K. Kudrat, “Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine,” *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 783–790, 2024, doi: 10.55338/saintek.v5i3.2702.
- [13] A. Supian, B. Tri Revaldo, N. Marhadi, L. Efrizoni, and R. Rahmaddeni, “Perbandingan Kinerja Naïve Bayes Dan Svm Pada Analisis Sentimen Twitter Ibukota Nusantara,” *J. Ilm. Inform.*, vol. 12, no. 01, pp. 15–21, 2024, doi: 10.33884/jif.v12i01.8721.
- [14] K. Kevin, M. Enjeli, and A. Wijaya, “Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes,” *J. Ilm. Comput. Sci.*, vol. 2, no. 2, pp. 89–98, 2024, doi: 10.58602/jics.v2i2.24.
- [15] D. Oktavia, Y. R. Ramadhan, and Minarto, “Analisis Sentimen Terhadap Penerapan Sistem E-Tilang

- Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.
- [16] A. Setiawan and R. R. Suryono, “Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 183–192, 2024, doi: 10.29408/edumatic.v8i1.25667.
- [17] W. Athira Luqyana, I. Cholissodin, and R. S. Perdana, “Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4704–4713, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [18] M. Taufik Sugandi, Martanto, and U. Hayati, “Analisis Sentimen Komentar Pengguna Youtube terhadap Kebijakan Baru Badan Penyelenggara Jaminan Kesehatan Sosial Menggunakan Naïve Bayes,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 218–227, 2024.
- [19] N. V. R. Jhosefhin and C. Dewi, “Analisis Sentimen Crawling Data dari Sosial Media X tentang Gaza Menggunakan Metode SVM dan Decision Tree,” *J. Indones. Manaj. Inform. DAN Komun.*, vol. 6, no. 1, pp. 427–437, 2025.
- [20] A. Ilham, “Analisis Sentimen Masyarakat Terhadap Kesehatan Mental Pada Twitter Menggunakan Algoritme K-Nearest Neighbor,” *Pros. Semin. Nas. Mhs. Fak. ...*, vol. 2, no. September, pp. 539–547, 2023, [Online]. Available: <http://senafti.budiluhur.ac.id/index.php/senafti/article/view/792%0A>
<http://senafti.budiluhur.ac.id/index.php/senafti/article/download/792/527>.