JNATIA Volume 3, Nomor 4, Agustus 2025 p-ISSN: 2986-3929
Jurnal Nasional Teknologi Informasi dan Aplikasinya e-ISSN: 3032-1948

Analisis Sentimen Fenomena FoMO Gaya Hidup Sehat Menggunakan Metode *Naïve Bayes Classifier*

Putu Mahdalika Intan Pratiwi[a1], Anak Agung Istri Ngurah Eka Karyawati[a2]

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]pratiwi.2308561028@student.unud.ac.id
[2]eka.karyawati@unud.ac.id

Abstract

The development of social media has given rise to the Fear of Missing Out (FoMO) phenomenon which can affect healthy lifestyles among the community, especially the younger generation. This study aims to conduct a sentiment analysis of the FoMO phenomenon related to a healthy lifestyle using the Naïve Bayes Classifier method. Data were collected from Twitter social media using a crawling technique with certain keywords in the period from January 2024 to June 2025. The data was then processed through the preprocessing stage, sentiment labeling, data balancing using SMOTE, and feature weighting using TF-IDF. The model was trained using 2,364 training data and tested on 591 data. The evaluation results showed that the model achieved an accuracy of 84.09%, precision 85.66%, recall 84.09%, and F1-score 83.57%. Validation using 5-Fold Cross Validation also showed stable model performance with an average accuracy of 84.50%. This study proves that Naïve Bayes is effective in analyzing social media sentiment towards the FoMO phenomenon, with potential for development through dataset expansion and exploration of other algorithms.

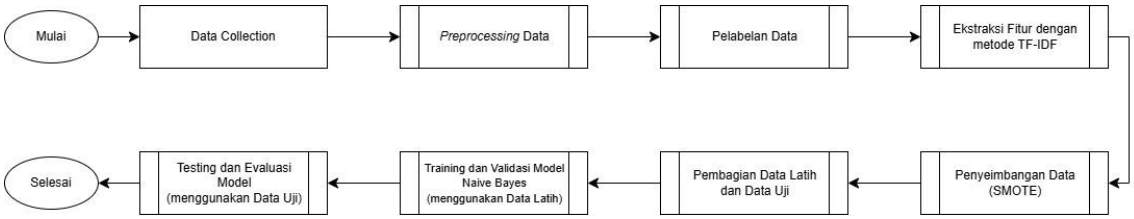
Keywords: *Sentiment Analysis, Fear of Missing Out, Twitter, Naïve Bayes Classifier, TF-IDF, SMOTE*

1. Pendahuluan

Pada era digital yang terus berkembang saat ini, media sosial seperti Twitter atau X menjadi salah satu *platform* yang memegang peran penting dalam penyebaran informasi gaya hidup sehat. Ini memungkinkan bagi individu maupun kelompok memposting konten seperti teks, gambar, dan video, serta berpartisipasi dalam percakapan global melalui media sosial. Dengan kemajuan teknologi, generasi muda lebih aktif dalam mencari informasi kesehatan mereka. Di mana menurut jurnal 'Digital Health', disebutkan 75% Gen Z melaporkan menggunakan *gadget* mereka untuk mengakses informasi tentang kesehatan dan kebugaran [1]. Media sosial ini dapat menjadi motivasi bagi pola hidup sehat, tetapi dari tren Kesehatan ini juga menimbulkan tekanan sosial [1]. Salah satu konsekuensi dari media sosial adalah munculnya fenomena *Fear of Missing Out* (FoMo), yaitu ketakutan ketinggalan informasi atau kegiatan yang sedang tren [2]. Misalnya, tren lari di kalangan Gen Z disebut-sebut sebagai bentuk FoMo terkait aktivitas olahraga yang viral [3]. Dalam konteks ini, penelitian sebelumnya mengenai fenomena FoMO menunjukkan manfaat analisis sentiment untuk memahami reaksi masyarakat. Seperti, penelitian yang diteliti oleh Henni dan Samidi (2025) menggunakan *Naïve Bayes* untuk mengklasifikasikan tingkat FoMO berdasarkan data penggunaan media sosial di Indonesia [4]. Penelitian serupa lainnya pada tahun 2023 yang dilakukan oleh Setyaningsih dkk dalam analisis sentiment *tweet* masyarakat mengenai kepopuleran produk *skincare* [5]. Penelitian dilakukan dengan menggunakan data Twitter, dengan metode *Naïve Bayes* menghasilkan akurasi sebesar 86% [5]. Berdasarkan uraian permasalahan dan referensi penelitian sebelumnya, penulis akan menganalisis sentiment menggunakan metode *Naïve Bayes* untuk analisis sentiment fenomena FoMO gaya hidup sehat berbasis data Twitter. Penelitian ini bertujuan untuk mengetahui sentiment masyarakat terhadap gaya hidup sehat, serta menguji apakah *Naïve Bayes* dapat digunakan untuk menganalisis sentiment tersebut.

2. Metode Penelitian

Pada bagian metodologi penelitian, penulis menerapkan serangkaian pendekatan atau metode untuk melakukan analisis sentimen. Pada penelitan ini, langkah-langkah yang diambil oleh penulis meliputi tahapan pencarian dan pengumpulan data melalui *crawling*, *preprocessing* data teks, dilanjutkan dengan tahap pelabelan data, dan pembobotan fitur dengan metode *Term Frequency – Inverse Document Frequency* (TF-IDF), penyeimbangan data menggunakan SMOTE, pemisahan data menjadi data uji (*test*) dan data latih (*training*). Selanjutnya, dilakukan *train model Naïve Bayes*, dan validasi dengan *K-Fold Cross Validation*, terakhir *testing* dan evaluasi model dengan *confusion matriks* menghasilkan seperti *accuracy*, *recall, precission,* dan skor F1. Gambar 1 di bawah ini memberikan tahpan-tahapan yang dilakukan untuk penelitian ini:



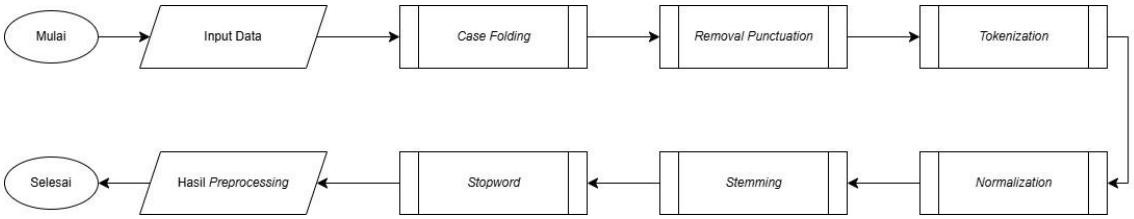
Gambar 1. Diagram Alir Penelitian

2.1. Pengumpulan Data

Pada penelitian ini, data yang digunakan adalah *tweet* masyarakat dari media sosial Twitter berbahasa Indonesia. Data dikumpulkan menggunakan metode *crawling* dengan *tool tweet-harvest* di Google Colab, serta diperlukan *auth_token* Twitter. Pencarian *tweet* dilakukan berdasarkan kata kunci “fomo,” “gaya hidup,” dan “hidup sehat,” untuk rentang waktu mulai bulan Januari 2024 hingga Juni 2025. Hasil *crawling* tersimpan dalam bentuk file CSV, dengan total data *tweet* yang diperoleh berjumlah 1703 *tweet*. Setiap *tweet* nantinya akan dilabeli otomatis dengan tiga kelas sentimen, yaitu “Negatif,” ‘Positif,” dan “Netral.”

2.2. *Preprocessing Text*

Preprocessing text adalah tahapan awal dalam pemrosesan bahasa alami (NLP). Agar proses dapat berlanjut dengan lancar ke tahap berikutnya, proses ini berupaya mengubah input teks asli yang tidak terstruktur menjadi data terstruktur dan menjamin formatnya [6]. Data *tweet* yang telah dikumpulkan, dibersihkan dan distandarisasi disiapkan untuk langkah berikutnya. Pada penelitian ini Gambar 2 di bawah ini menunjukkan langkah-langkah yang dilakukan dalam penelitian ini.



Gambar 2. Alur *Preprocessing*

Berikut ini adalah penjelasan terperinci mengenai proses *preprocessing* yang digunakan dalam penelitian ini.

a. *Case Folding*
   Case folding adalah proses mengubah setiap huruf dalam kumpulan data menjadi huruf kecil dikenal sebagai pelipatan huruf, sehingga keseluruhan teks memiliki bentuk standar [7].

b. *Removal Punctuation*
Pembersihan teks mencakup penghapusan karakter yang ada dalam teks berupa hastag (‘#’), URL, mention (‘@’), dan angka yang tidak diperlukan, sehingga mengurangi noise dan menghasilkan data cuitan yang asli [6].

c. *Tokenization*
Tokenisasi adalah memotong atau memecah teks menjadi unit-unit kata (token) pada dokumen yang bertujuan untuk dianalisis secara terpisah, ini memudahkan bagi computer untuk memproses daftar token dibandingkan teks panjang [7].

d. *Normalization*
Normalization adalah tahap dalam pemrosesan teks yang bertujuan untuk mengubah kata-kata non-baku, slang, atau singkatan menjadi bentuk baku yang sesuai dengan tata bahasa, seperti yang terdapat dalam Kamus Besar Bahasa Indonesia (KBBI) [6].

e. *Stemming*
Stemming adalah proses melibatkan penghapusan afiks dari kata yang telah dibersihkan untuk mengembalikannya ke bentuk dasarnya, memetakan semua variasi kata ke bentuk yang sama [7].

f. *Stopword*
Stopword removal adalah tahapan untuk menghilangkan kata-kata yang tidak penting atau sangat umum tetapi cenderung kurang membawa makna spesifik. Ini bertujuan untuk mengurangi dimensi data dengan menyingkirkan token yang sering muncul [7].

2.3. Pelabelan Data

Penelitian ini, untuk setiap *tweet* diberi label sentiment secara otomatis. Proses pelabelan dilakukan dengan menerjemahkan teks *tweet* ke Bahasa Inggirs dengan library *translate* dan menghitung skor sentiment dengan library *Textblob.* Skor akan digunakan sebagai indikator pelabelan sentiment (‘negatif, ‘positif, ‘netral’). Pada penelitian ini, analisis sentiment tetap menggunakan bahasa Indonesia, walaupun penetuan skor awal menggunakan alat Bahasa Inggris.

2.4. Penyeimbangan Data

Pada penelitian ini, dilakukan *oversampling* menggunakan pendekatan SMOTE, agar distribusi kelas menjadi lebih seimbang. SMOTE (*Synthetic Minority Oversampling Technique)* adalah teknik mengatasi masalah ketidakseimbangan data, dengan cara membuat contoh baru secara sintetis di ruang fitur [8]. SMOTE banyak digunakan dalam penelitian karena terbukti mampu memperbaiki keseimbangan data secara otomatis dalam banyak skenario.

2.5. *Term Frequency – Inverse Document Frequency*

Salah satu pendekatan pembobotan paling umum digunakan dalam penelitian analisis teks adalah TF-IDF. TF-IDF (*Term Frequency – Inverse Document Frequency*) adalah suatu teknik untuk menimbang frekuensi istilah dalam dokumen teks. Salah satu tujuan pembobotan TF-IDF adalah mengubah data dari data tekstual menjadi data numerik sehingga pembobotan dapat dilakukan pada setiap kata atau fitur [9]. Persamaan yang digunakan dalam TF-IDF dapat dilihat pada persamaan 1, 2 dan 3 sebagai berikut.

$$tf(t,d) = \frac{f_{(t,d)}}{n} \quad (1)$$

$$idf(\mathrm{t}) = log\left(\frac{n}{1+df(t)}\right) \quad (2)$$

$$w_{t,d} = tf_{t,d}\ x\ idf_t \quad (3)$$

2.6. *Naïve Baiyes Classifier*

Algoritma *Naïve Bayes Classifier* (NBC) adalah model klasifikasi probabilistik yang menggunakan Teorema Bayes. *Naive Bayes Classifier* sendiri merupakan metode pembelajaran probabilistik di mana setiap kata memiliki probabilitas kemunculan yang independen. Ini berarti bahwa nilai atribut kategori tidak saling mempengaruhi atau dipengaruhi oleh nilai atribut lainnya. *Naïve Bayes* memilih kelas dokumen berdasarkan probabilitas tertinggi [10]. Sebagaimana ditunjukkan dalam permasamaan 4 berikut.

$$P(\mathrm{c}) = \frac{Nc}{N_{doc}} \tag{4}$$

Persamaan (5) menyatakan bahwa N_c sebagai jumlah kelas C untuk keseluruhan dokumen dan N sebagai jumlah total dokumen [10]. Persamaan 5 digunakan untuk mendapatkan rumus probabilitas untuk kata ke-n.

$$P(X_n \mid C) = \frac{N_{x_n,c} + \alpha}{N(C) + V} \tag{5}$$

Dalam kasus ini, N_c adalah jumlah total *term* dalam semua data latih untuk kelas C dan $N_{x_n,c}$ adalah jumlah kemunculan *term* X_n dalam semua data latih. Kata-kata yang tidak ada dalam data latih ditangani menggunakan parameter pemulusan *Laplace* α. Lebih lanjut, jumlah total kata dalam data latih dilambangkan dengan V [10]. Berikut ini adalah rumus *Multinomial Naïve Bayes* yang digunakan dalam pembobotan TF-IDF. Dengan $\sum tf(X_n, d \in C)$ menunjukkan total bobot TF (frekuensi istilah) Xn pada semua dokumen kelas C. Serta, $\sum N_{d \in C}$ menunjukkan jumlah bobot dari seluruh *term* dalam kelas C.

$$P(X_n \mid C) = \frac{\sum tf(X_n, d \in C) + \alpha}{\sum N_{d \in C} + V} \tag{6}$$

2.7. *K-Fold Cross Validation*

K-Fold Cross Validation adalah metode validasi yang digunakan untuk mendapatkan nilai akurasi optimal yang memisahkan data menjadi data uji dan data latih [11]. *K-Fold Cross Validation*, di mana dataset dibagi menjadi K partisi acak, adalah salah satu jenis yang paling popular. Satu partisi digunakan sebagai data uji dan yang lainnya sebagai data latih untuk prosedur evaluasi, yang dijalankan sebanyak K kali [11]. Hal ini berkaitan dengan penggunaan dataset lengkap untuk evaluasi dan memberikan estimasi akurasi model yang lebih akurat.

2.8. Skenario Pengujian dan Evaluasi

Evaluasi performa model pada penelitian ini, dilakukan menggunakan *confusion matriks. Confusion matriks* adalah metode yang memuat informasi mengenai jumlah prediksi yang benar dan salah yang dilakukan oleh model untuk setiap kelas dalam data. Nilai yang dihitung meliputi *accuracy, precision, recall,* dan F1-*score* [12]. Evaluasi dilakukan setelah tahap pengujian menggunakan data testing yang telah dibagi sebelumnya. Untuk menghitung performa model melalui *accuracy, precision, recall,* dan F1-*score,* menggunakan rumus-rumus pada persamaan 7 hingga 10 berikut.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

$$\mathrm{F1-Score} = 2\ \mathrm{X}\ \frac{precision\ x\ recall}{precision + recall} \tag{10}$$

Keterangan:

- *True Positive* (TP) = jumlah data yang positif yang terprediksi benar (positif).
- *True Negative* (TN) = jumlah data yang negatif yang terprediksi benar (negatif).
- *False Positive* (FP) = jumlah data yang positif namun terprediksi salah (negatif).
- *False Negative* (FN) = jumlah data yang berlabel negatif tetapi terprediksi salah (positif).

3. Hasil dan Diskusi

Bagian ini memuat hasil dan diskusi penelitian yang diperiksa menggunakan *confusion matriks*. Hasil evaluasi mencakup nilai F1-Score, *recall*, akurasi, dan presisi. Angka-angka ini menunjukkan seberapa efektif dan tepat prosedur kategorisasi sentimen tersebut.

3.1 Pengumpulan Data

Data penelitian harus diolah terlebih dahulu karena masih mentah dan mengandung karakter asing serta singkatan. Data tersebut berasal dari istilah "fomo," "gaya hidup," dan "hidup sehat," yang menghasilkan 1703 tweet. Data mentah tersebut memiliki 15 karakteristik, termasuk *conversation_id_str, created_at, favorite_count, full_text, id_str, image_url, in_reply_to_screen_name, lang, location, quote_count, reply_count, retweet_count, tweet_url, user_id_str,* dan *username.* Berikut contoh dataset yang sudah dibersihkan.

Tabel 1. Dataset

stemming	**sentimen**
seproper si lengkap timsar nya teknologi kalah cuaca buruk dn kondisi alam emng daki mula dn fomo hiking takut mati larang hiking gunung ky rinjani	negatif
ngomongin much love staff fomo bilang hai staffs tau stalk sayang banget makasih ya hadir hidup kalo memutarbalikan would choose every timeline	positif
portofolio sehat niat ajar fomo	positif
hidup fomo	netral

Selanjutnya, data yang berlebihan dan kualitas yang tidak diperlukan dihilangkan, sehingga tersisa 1.396 data yang selanjutnya diklasifikasikan sebagai "positif," "negatif," dan "netral" dan dilakukan penyeimbangan data. Berikut pada Gambar 3 ditampilkan pemetaan dari skor sentiment dengan label sentiment. Di mana terdapat dua nilai utama, yaitu nilai subjektivitas dan nilai polaritas. Nilai polaritas menunjukkan kecenderungan sentimen teks dalam rentang -1,0 hingga +1,0, di mana -1,0 menunjukkan emosi yang sangat negatif, 0 netral, dan +1,0 menunjukkan sentimen yang sangat positif. Sedangkan, nilai subjektivitas menunjukkan tingkat di mana suatu teks merupakan opini atau fakta. Pemetaan skor sentimen menjadi tiga label kategorisasi dilakukan berdasarkan nilai polaritas ini: "positif" untuk nilai polaritas > 0, "negatif" untuk nilai polaritas < 0, dan "netral" untuk nilai polaritas = 0.

translated_text	subjektivitas	polaritas	sentimen
	0.0	0.0	netral
ınology can t be a bad weather and the conditions of nature mng daki daki scared d mountain	0.6083333333333334	-0.17499999999999993	negatif
›ve staff fomo said that hai staffs know stalk really love thank you if you re coming for the would be choose every timeline	0.5333333333333333	0.3666666666666667	positif

Gambar 3. Hasil *Labeling*

Sebelum implementasi algoritma *Naïve Bayes*, dataset total sejumlah 1396 data dilakukan penyeimbangan menggunakan SMOTE, sehingga menghasilkan 2955 data. Selanjutnya, pembagian data menjadi 80% data latih dan 20% data uji. Ini mengikuti prinsip *Pareto* yang banyak digunakan dalam pemodelan data, dengan tujuan menyediakan cukup banyak data untuk melatih model serta menyisakan data untuk menguji kinerja [13]. Pembagian data dilakukan

p-ISSN: 2986-3929
e-ISSN: 3032-1948

secara acak agar data uji benar-benar menggambarkan keadaan data baru yang belum pernah dilihat oleh model. Data latih terdiri dari 2364 data terbagi merata ke dalam tiga kelas, yaitu negatif, positif, dan netral, masing-masing sebanyak 788 data. Sedangkan, data uji terdiri dari 591 data, dengan distribusi masing-masing kelas, yaitu negatif, positif, dan netral sebanyak 197 data.

3.2 *Preprocessing* Data

Setelah tahapan *crawling*, selanjutnya masuk ketahapan *preprocessing*. Pada tahapan ini, dilakukan pembersihan data agar konsisten dan bisa digunakan untuk *train* model. Berikut ini merupakan hasil *preprocessing* yang meliputi *tweet, lowercase, cleanpunct, tokens, normalization, stemming,* dan *stopword.*

tweet	lowercase	cleanpunct
@sapphicsfess dia yg lesbi fomo banyak gaya	@sapphicsfess dia yg lesbi fomo banyak gaya	dia yg lesbi fomo banyak gaya
@Dillakodildap2 @NFakhroni @iwontmove Mau se-proper apalagi si perlengkapan timsar nya? Teknologi ttp bakal kalah sm cuaca buruk dn kondisi alam emng lebih baik pendaki yg pemula dn fomo hiking tp takut mati dilarang hiking di gunung ky rinjani	@dillakodildap2 @nfakhroni @iwontmove mau se-proper apalagi si perlengkapan timsar nya? teknologi ttp bakal kalah sm cuaca buruk dn kondisi alam emng lebih baik pendaki yg pemula dn fomo hiking tp takut mati dilarang hiking di gunung ky rinjani	mau seproper apalagi si perlengkapan timsar nya teknologi ttp bakal kalah sm cuaca buruk dn kondisi alam emng lebih baik pendaki yg pemula dn fomo hiking tp takut mati dilarang hiking di gunung ky rinjani
pada ngomongin how much they love their staff jadi gue fomo dan gue mau bilang hai staffs! gue tau beberapa dari kalian stalk gue gue sayang banget sama kalian makasih ya udah hadir di hidup gue. kalo gue bisa memutarbalikan waktu i would choose this again on every timeline	pada ngomongin how much they love their staff jadi gue fomo dan gue mau bilang hai staffs! gue tau beberapa dari kalian stalk gue gue sayang banget sama kalian makasih ya udah hadir di hidup gue. kalo gue bisa memutarbalikan waktu i would choose this again on every timeline	pada ngomongin how much they love their staff jadi gue fomo dan gue mau bilang hai staffs gue tau beberapa dari kalian stalk gue gue sayang banget sama kalian makasih ya udah hadir di hidup gue kalo gue bisa memutarbalikan waktu i would choose this again on every timeline

Gambar 4. Hasil *tweet, lowercase, cleanpunct*

tokens	normalization	stemming	stopword
['dia', 'yg', 'lesbi', 'fomo', 'banyak', 'gaya']	['dia', 'yang', 'lesbi', 'fomo', 'banyak', 'gaya']	dia yang lesbi fomo banyak gaya	['lesbi', 'fomo', 'gaya']
['mau', 'seproper', 'apalagi', 'si', 'perlengkapan', 'timsar', 'nya', 'teknologi', 'ttp', 'bakal', 'kalah', 'sm', 'cuaca', 'buruk', 'dn', 'kondisi', 'alam', 'emng', 'lebih', 'baik', 'pendaki', 'yg', 'pemula', 'dn', 'fomo', 'hiking', 'tp', 'takut', 'mati', 'dilarang', 'hiking', 'di', 'gunung', 'ky', 'rinjani']	['mau', 'seproper', 'apalagi', 'si', 'perlengkapan', 'timsar', 'nya', 'teknologi', 'tetap', 'bakal', 'kalah', 'sama', 'cuaca', 'buruk', 'dn', 'kondisi', 'alam', 'emng', 'lebih', 'baik', 'pendaki', 'yang', 'pemula', 'dn', 'fomo', 'hiking', 'tapi', 'takut', 'mati', 'dilarang', 'hiking', 'di', 'gunung', 'ky', 'rinjani']	mau seproper apalagi si lengkap timsar nya teknologi tetap bakal kalah sama cuaca buruk dn kondisi alam emng lebih baik daki yang mula dn fomo hiking tapi takut mati larang hiking di gunung ky rinjani	['seproper', 'si', 'lengkap', 'timsar', 'nya', 'teknologi', 'kalah', 'cuaca', 'buruk', 'dn', 'kondisi', 'alam', 'emng', 'daki', 'dn', 'fomo', 'hiking', 'takut', 'mati', 'larang', 'hiking', 'gunung', 'ky', 'rinjani']
['pada', 'ngomongin', 'how', 'much', 'they', 'love', 'their', 'staff', 'jadi', 'gue', 'fomo', 'dan', 'gue', 'mau', 'bilang', 'hai', 'staffs', 'gue', 'tau', 'beberapa', 'dari', 'kalian', 'stalk', 'gue', 'gue', 'sayang', 'banget', 'sama', 'kalian', 'makasih', 'ya', 'udah', 'hadir', 'di', 'hidup', 'gue', 'kalo', 'gue', 'bisa', 'memutarbalikan', 'waktu', 'i', 'would', 'choose', 'this', 'again', 'on', 'every', 'timeline']	['pada', 'ngomongin', 'how', 'much', 'they', 'love', 'their', 'staff', 'jadi', 'saya', 'fomo', 'dan', 'saya', 'mau', 'bilang', 'hai', 'staffs', 'saya', 'tau', 'beberapa', 'dari', 'kalian', 'stalk', 'saya', 'saya', 'sayang', 'banget', 'sama', 'kalian', 'makasih', 'ya', 'sudah', 'hadir', 'di', 'hidup', 'saya', 'kalo', 'saya', 'bisa', 'memutarbalikan', 'waktu', 'i', 'would', 'choose', 'this', 'again', 'on', 'every', 'timeline']	pada ngomongin how much they love their staff jadi saya fomo dan saya mau bilang hai staffs saya tau beberapa dari kalian stalk saya saya sayang banget sama kalian makasih ya sudah hadir di hidup saya kalo saya bisa memutarbalikan waktu i would choose this again on every timeline	['ngomongin', 'much', 'love', 'staff', 'fomo', 'bilang', 'hai', 'staffs', 'tau', 'stalk', 'sayang', 'banget', 'makasih', 'ya', 'hadir', 'hidup', 'kalo', 'memutarbalikan', 'would', 'choose', 'every', 'timeline']

Gambar 5. Hasil *tokens, normalization, stemming, stopword*

p-ISSN: 2986-3929
e-ISSN: 3032-1948

3.3 Ekstraksi Fitur

Pada tahapan ini, fitur diekstraksi fitur menggunakan TF-IDF. Proses ini dilakukan setelah data melalui tahapan *preprocessing* ini bertujuan untuk mengubah data teks menjadi numerik. Pembagoan juga dibedakan dari 2 hal, yaitu nilai X sebagai fitur ekstraksi TF-IDF yang diambil dari hasil *stemming* dan Y sebagai label sentiment. Sehingga didapat 20 kata yang memiliki skor bobot tertinggi sebagai berikut.

20 Kata dengan skor TF-IDF tertinggi:

	kata	skor_tfidf
1029	hidup	0.061475
1084	hidup sehat	0.057281
2858	takut	0.039135
1645	makan	0.031115
808	fomo	0.030968
2075	orang	0.024050
3218	ya	0.021835
261	banget	0.020863
2042	olahraga	0.018756
1815	moga	0.016219
2256	pola	0.016154
2257	pola hidup	0.013920
890	gaya	0.013835
1319	kalo	0.013654
226	bahagia	0.012596
1669	makan sehat	0.011478
892	gaya hidup	0.011414
2428	sakit	0.011306
397	biar	0.011306
118	anak	0.010071

Gambar 6. Hasil *TF-IDF*

3.4 Implementasi *Naïve Bayes Classifier*

Setelah tahap ektraksi fitur, dilakukan implementasi *Naïve Bayes* pada data latih. Performa model dipengaruhi oleh sejumlah *hyperparameter* utama selama implementasi, yaitu untuk menghindari probabilitas nol saat kata tertentu tidak muncul di kelas tertentu, *alpha* (perataan *Laplace*) digunakan. Model dibuat lebih stabil dengan efek *smoothing* dengan nilai *default* 1,0. Kemudian, *fit_prior* disetel ke *True*, maka model akan menggunakan distribusi data untuk menentukan prior probabilitas sebelumnya dari masing-masing kelas. Sebaliknya, jika *False* akan mengabaikan distribusi data sebenarnya dan menganggap bahwa setiap kelas memiliki prior yang sama. Serta, *class_prior* ini nilai probabilitas ditentukan secara manual, hanya digunakan saat *fit_prior* disetel *False*, parameter ini menjadi aktif. Pada tahap ini, model dilatih menggunakan data latih sebanyak 2364 data yang sudah seimbang untuk tiga jenis kelas sentiment. Proses pelatihan bertujuan untuk menghasilkan model klasifikasi berbasi probabilitas yang memanfaatkan teorema Bayes. Hasil *train* model menunjukkan akurasi sebesar 92.22%. Diikuti dengan hasil *precision, recall,* dan F1-*score* pada gambar 7 berikut.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

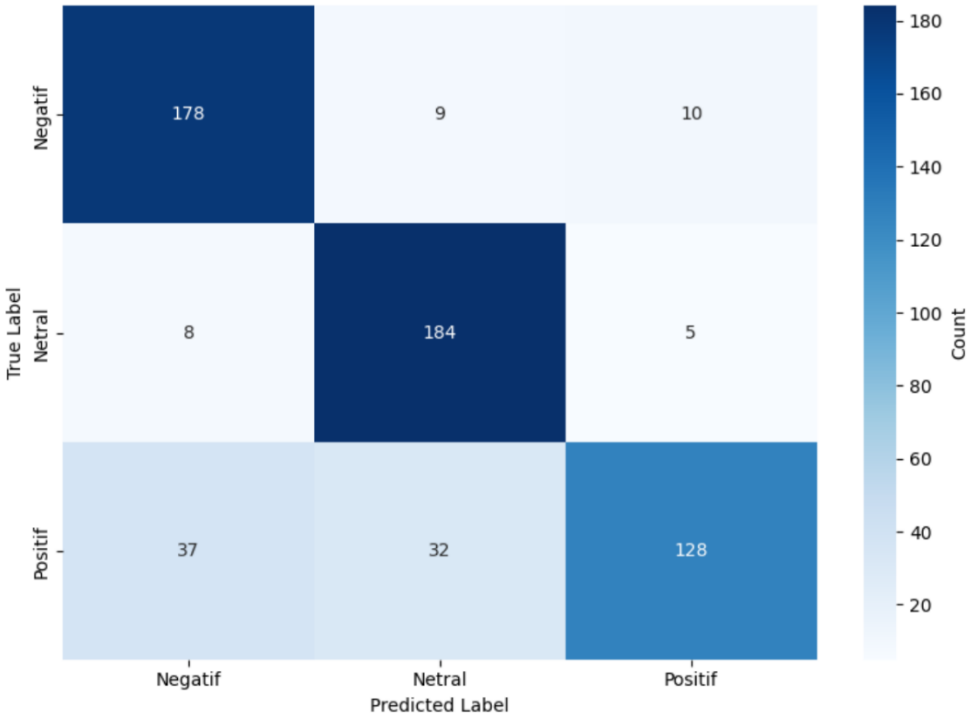
Accuracy: 0.9222 (92.22%)

	precision	recall	f1-score	support
Negatif	0.9048	0.9530	0.9283	788
Netral	0.8976	0.9898	0.9415	788
Positif	0.9759	0.8236	0.8933	788
accuracy			0.9222	2364
macro avg	0.9261	0.9222	0.9210	2364
weighted avg	0.9261	0.9222	0.9210	2364

Gambar 7. Hasil Akurasi Data Latih

3.5 Implementasi Data Uji dan Evaluasi Model

Pada tahap evaluasi dilakukan pengujian data menggunakan *confusion matriks*. Evaluasi model akan menunjukkan hasil *precision*, *recall*, dan f1-*score*, dan *accuracy* terhadap hasil data uji yang sebelumnya dipisahkan dari data latih. Sehingga mendapatkan hasil pengujian model mencapai akurasi evaluasi performa sebesar 0.8409, *precision* sebesar 0.8566, *recall* sebesar 0.8409 dan f1-score sebesar 0.8357. Hasil *confusion matriks* ditampilka pada gambar 8 menunjukkan distribusi prediksi model terhadap kelas. Dari total 591 data uji, menunjukkan kelas Positif diprediksi benar sebanyak 128 data, namun mengalami kesalahan, yaitu 37 data diprediksi sebagai Negatif, dan 32 data sebagai Netral. Kelas Negatif diprediksi benar sebanyak 178 data, dengan 9 data diprediksi sebagai Netral, dan 10 data sebagai Positif. Dan kelas Netral diprediksi sebanyak 184 data yang benar, dengan 8 data sebagai Negatif, dan 5 data salah sebagai positif. Distribusi ini menunjukkan model cukup baik dalam mengenal kelas Negatif dan Netral, namun masih kurang dalam membedakan sentiment Positif.



Gambar 8. *Confusion Matrix* Data Uji

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Pada gambar 9, ditunjukkan *classification report* hasil data uji. Dengan hasil *precision* tertinggi diperoleh pada kelas Positif sebesar 0.9552, menandakan bahwa sebagaian besar prediksi positf memang benar. Kemudian, *recall* diperoleh oleh kelas Netral sebesar 0.9746, menunjukkan bahwa model sangat baik dalam mengenali semua data yang benar-benar netral. Dan F1-*score* tertinggi dicapai oleh kelas Netral sebesar 0.8868. Berikut *classification report* evaluasi model dapat dilihat pada gambar 9.

	precision	recall	f1-score	support
Negatif	0.8009	0.8985	0.8469	197
Netral	0.8136	0.9746	0.8868	197
Positif	0.9552	0.6497	0.7734	197
accuracy			0.8409	591
macro avg	0.8566	0.8409	0.8357	591
weighted avg	0.8566	0.8409	0.8357	591

Gambar 9. *Clasification Report* Data Uji

Hasil *K-Fold Cross Validation* dengan K=5, ini dilakukan untuk menguji stabilitas dan konsistensi model. Hasil validasi silang menunjukkan hasil sebagai berikut pada gambar 10. Rata-rata akurasi dari 5-*fold* adalah sebesar 0.8450 dengan standar deviasi + 0.0120. Ini menunjukkan model memiliki performa yang cukup stabil.

```
Melakukan 5-fold Cross Validation untuk analisis stabilitas model
Fold 1/5 Accuracy: 0.8528
Fold 2/5 Accuracy: 0.8291
Fold 3/5 Accuracy: 0.8325
Fold 4/5 Accuracy: 0.8511
Fold 5/5 Accuracy: 0.8596
Akurasi rata-rata: 0.8450 ± 0.0120
```

Gambar 10. *K-Fold Validation*

4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dapat dikatakan bahwa penggunaan metode *Naïve Bayes* untuk analisis sentiment *Fear of Missing Out (FoMO)* pada media Twitter berhasil diterapkan dengan baik. Algoritma Naïve Bayes, yang diterapkan untuk mengklasifikasikan tingkat FoMO menggunakan *confusion matrix*, mencapai akurasi tertinggi sebesar 84.09%, precision 85,66%, recall 84,09%, dan F1-score 83,57%. Hasil evaluasi menunjukkan bahwa metode *Naïve Bayes* efektif untuk digunakan dalam analisis sentimen teks media sosial dengan *preprocessing* dan penyeimbangan data yang memadai. Selain itu, penggunaan metode SMOTE untuk *oversampling* menunjukkan kemampuan untuk memperbaiki distribusi data yang tidak seimbang, yang meningkatkan stabilitas model. Dalam mengoptimalkan model klasifikasi tingkat kecenderungan FoMO, perbaikan yang dapat dipertimbangkan. Seperti, eksplorasi metode alternatif seperti algoritma *Random Forest, KNN, Logistic Regression,* maupun metode lainnya. Kemudian, dapat memperluas dataset yang digunakan agar model memiliki cukup data yang lebih representastif terhadap berbagai variasi pola kecenderungan FoMO.

Daftar Pustaka

[1] Pengetahuan Umum, “Generasi Z dan Transformasi Gaya Hidup Sehat di Era Digital,” accessed Jun. 26, 2025. [Online]. Available: https://kumparan.com/pengetahuan-umum/generasi-z-dan-transformasi-gaya-hidup-sehat-di-era-digital-21yfPMGhW T9.

[2] Nurul Iffah Majidah, “Fenomena FOMO (Fear of Missing Out),” accessed Jun. 26, 2025. [Online]. Available: https://kumparan.com/nurul-iffah-1729041693881694630/fenomena-fomo-fear-of-missing-out-23kOl1XUp471.

[3] Serly Putri Jumbadi, “Lari Jadi Tren di Jogja, Gen Z Akui FOMO tapi Bawa Dampak Positif,” accessed Jun. 26, 2025. [Online]. Available: https://www.detik.com/jogja/berita/d7824516/lari-jadi-tren-di-jogja-gen-z-akui-fomo-tapi-bawa-dampak-positif.

[4] C. Novella Henni dan Samidi, “Optimizing Marketing Strategy by Predicting Fear of Missing Out: Machine Learning Approach,” *Ekonomis: Journal of Economics and Business*, vol. 9, no. 1, hlm. 659–670, 2025, doi: 10.33087/ekonomis.v9i1.2326.

[5] A. F. Setyaningsih, D. Septiyani, dan S. R. Widiasari, “Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepuasan Produk Skincare di Indonesia,” *Jurnal Teknologi Informatika dan Komputer MH. Thamrin*, vol. 9, no. 1, hlm. 224–235, Mar 2023, doi: 10.37012/jtik.v9i1.1409.

[6] Raden Budiarto Hadiprakoso, H. Setiawan, R. N. Yasa, and None Girinoto, “Text preprocessing for optimal accuracy in Indonesian sentiment analysis using a deep learning model with word embedding,” *AIP conference proceedings*, vol. 2879, pp. 020050–020050, Jan. 2023, doi: https://doi.org/10.1063/5.0126116.

[7] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, “Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation,” *Journal of Big Data*, vol. 8, no. 1, Jan. 2021, doi: https://doi.org/10.1186/s40537-021-00413-1.

[8] C. Maklin, “Synthetic Minority Over-sampling TEchnique (SMOTE),” accessed Jun. 26, 2025. [Online]. Available: https://medium.com/@corymaklin/synthetic-minority-over-sampling-technique-smote-7d419696b88c.

[9] C. Liu, Y. Sheng, Z. Wei, and Y. Yang, “Research of Text Classification Based on Improved TF-IDF Algorithm,” *IEEE Xplore*, Aug. 01, 2018. https://ieeexplore.ieee.org/abstract/document/8492945.

[10] S. Anggina, N. Y. Setiawan, and F. A. Bachtiar, “Analisis Ulasan Pelanggan Menggunakan Multinomial Naïve Bayes Classifier dengan Lexicon-Based dan TF-IDF Pada Formaggio Coffee and Resto,” *Accounting Information Systems and Information Technology Business Enterprise*, vol. 7, no. 1, pp. 76–90, Sep. 2022, doi: https://doi.org/10.34010/aisthebest.v7i1.7072.

[11] O. Rainio, J. Teuho, and R. Klén, “Evaluation metrics and statistical tests for machine learning,” *Scientific Reports*, vol. 14, no. 1, pp. 1–14, Mar. 2024, doi: https://doi.org/10.1038/s41598-024-56706-x.

[12] S. Sathyanarayanan, “Confusion Matrix-Based Performance Evaluation Metrics,” *African Journal of Biomedical Research*, vol. 27, no. 4s, pp. 4023–4031, Nov. 2024, doi: https://doi.org/10.53555/ajbr.v27i4s.4345.

[13] Sunarto Sunarto and H. Santoso, “Buku Saku Analisis Pareto,” Jul. 01, 2020. https://www.researchgate.net/publication/342586675_Buku_Saku_Analisis_Pareto