

Klasifikasi Model Konten Kuliner Viral UMKM di TikTok Menggunakan Algoritma Naive Bayes

Ziyad Habibul Mikraj*, Raissa Amanda Putri

Fakultas Sains dan Teknologi, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ¹*ziyadvivo90@gmail.com, ²raissa.ap@uinsu.ac.id

Email Penulis Korespondensi: ziyadvivo90@gmail.com

Submitted 26-01-2025; Accepted 19-02-2026; Published 25-02-2026

Abstrak

Penelitian ini membahas penyusunan model klasifikasi konten kuliner viral UMKM pada platform TikTok berdasarkan teks caption unggahan video. Permasalahan utama terletak pada tingginya variasi bahasa promosi, penggunaan istilah tren, serta struktur teks yang tidak baku sehingga identifikasi kategori kuliner viral sulit dilakukan secara manual dan tidak konsisten. Penelitian ini bertujuan merancang model klasifikasi yang sistematis untuk mengelompokkan caption TikTok UMKM ke dalam kategori kuliner viral secara otomatis dan terukur. Data yang digunakan berjumlah 800 caption yang diperoleh melalui scraping menggunakan Apify API. Proses penyusunan model meliputi tahap preprocessing berupa cleaning, case folding, tokenizing, normalisasi, stopword removal, stemming, dan detokenisasi untuk menghasilkan teks terstandar. Selanjutnya dilakukan pembobotan fitur menggunakan TF-IDF dan pembangunan model klasifikasi dengan algoritma Naïve Bayes. Model yang dihasilkan mampu mengelompokkan data ke dalam sepuluh kategori kuliner viral yaitu Donat Mochi, Cireng, Risol, Kentang Curly, Lukchup, Dimsum Mentai, Es Teh Jumbo, Indomie Telur, Mochi Daifuku, dan Dessert Box. Evaluasi menggunakan confusion matrix dan classification report menunjukkan akurasi sebesar 0,74 atau 74 persen. Hasil ini menunjukkan model mampu mendukung analisis tren kuliner viral UMKM secara otomatis.

Kata Kunci: TikTok; UMKM; Text Mining; TF-IDF; Naïve Bayes; Klasifikasi

Abstract

This study examines the development of a classification model for viral culinary content of Micro, Small, and Medium Enterprises on the TikTok platform based on video caption text. The main problem lies in the high variation of promotional language, the use of trending terms, and unstructured text formats, which make the identification of viral culinary categories difficult to perform manually and inconsistently. This study aims to design a systematic classification model to automatically and measurably group TikTok captions of enterprises into viral culinary categories. The dataset consists of 800 captions collected through scraping using the Apify API. The model development process includes preprocessing stages such as cleaning, case folding, tokenizing, normalization, stopword removal, stemming, and detokenization to produce standardized text. Feature weighting is then performed using TF-IDF, followed by model construction using the Naïve Bayes algorithm. The resulting model classifies data into ten viral culinary categories, namely Donat Mochi, Cireng, Risol, Kentang Curly, Lukchup, Dimsum Mentai, Es Teh Jumbo, Indomie Telur, Mochi Daifuku, and Dessert Box. Evaluation using a confusion matrix and classification report shows an accuracy of 0.74 or 74 percent. These results indicate the model supports automated analysis of viral culinary trends.

Keywords: TikTok; MSMEs; Text Mining; TF-IDF; Naïve Bayes; Classification

1. PENDAHULUAN

Perkembangan media sosial dalam beberapa tahun terakhir telah mengubah cara pelaku Usaha Mikro, Kecil, dan Menengah (UMKM) memasarkan produk kepada masyarakat [1]. TikTok menjadi salah satu platform yang paling banyak dimanfaatkan karena mampu menyebarkan informasi secara cepat melalui konten video pendek yang disertai *caption*[2]. *Caption* pada konten TikTok sering digunakan sebagai media promosi, penyampaian informasi produk, sekaligus penguat daya tarik konten untuk mendorong interaksi *audiens* [3].

Di sisi lain, tingginya intensitas unggahan konten pada TikTok menyebabkan munculnya berbagai fenomena kuliner viral yang cepat berubah dalam waktu singkat [4]. Kondisi tersebut berdampak pada sulitnya melakukan pemantauan tren secara manual, khususnya bagi UMKM yang membutuhkan informasi tren untuk menyesuaikan strategi pemasaran dan penawaran produk. Oleh karena itu, diperlukan pendekatan yang mampu memproses data teks dalam jumlah besar sehingga tren kuliner viral dapat diidentifikasi secara otomatis dan lebih objektif [5].

Pendekatan *text mining* dapat digunakan untuk mengolah *caption* TikTok menjadi informasi yang terstruktur melalui tahapan *preprocessing* dan pembobotan kata [6]. Tahap *preprocessing* penting untuk memastikan teks menjadi lebih bersih dan seragam agar dapat diproses oleh model klasifikasi[7]. Setelah teks diproses, metode pembobotan seperti TF-IDF dapat digunakan untuk merepresentasikan kata dalam bentuk nilai numerik, sehingga algoritma klasifikasi dapat bekerja berdasarkan pola kemunculan kata pada setiap *caption*[8].

Algoritma Naïve Bayes merupakan salah satu metode klasifikasi yang banyak digunakan dalam pengolahan teks karena sederhana, efisien, dan mampu memberikan hasil yang baik pada data berbasis dokumen[9]. Naïve Bayes bekerja dengan menghitung probabilitas suatu dokumen masuk ke dalam kelas tertentu berdasarkan kata-kata yang muncul pada dokumen tersebut[10]. Dengan kombinasi TF-IDF dan Naïve Bayes, *caption* TikTok dapat dikelompokkan ke dalam kategori kuliner viral sesuai pola kata yang dominan[11].

Pemanfaatan klasifikasi otomatis pada *caption* TikTok dapat membantu mempercepat proses analisis tren kuliner viral UMKM, karena setiap *caption* dapat dipetakan ke kategori tertentu berdasarkan konten yang dibahas[12]. Dalam konteks penelitian ini, klasifikasi tren kuliner viral dibentuk menjadi sepuluh kategori, sehingga hasil klasifikasi tidak

hanya memberikan informasi populer atau tidak populer, tetapi juga menunjukkan jenis produk kuliner viral yang sedang banyak dibahas pada TikTok[13].

Beberapa penelitian terdahulu menunjukkan bahwa teknik text mining dengan algoritma Naïve Bayes cukup efektif dalam mengolah data teks dari platform digital. Sari dan Pratama (2025) melakukan penelitian analisis sentimen produk pada *TikTok Shop* menggunakan pembobotan TF-IDF dan klasifikasi Naïve Bayes untuk mengidentifikasi kecenderungan opini pengguna[14]. Sudirman dan Ramadhan (2024) mengklasifikasikan sentimen masyarakat terhadap aplikasi TikTok menggunakan algoritma Naïve Bayes berdasarkan ulasan pengguna Google Play Store yang diperoleh melalui *web scraping*[15]. Selanjutnya, Putra dan rekan-rekan (2024) melakukan analisis sentimen *TikTok Shop* pada media sosial Twitter menggunakan algoritma Naïve Bayes untuk mengetahui kecenderungan opini publik terhadap fenomena *TikTok Shop*[16]. Selain itu, Wahyu Tisno, Ericka Keisya, dan Afifah Trista (2025) mengimplementasikan algoritma Naïve Bayes dalam analisis sentimen terhadap tren TikTok untuk mengukur performa klasifikasi berdasarkan data teks yang digunakan[17].

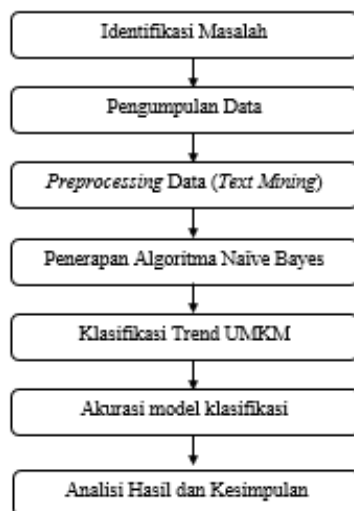
Namun, penelitian-penelitian tersebut lebih dominan berfokus pada analisis sentimen ulasan atau komentar, sementara kebutuhan UMKM tidak hanya terbatas pada sentimen, tetapi juga memerlukan pengelompokan tren berdasarkan jenis produk yang sedang viral. Perbedaan fokus tersebut menunjukkan adanya gap penelitian, yaitu perlunya klasifikasi berbasis kategori tren kuliner viral UMKM pada TikTok yang memanfaatkan teks caption sebagai sumber data utama agar informasi tren dapat diperoleh lebih spesifik dan dapat digunakan sebagai dasar pengambilan keputusan pemasaran.

Berdasarkan kondisi tersebut, penelitian ini bertujuan menyusun model klasifikasi konten kuliner viral UMKM di TikTok melalui tahapan preprocessing, pembobotan TF-IDF, serta penerapan algoritma Naïve Bayes pada 800 data caption hasil scraping. Model yang dikembangkan mengelompokkan data ke dalam sepuluh kategori kuliner viral dan dievaluasi menggunakan confusion matrix serta classification report untuk mengukur performa, dengan akurasi 0,74 atau 74% sebagai indikator bahwa model mampu melakukan klasifikasi secara efektif. Penelitian ini diharapkan dapat menjadi dasar pengembangan sistem analisis tren kuliner berbasis teks yang lebih akurat, mendukung pengambilan keputusan pelaku UMKM dalam strategi promosi digital, serta memberikan kontribusi metodologis dalam penerapan teknik klasifikasi teks pada konteks media sosial.

2. METODOLOGI PENELITIAN

2.1 Kerangka Penelitian dan Flowchart Penelitian

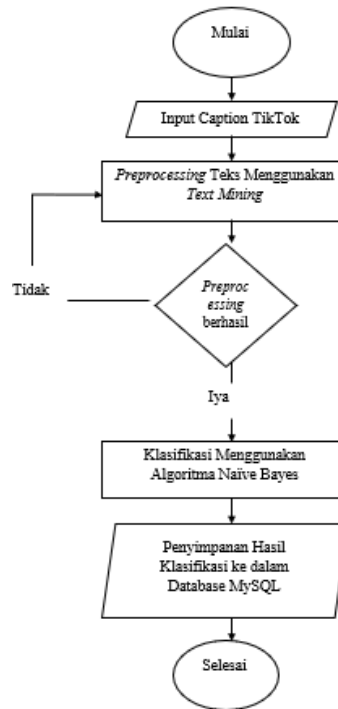
Pada penelitian ini, tahapan metodologi disusun secara sistematis agar proses klasifikasi konten kuliner viral UMKM pada platform TikTok dapat dilakukan secara terarah. Kerangka penelitian digunakan untuk menggambarkan alur umum mulai dari penentuan permasalahan, pengolahan data teks, penerapan metode klasifikasi, hingga analisis hasil akhir. Selain itu, *flowchart* penelitian digunakan untuk memperjelas urutan proses yang berjalan dari input data caption TikTok sampai menghasilkan keluaran klasifikasi yang tersimpan pada basis data.



Gambar 1. Kerangka Penelitian

Kerangka penelitian pada Gambar 1 menunjukkan bahwa penelitian dimulai dari tahap identifikasi masalah yang berfokus pada kebutuhan klasifikasi tren UMKM berdasarkan caption TikTok. Setelah masalah ditetapkan, tahap berikutnya adalah pengumpulan data, yaitu mengambil caption TikTok sebagai dataset penelitian. Data yang diperoleh kemudian diproses melalui *preprocessing* data berbasis text mining untuk mengubah teks mentah menjadi data yang bersih dan siap dianalisis. Tahap selanjutnya adalah penerapan algoritma Naïve Bayes sebagai metode klasifikasi utama untuk mengelompokkan caption ke kategori yang sesuai. Hasil dari proses tersebut menghasilkan klasifikasi tren UMKM

yang kemudian dievaluasi untuk memperoleh nilai akurasi model. Tahap akhir adalah analisis hasil dan penarikan kesimpulan berdasarkan performa model serta keluaran klasifikasi yang dihasilkan.



Gambar 2. Flowchart Penelitian

Flowchart pada Gambar 2 menggambarkan alur proses penelitian secara lebih detail dimulai dari kondisi “Mulai”, kemudian sistem menerima input berupa caption TikTok. Setelah data masuk, dilakukan *preprocessing* teks menggunakan tahapan *text mining* agar teks menjadi lebih terstruktur. Selanjutnya terdapat proses pengecekan keberhasilan *preprocessing*, jika *preprocessing* tidak berhasil maka sistem akan mengulang kembali proses *preprocessing* sampai data dapat diproses dengan benar. Setelah *preprocessing* berhasil, *caption* akan masuk ke tahap klasifikasi menggunakan algoritma Naïve Bayes untuk menentukan kategori kuliner viral yang sesuai. Hasil klasifikasi tersebut kemudian disimpan ke dalam database MySQL sebagai output sistem, dan proses penelitian berakhir pada tahap “Selesai”.

2.2 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan untuk memperoleh dataset berupa teks caption dari konten TikTok yang berkaitan dengan tren kuliner viral UMKM[18], dimana proses pengambilan data dilakukan secara otomatis menggunakan teknik scraping melalui Apify API sehingga lebih efisien dibandingkan pengumpulan manual, dengan rincian:

- Sumber data berasal dari platform TikTok.
- Data yang diambil berupa caption unggahan video.
- Jumlah data yang digunakan sebanyak 800 *caption*.
- Dataset yang telah terkumpul selanjutnya digunakan sebagai data masukan untuk tahap *preprocessing*, pembobotan TF-IDF, serta klasifikasi menggunakan algoritma Naïve Bayes.

2.3 Preprocessing Data

Preprocessing data dilakukan untuk membersihkan dan menyiapkan teks caption TikTok agar dapat diproses pada tahap pembobotan dan klasifikasi, dimana data teks mentah biasanya masih mengandung karakter tidak penting, variasi penulisan, serta kata-kata yang tidak relevan sehingga dapat memengaruhi hasil analisis[19]. Tahapan *preprocessing* dalam penelitian ini meliputi *cleaning* untuk menghilangkan tanda baca, angka, dan karakter khusus, *case folding* untuk menyeragamkan teks menjadi huruf kecil, *tokenizing* untuk memisahkan teks menjadi kata, normalisasi untuk mengubah kata tidak baku menjadi kata baku, *stopword removal* untuk menghapus kata umum yang kurang bermakna, *stemming* menggunakan Sastrawi untuk mengubah kata ke bentuk dasar, serta detokenisasi untuk menyusun kembali hasil pemrosesan menjadi teks yang siap digunakan pada tahap pembobotan TF-IDF dan klasifikasi Naïve Bayes.

2.4 Klasifikasi Naïve Bayes

Tahap klasifikasi dilakukan untuk mengelompokkan caption TikTok yang telah direpresentasikan dalam bentuk bobot TF-IDF ke dalam kategori kuliner viral yang sesuai[20]. Pada penelitian ini digunakan algoritma Naïve Bayes karena metode ini mampu melakukan klasifikasi teks secara efektif dengan menghitung probabilitas suatu dokumen termasuk ke

dalam kelas tertentu berdasarkan kata-kata yang terdapat di dalamnya. Naïve Bayes bekerja menggunakan prinsip probabilitas Bayes dengan asumsi bahwa setiap fitur (kata) bersifat independen, sehingga perhitungan probabilitas dapat dilakukan lebih sederhana dan efisien.

Rumus dasar Naïve Bayes yang digunakan adalah sebagai berikut.

$$P(C|d) = \frac{P(d|C) \times P(C)}{P(d)} \quad (1)$$

$$C = \arg \max_c (P(C) \times \prod_{i=1}^n P(t_i|C)) \quad (2)$$

Keterangan rumus Naïve Bayes pada penelitian ini adalah sebagai berikut:

- ($P(C|d)$) adalah probabilitas *posterior*, yaitu peluang dokumen (d) termasuk ke dalam kelas (C).
- ($P(d|C)$) adalah probabilitas *likelihood*, yaitu peluang dokumen (d) muncul pada kelas (C).
- ($P(C)$) adalah probabilitas *prior*, yaitu peluang awal suatu kelas (C) berdasarkan data latih.
- ($P(d)$) adalah probabilitas *evidence*, yaitu peluang dokumen (d) muncul dalam seluruh dataset.
- (t_i) adalah *term* atau kata ke-(i) yang terdapat pada dokumen (d).
- (n) adalah jumlah seluruh term pada dokumen (d).
- (C) merupakan kelas atau kategori yang diprediksi, dimana kelas dipilih berdasarkan nilai probabilitas terbesar melalui ($\arg \max$).

2.6 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kinerja algoritma Naïve Bayes dalam mengklasifikasikan *caption* TikTok ke dalam kategori kuliner viral yang telah ditentukan. Pada penelitian ini, evaluasi dilakukan *menggunakan confusion matrix* dan *classification report* untuk melihat nilai *precision*, *recall*, dan *f1-score* pada setiap kelas, serta mengukur tingkat akurasi keseluruhan dari model yang dibangun. Hasil evaluasi menunjukkan bahwa model mampu melakukan klasifikasi dengan akurasi sebesar 0,74 atau 74%, sehingga metode yang digunakan dinilai cukup efektif dalam melakukan klasifikasi konten kuliner viral UMKM pada platform TikTok.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data

Tahap analisis data dilakukan untuk menggambarkan karakteristik awal data yang digunakan, yaitu teks *caption* atau komentar pada konten TikTok terkait aktivitas UMKM, khususnya kuliner viral. Analisis ini bertujuan memastikan data sesuai kebutuhan penelitian dan dapat diproses pada tahap preprocessing, pembobotan TF-IDF, serta klasifikasi menggunakan algoritma Naïve Bayes. Hasil analisis menunjukkan bahwa *caption* TikTok memiliki variasi penulisan yang tinggi, mengandung kata tidak baku, tanda baca, dan gaya promosi yang beragam, sehingga diperlukan pembersihan data sebelum proses klasifikasi.

Tabel 1. Sampel Data Komentar

No	text
1	Resep Bolu ketan hitam tanpa mixer dan takaran sendok, cocok untuk temen ngopi hujan-hujan hee... #boluketanhitam #bolukukus #bolutanpamixer
2	Emang cireng kuahnya @Uma.food_ nggak ada tandingannya 😊😂😂 #monggonedhobyhilmi #mukbang #asmr #mukbangindonesia #cireng #cirengisi #cirengkuahpedes #umafood #cirengviral #makananviral #bikinngiler #pedasnyanaiklevel #viraltiktok #masukberandafyp #fyp #foryou
3	ENAKNYA KIRIM KE SIAPA YA GESS BUAT DI REVIEW? #dimsummentai BOLU KETAN HITAM KEJU LUMER YANG LAGI HITS DI JAKTIM 📍📍📍 IG @gulajawabakery
4	🏠 GULA JAWA BAKERY 🕒 08.00-21.00 📍 Jl. I Gusti Ngurah Rai No.21, Cipinang Muara, Kec. Duren Sawit
5	ENAKNYA KIRIM KE SIAPA YA GESS BUAT DI REVIEW? #dimsummentai Enak banget lagiii. udah bikin 2x dalam sehari wkwk. cocok banget buat cemilan kalo lagi iseng dirumah.
6	bikinnya gampang & satset banget. rasanya pedesss, gurih, kenyel2 ada garing nya juga 😊👍 #cidog #viraltiktok #fyp

7	Donat Mochi ini definisi comfort food, buatnya gampang gak perlu di resting adonan donat nya ☺🌟☐☐ #donat #donatmochi #mochi
8	Donat mochi hadir di Happy Bites Dessert Cafe baru di Cikarang, wajib cobain !! #kulinerbekasi #donatmochi #donat #bekasi
9	Isi 6 atau 10 yg penting wajib punya stock risolusi ☺ #bakulrisol #risol #risolragout #risolmayo #risolmayomargo #smallbusiness #belidong #risolmargo #risolusi #risoles #cemilan #frozenfood CHEESECAKE TOM AND JERRY 🌟
10	Tom & jerry cheesecake bentuknya cantik unik rasanya enak ngeju dan creamy banget. #cheesecake #tomandjerry #resepati #cheesecaketomandjerry #tomandjerrycheesecake

Berdasarkan sampel data pada Tabel 1, dapat terlihat bahwa caption atau komentar TikTok memiliki pola kalimat yang singkat, menggunakan bahasa sehari-hari, dan sering memuat kata-kata yang menunjukkan jenis produk kuliner tertentu. Hal tersebut menunjukkan bahwa teks caption TikTok dapat dimanfaatkan sebagai sumber informasi untuk mengidentifikasi tren kuliner viral UMKM, namun tetap memerlukan pengolahan lebih lanjut agar hasil klasifikasi yang diperoleh menjadi lebih akurat dan terstruktur.

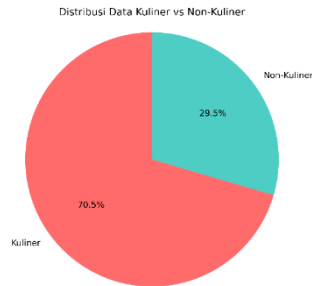
3.2 Pelabelan dan Visualisasi Dataset

Tahap pelabelan dilakukan untuk memberikan kategori pada setiap data caption TikTok sehingga data memiliki kelas yang jelas dan dapat digunakan dalam proses klasifikasi. Pelabelan pada penelitian ini bertujuan untuk memetakan data sesuai dengan kebutuhan analisis, yaitu membedakan data yang termasuk kategori UMKM serta mengidentifikasi keterkaitan caption dengan konten kuliner. Data yang telah diberi label selanjutnya divisualisasikan untuk melihat distribusi kategori serta kecenderungan sentimen pada dataset yang digunakan.

Tabel 2. Pelabelan Dataset UMKM

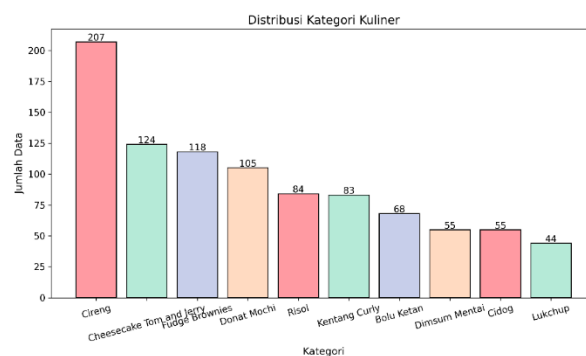
No.	prosescomment	kategori_kuliner
1	resep bolu ketan hitam mixer takar sendok cocok teman ngopi hujanhujan hee boluketanhitam bolukukus bolutanpamixer	Bolu Ketan
2	cireng kuah food tandingannyamongonedhobyhilmi mukbang asmr mukbangindonesia cireng cirengisi cirengkuahpedes umafood cirengviral makananviral bikinngiler pedasnyanaiklevel viraltiktok masukberandafyp fyp foryou	Cireng
3	enak kirim ya gess review dimsummentai	Dimsum Mentai
4	bolu ketan hitam keju lumer hits jaktim ig gula jawa bakery jl i gusti ngurah rai no cipinang muara kec duren sawit	Bolu Ketan
5	enak kirim ya gess review dimsummentai	Dimsum Mentai
6	enak banget bikin kali hari wkwk cocok banget cemilan kalo iseng rumah bikin gampang satset banget pedesss gurih kenyel garing cidog viraltiktok fyp	Cidog
7	donat mochi definisi comfort food buat gampang resting adon donat donat donatmochi mochi	Donat Mochi
8	donat mochi hadir happy bites dessert cafe cikarang wajib coba kulinerbekasi donatmochi donat bekas	Donat Mochi
9	isi wajib stock risolusi bakulrisol risol risolragout risolmayo risolmayomargo smallbusiness belidong risolmargo risolusi risoles cemilan frozenfood	Risol
10	cheesecake tom and jerrytom jerry cheesecake bentuk cantik unik enak ngeju creamy bangetcheesecake tomandjerry resepati cheesecaketomandjerry tomandjerrycheesecake	Cheesecake Tom and Jerry

Berdasarkan Tabel 2, pelabelan dataset dilakukan untuk memastikan setiap *caption* memiliki kategori yang sesuai sehingga proses analisis dan klasifikasi dapat berjalan secara terstruktur. Pelabelan ini menjadi dasar bagi tahapan berikutnya, terutama untuk menentukan fitur dan kelas yang akan diproses oleh model Naïve Bayes.



Gambar 3. Visualisasi Perbandingan Kategori Kuliner dan Non Kuliner

Berdasarkan gambar tersebut, distribusi dataset menunjukkan bahwa data kuliner mendominasi dengan persentase 70,5%, sedangkan data non-kuliner sebesar 29,5%. Hasil ini menunjukkan bahwa caption TikTok yang dikumpulkan lebih banyak membahas konten kuliner, sehingga dataset penelitian lebih berfokus pada tren kuliner viral UMKM dan dapat memengaruhi proses klasifikasi karena jumlah data kuliner lebih besar dibandingkan non-kuliner.



Gambar 4. Visualisasi Jumlah Analisis Sentimen

Berdasarkan Gambar 4, hasil visualisasi analisis sentimen memperlihatkan jumlah data yang termasuk ke dalam sentimen positif dan negatif. Visualisasi ini membantu memahami kecenderungan respons pada caption TikTok serta dapat menjadi informasi tambahan dalam melihat karakteristik dataset yang digunakan dalam penelitian..

3.3 Hasil Pembobotan TF-IDF

Tahap pembobotan TF-IDF digunakan untuk mengubah teks caption TikTok yang telah di-*preprocessing* menjadi fitur numerik agar dapat dijadikan input pada klasifikasi menggunakan algoritma Naïve Bayes. Metode TF-IDF memberikan nilai bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen serta tingkat kekhasannya pada keseluruhan data, sehingga kata yang lebih spesifik dan relevan terhadap suatu kategori akan memiliki bobot lebih tinggi dibandingkan kata yang umum muncul pada banyak caption.

Tabel 3. Tabel TF-IDF.

Kata	Dokumen Muncul	TF-IDF (Setiap Dokumen)
produk	D1, D2	0.066, 0.079
lokal	D1	0.116
UMKM	D1–D5	0.000
kreatif	D1	0.116
inovatif	D1	0.116
keren	D1	0.116
harga	D2	0.1398
mahal	D2	0.1398
toko	D2	0.1398
besar	D2	0.1398
dukungan	D3	0.1398
pemerintah	D3	0.1398
membantu	D3	0.1398
berkembang	D3	0.1398

banyak	D4	0.1747
promosi	D4	0.1747
gulung	D4	0.1747
tikar	D4	0.1747
tulang	D5	0.1747
ekonomi	D5	0.1747
masyarakat	D5	0.1747
bangkit	D5	0.1747

Berdasarkan Tabel 3., nilai TF-IDF menunjukkan tingkat kepentingan setiap kata pada dokumen tertentu dalam dataset, dimana kata yang muncul pada dokumen yang lebih spesifik akan menghasilkan bobot yang lebih besar. Pada tabel terlihat bahwa kata seperti “banyak”, “promosi”, “gluten”, “tiktok”, “tulisan”, “ekonomi”, “masyarakat”, dan “bangkit” memiliki nilai TF-IDF tertinggi yaitu 0,1747, sehingga kata-kata tersebut lebih dominan dalam merepresentasikan dokumen tempat kata tersebut muncul. Sementara itu kata seperti “lokal”, “kreatif”, “inovatif”, dan “keren” memiliki bobot 0,116, sedangkan kata “produk” memiliki nilai lebih kecil karena muncul pada lebih dari satu dokumen sehingga bobotnya terbagi. Hasil pembobotan ini digunakan sebagai fitur numerik yang menjadi dasar bagi algoritma Naïve Bayes dalam mengenali pola kata pada caption TikTok dan melakukan klasifikasi kategori kuliner viral UMKM secara lebih terstruktur.

3.4 Wordcloud

Pada bagian ini disajikan visualisasi wordcloud untuk menggambarkan distribusi kata yang dominan dalam keseluruhan data caption. Visualisasi ini bertujuan memberikan gambaran awal mengenai karakteristik teks setelah melalui tahapan preprocessing, sehingga pola kemunculan istilah yang paling sering digunakan dalam promosi kuliner viral UMKM di TikTok dapat diidentifikasi secara visual. Penyajian wordcloud berfungsi sebagai analisis eksploratif untuk mendukung pemahaman terhadap struktur dan kecenderungan kata sebelum dilakukan pembahasan lebih lanjut pada tahap klasifikasi.



Gambar 5. Wordcloud

Berdasarkan visualisasi *wordcloud* pada gambar 5., terlihat bahwa beberapa kata muncul lebih dominan dibandingkan kata lainnya, yang menunjukkan adanya kecenderungan topik dan istilah tertentu yang sering digunakan pada caption TikTok. Kata-kata dominan tersebut menjadi indikasi bahwa dataset memiliki karakteristik yang relevan dengan pembahasan tren kuliner viral UMKM, serta mendukung proses klasifikasi karena kata-kata yang sering muncul dapat menjadi fitur penting dalam pembobotan TF-IDF dan perhitungan probabilitas pada algoritma Naïve Bayes.

3.5 Hasil Evaluasi Confusion Matrix

Pada bagian ini disajikan hasil evaluasi model klasifikasi menggunakan confusion matrix untuk mengukur kinerja model secara terperinci. Confusion matrix digunakan untuk menunjukkan distribusi prediksi yang benar dan salah pada setiap kategori kuliner viral, sehingga tingkat ketepatan klasifikasi dapat dianalisis berdasarkan masing-masing kelas. Melalui evaluasi ini, dapat diidentifikasi pola kesalahan prediksi serta kategori yang memiliki tingkat akurasi, presisi, dan recall

yang tinggi maupun rendah, sehingga memberikan gambaran menyeluruh terhadap performa model yang telah dikembangkan.

Classification Report:				
	precision	recall	f1-score	support
Bolu Ketan	0.75	0.69	0.72	13
Cheesecake Tom and Jerry	0.89	0.96	0.92	25
Cidog	1.00	0.45	0.62	11
Cireng	0.49	0.80	0.61	41
Dimsum Mentai	1.00	0.27	0.43	11
Donat Mochi	1.00	0.86	0.92	21
Fudge Brownies	0.79	0.96	0.87	24
Kentang Curly	0.87	0.76	0.81	17
Lukchup	1.00	0.44	0.62	9
Risol	1.00	0.47	0.64	17
accuracy			0.74	189
macro avg	0.88	0.67	0.72	189
weighted avg	0.82	0.74	0.74	189

Gambar 6. Hasil *Confusion Matrix / Classification Report*

Berdasarkan hasil evaluasi pada gambar 6., model Naïve Bayes memperoleh akurasi sebesar 0,74 atau 74% dari total 189 data uji, yang menunjukkan bahwa model mampu mengklasifikasikan caption TikTok dengan cukup baik. Kelas *Cheesecake Tom and Jerry* dan Donat Mochi memiliki performa tinggi dengan nilai f1-score sebesar 0,92, sementara kelas Fudge Brownies juga menunjukkan hasil f1-score 0,87. Namun, beberapa kelas seperti Dimsum Mentai memiliki nilai recall yang rendah yaitu 0,27 dan f1-score 0,43, yang mengindikasikan masih adanya kesalahan prediksi pada kelas tersebut. Hasil ini menunjukkan bahwa metode TF-IDF dan Naïve Bayes efektif digunakan untuk klasifikasi tren kuliner viral UMKM pada TikTok, tetapi performa pada beberapa kategori masih dapat ditingkatkan melalui penambahan data latih atau pengembangan metode klasifikasi yang lebih optimal.

4. KESIMPULAN

Dari hasil yang telah dilakukan, dapat didapati bahwa penerapan algoritma Naïve Bayes mampu digunakan untuk melakukan klasifikasi konten kuliner viral UMKM pada platform TikTok berdasarkan teks caption. Data penelitian dikumpulkan melalui proses scraping menggunakan Apify API dengan jumlah 800 caption TikTok, kemudian diproses menggunakan tahapan preprocessing text mining yang terdiri dari cleaning, case folding, tokenizing, normalisasi, stopword removal, stemming, dan detokenisasi sehingga data menjadi lebih bersih dan terstruktur. Selanjutnya, pembobotan TF-IDF digunakan untuk mengubah caption menjadi fitur numerik yang dapat diproses oleh model klasifikasi, sehingga kata-kata penting pada caption dapat memberikan kontribusi dalam menentukan kategori. Hasil klasifikasi menghasilkan beberapa kategori kuliner viral yang dapat menggambarkan tren yang sedang banyak dibahas pada TikTok, serta menunjukkan bahwa pendekatan ini dapat membantu identifikasi tren secara otomatis dan lebih efisien dibandingkan analisis manual. Berdasarkan evaluasi menggunakan confusion matrix dan classification report, model memperoleh nilai akurasi sebesar 0,74 atau 74% yang menunjukkan bahwa metode yang digunakan memiliki performa cukup baik dalam mengelompokkan caption ke dalam kelas yang sesuai. Namun demikian, penelitian ini masih memiliki keterbatasan, terutama pada distribusi jumlah data antar kategori yang belum seimbang dan adanya beberapa kelas yang memiliki nilai recall rendah sehingga memengaruhi ketepatan prediksi pada kategori tertentu. Selain itu, variasi bahasa pada caption TikTok yang sangat beragam juga dapat memengaruhi hasil klasifikasi meskipun preprocessing telah dilakukan. Oleh karena itu, penelitian selanjutnya dapat dikembangkan dengan menambah jumlah dataset, melakukan penyeimbangan data pada setiap kategori, serta mencoba penggunaan metode klasifikasi lain agar performa model dapat ditingkatkan.

REFERENCES

- [1] C. Yolanda *et al.*, “peran usaha mikro, kecil dan menengah (UMKM) dalam pengembangan ekonomi indonesia,” vol. 2, no. 3, hal. 170–186, 2024.
- [2] T. Sudrartono *et al.*, *Kewirausahaan Umkm Di Era Digital*. 2022.
- [3] I. K. Dewi, M. Y. R.Pandin, dan A. D. GS, “Peningkatan Kinerja UMKM Melalui Pengelolaan Keuangan,” vol. 7, no. April, hal. 23–36, 2022.
- [4] F. Gilang dan R. Panggabean, “Application Of Naive Bayes Algorithm For Sentiment Analysis On Economic Recession Threat,” vol. 1, no. 1, hal. 25–32, 2025.
- [5] M.Husaini, S. Raudah, dan M. Amaliya, “Implementasi Prorgam Perluasan Jangkauan UMKM Di Kabupaten Balangan,” vol. 2, no. 6, hal. 2134–2139, 2023.
- [6] Iis Dewi Ratih, S M Retnaningsih, dan V M Dewi, “Klasifikasi Kualitas Tanah Menggunakan Metode Naive Bayes Classifier,” *J. Apl. Mat. dan Stat.*, vol. 1, no. 1, hal. 11–20, 2022, doi: 10.53625/jams.v1i1.4427.
- [7] Armansyah dan R. K. Ramli, “Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes,” vol. 6, no. 1, hal. 1–10, 2022, doi: 10.29408/edumatic.v6i1.4789.
- [8] A. Rahman, “Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes,” *J. SAINTEKOM*,

- vol. 13, no. 1, hal. 22–31, 2023, doi: 10.33020/saintekom.v13i1.349.
- [9] A. F. Riany, G. Testiana, S. S. Informasi, dan K. Palembang, “Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes Sedangkan Provinsi di Indonesia dengan,” vol. 9, hal. 42–54, 2023.
 - [10] M. S. Hasibuan dan A. Serdano, “Analisis Sentimen Kebijakan Pembelajaran Tatap Muka Menggunakan Support Vector Machine dan Naive Bayes Policy Sentiment Analysis Face-to-face Learning Using Supports Vector and Naive Bayes Engines,” vol. 6, no. 2, hal. 199–204, 2022.
 - [11] B. Imran, M. N. Karim, dan N. I. Ningsih, “Klasifikasi Berita Hoax Terkait Pemilihan Umum Presiden Republik Indonesia Tahun 2024 Menggunakan Naïve Bayes Dan Svm,” vol. 20, 2024.
 - [12] A. A. A. Arifin, W. Handoko, dan Z. Efendi, “Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan,” *J-Com (Journal Comput.*, vol. 2, no. 1, hal. 21–26, 2022, doi: 10.33330/j-com.v2i1.1577.
 - [13] Sutisna dan N. M. Yuniar, “Klasifikasi Kualitas Air Bersih Menggunakan Metode Naïve baiyes,” *J. Sains dan Teknol.*, vol. 5, no. 1, hal. 243–246, 2023, [Daring]. Tersedia pada: <https://doi.org/10.55338/saintek.v5i1.1383>.
 - [14] M. Ziddan, H. F. Adiyatma, dan A. P. Sari, “Analisis Sentimen Produk pada Tiktok Shop dengan Metode Naive Bayes,” vol. 9, hal. 22098–22106, 2025.
 - [15] M. Nahdhudin, N. Cahyo, H. Wibowo, M. R. Handayani, dan K. Umam, “Klasifikasi Sentimen Masyarakat Terhadap Aplikasi Tiktok Menggunakan Algoritma Naive Bayes,” vol. 1, no. 3, hal. 107–112, 2025.
 - [16] D. Kariman *et al.*, “Analisis Sentimen TikTok Shop pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes,” 2024.
 - [17] W. T. Atmojo, E. Keisya, dan A. T. Ayunda, “Mplementasi Algoritma Naïve Bayes Dalam Analisis Sentimen Terhadap Trend Tiktok,” vol. 13, no. 2, 2025.
 - [18] R. Setiawan dan A. Triayudi, “Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K- Nearest Neighbor Berbasis Web,” vol. 6, no. 2, hal. 777–785, 2022, doi: 10.30865/mib.v6i2.3566.
 - [19] A. Tangkelayuk dan E. Mailoa, “Klasifikasi Kualitas Air Menggunakan Metode KNN , Naïve Bayes Dan Decision Tree,” vol. 9, no. 2, hal. 1109–1119, 2022.
 - [20] N. Christian, R. Riego, dan D. B. Villarba, “Utilization of Multinomial Naive Bayes Algorithm and Term Frequency-Inverse Document Frequency (TF-IDF Vectorizer) in Checking the Credibility of News Tweet in the Philippines,” *arXiv Prepr. arXiv2306.00018*, 2023.