

KLASIFIKASI KERAMAIAAN VIDEO KULINER BERBASIS FITUR ENGAGEMENT, SENTIMEN, DAN KONTEN MENGUNAKAN METODOLOGI CRISP-DM

Oleh : Adinda Rahimah Azzahra¹, Sukamto²

Jurusan Teknik Elektro^{1,2}, Politeknik Negeri Semarang^{1,2}

E-mail : adinda.43323201@mhs.polines.ac.id¹, sukamto@polines.ac.id²

Abstrak

Penelitian ini bertujuan menganalisis faktor yang memengaruhi keramaian video kuliner TikTok dengan membandingkan konten bertema keju dan non-keju. Metode yang digunakan adalah Cross-Industry Standard Process for Data Mining (CRISP-DM) sebagai kerangka kerja utama. Dataset mencakup fitur engagement serta sentimen komentar, kemudian diproses menggunakan dua algoritma klasifikasi, yaitu Decision Tree (DT) dan Random Forest (RF). Evaluasi model dilakukan menggunakan cross-validation untuk memperoleh performa yang stabil. Hasil menunjukkan bahwa konten kuliner berbahan keju cenderung memperoleh interaksi lebih tinggi, khususnya melalui likes dan shares. Model RF memberikan akurasi terbaik dibandingkan DT. Penelitian ini membuktikan bahwa pemilihan bahan makanan, terutama keju, dapat menjadi indikator penting dalam strategi penyusunan konten kuliner di TikTok.

Kata kunci : CRISP-DM, TikTok, Keju, Random Forest, Decision Tree.

Abstract

This study aims to analyze the factors that influence the popularity of culinary TikTok videos by comparing cheese-based content with non-cheese content. The Cross-Industry Standard Process for Data Mining (CRISP-DM) was applied as the main methodological framework. The dataset includes engagement metrics and comment sentiment features, which were processed using two classification algorithms: Decision Tree (DT) and Random Forest (RF). Model evaluation was performed using cross-validation to obtain stable performance results. The findings indicate that cheese-related content tends to generate higher engagement, particularly through likes and shares. RF achieved better accuracy than DT in predicting whether a video becomes “viral” or not. This research demonstrates that food ingredients, especially cheese, can serve as important indicators in developing effective culinary content strategies on TikTok.

Keywords : CRISP-DM, TikTok, cheese content, Random Forest, Decision Tree.

1. PENDAHULUAN

Perkembangan media sosial telah mengubah pola konsumsi konten digital masyarakat, termasuk dalam bidang kuliner. Platform berbasis video pendek seperti TikTok memungkinkan pengguna untuk membagikan pengalaman kuliner secara visual, interaktif, dan cepat, sehingga berperan besar dalam pembentukan tren makanan dan perilaku konsumsi masyarakat (Kaye et al., 2022). TikTok tidak hanya menjadi media hiburan, tetapi juga sarana promosi kuliner yang efektif

karena menggabungkan unsur visual, audio, dan interaksi pengguna dalam satu konten (Zeng & Abidin, 2021).

Popularitas suatu video kuliner di TikTok umumnya diukur melalui tingkat engagement pengguna, seperti jumlah likes, comments, dan shares. Penelitian sebelumnya menunjukkan bahwa engagement dipengaruhi oleh berbagai faktor, antara lain tema konten, bahan makanan yang digunakan, gaya penyajian, durasi video, serta emosi yang ditampilkan dalam konten (Hu et al., 2023; Omar & Dequan, 2020).

Tingginya engagement mencerminkan ketertarikan audiens dan menjadi indikator penting dalam menilai keramaian atau viralitas suatu video.

Salah satu bahan makanan yang sering muncul dan menjadi daya tarik utama dalam konten kuliner adalah keju. Keju kerap diasosiasikan dengan rasa gurih, tekstur meleleh, dan tampilan visual yang menggugah selera, sehingga mampu meningkatkan respons emosional audiens (Spence, 2020). Beberapa studi menyebutkan bahwa konten makanan dengan visual yang kuat cenderung mendapatkan tingkat interaksi yang lebih tinggi dibandingkan konten dengan visual biasa (Velasco et al., 2018). Namun demikian, penelitian yang secara khusus membandingkan tingkat engagement antara konten kuliner berbasis keju dan non-keju di TikTok masih terbatas.

Analisis konten media sosial dalam skala besar memerlukan pendekatan data mining untuk mengekstraksi pola dan pengetahuan dari data yang kompleks dan berjumlah besar (Han et al., 2022). Data mining memungkinkan pengolahan metadata video dan teks komentar pengguna untuk mengidentifikasi faktor-faktor yang memengaruhi popularitas konten (Aggarwal, 2015). Salah satu kerangka kerja yang banyak digunakan dalam proses data mining adalah Cross-Industry Standard Process for Data Mining (CRISP-DM) karena memberikan tahapan analisis yang sistematis dan terstruktur (Wirth & Hipp, 2000).

Dalam penelitian ini, fitur engagement berupa likes, comments, dan shares, serta sentimen komentar pengguna digunakan untuk membangun model klasifikasi tingkat keramaian video kuliner di TikTok. Dua algoritma machine learning, yaitu Decision Tree dan Random Forest, digunakan dan dibandingkan untuk mengetahui algoritma dengan performa terbaik

dalam mengklasifikasikan tingkat engagement video. Dengan demikian, penelitian ini diharapkan dapat memberikan gambaran mengenai pengaruh tema kuliner berbasis keju terhadap popularitas konten TikTok serta kontribusi data mining dalam analisis tren kuliner digital.

2. Tinjauan Pustaka

Media sosial telah menjadi sumber data penting dalam menganalisis perilaku dan preferensi masyarakat. Platform seperti TikTok memiliki karakteristik berupa konten video pendek, algoritma rekomendasi yang kuat, serta tingkat interaksi pengguna yang tinggi, sehingga banyak dimanfaatkan dalam penelitian terkait tren, pemasaran digital, dan perilaku konsumen (Montag et al., 2021; Kaye et al., 2022). Konten kuliner di TikTok sering kali menampilkan proses memasak, review makanan, dan eksplorasi bahan tertentu yang mampu membentuk persepsi dan minat audiens.

Engagement pengguna merupakan indikator utama dalam menilai keberhasilan sebuah konten media sosial. Likes, comments, dan shares mencerminkan tingkat keterlibatan audiens dan sering digunakan sebagai parameter popularitas konten (Cvijikj & Michahelles, 2013). Penelitian oleh Hu et al. (2023) menunjukkan bahwa konten dengan visual menarik dan tema yang relevan memiliki korelasi positif terhadap peningkatan engagement. Selain itu, analisis sentimen komentar dapat memberikan gambaran respon emosional pengguna terhadap suatu konten (Liu, 2012).

Dalam konteks kuliner, visual makanan berperan penting dalam menarik perhatian audiens. Spence (2020) menjelaskan bahwa aspek visual seperti warna, tekstur, dan efek

lelahan dapat memicu respons sensorik dan emosional. Keju sebagai bahan makanan memiliki karakteristik visual yang kuat, sehingga sering digunakan sebagai elemen utama dalam konten makanan viral (Velasco et al., 2018). Namun, perbandingan kuantitatif antara konten berbasis keju dan non-keju terhadap tingkat engagement masih jarang dibahas secara mendalam.

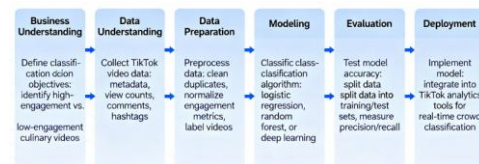
Analisis data media sosial umumnya memanfaatkan pendekatan data mining dan text mining untuk mengekstraksi informasi dari data terstruktur maupun tidak terstruktur (Aggarwal, 2015). Teknik text mining seperti TF-IDF dan analisis sentimen telah terbukti efektif dalam mengolah komentar pengguna pada media sosial (Haris et al., 2025; Liu, 2012). Proses data mining sering mengikuti kerangka kerja CRISP-DM yang terdiri dari enam tahap, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment (Wirth & Hipp, 2000).

Dalam tahap modeling, algoritma klasifikasi banyak digunakan untuk memprediksi kategori atau tingkat popularitas konten. Decision Tree dikenal mudah diinterpretasikan dan efektif untuk dataset dengan aturan yang jelas (Quinlan, 1993). Sementara itu, Random Forest merupakan pengembangan dari Decision Tree yang menggabungkan banyak pohon keputusan sehingga lebih stabil dan memiliki akurasi yang lebih tinggi (Breiman, 2001). Penelitian sebelumnya menunjukkan bahwa Random Forest unggul dalam menangani kombinasi fitur numerik dan teks serta tahan terhadap noise data (Anggraeni, 2024; Han et al., 2022).

Berdasarkan kajian pustaka tersebut,

dapat disimpulkan bahwa integrasi media sosial TikTok, data mining, CRISP-DM, serta algoritma Decision Tree dan Random Forest memberikan landasan yang kuat untuk menganalisis tingkat keramaian konten kuliner. Penelitian ini memposisikan diri untuk mengisi celah penelitian dengan membandingkan konten kuliner berbasis keju dan non-keju serta mengevaluasi performa model klasifikasi dalam memprediksi tingkat engagement video TikTok.

3. METODE PENELITIAN



Gambar 1. Diagram alir metode CRISP-DM yang dimodifikasi oleh penulis

Penelitian ini menggunakan metode CRISP-DM (Cross-Industry Standard Process for Data Mining) sebagai kerangka kerja utama dalam proses analisis data. Metode CRISP-DM dipilih karena mampu memberikan alur kerja yang sistematis, fleksibel, dan terstruktur dalam menyelesaikan permasalahan data mining, khususnya pada analisis data media sosial yang bersifat numerik dan semi-terstruktur. Menurut Wirth dan Hipp (2000), CRISP-DM terdiri dari enam tahapan utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment.

3.1 Business Understanding

berangkat dari kebutuhan untuk memahami faktor apa saja yang membuat video kuliner di TikTok menjadi ramai ditonton dan tinggi interaksi, sehingga kreator dan brand dapat mengoptimalkan strategi konten secara terukur. Fokus utama diarahkan pada perbedaan performa antara konten bertema keju dan non-keju, dengan dugaan bahwa unsur

bahan (ingredient appeal) memengaruhi minat dan keterlibatan audiens. Tujuan analitis yang ingin dicapai meliputi:

1. Mengidentifikasi fitur-fitur kunci yang berkontribusi terhadap keramaian video, khususnya metrik engagement (likes, comments, shares) dan sentimen komentar yaitu positif, negatif, dan netral.
2. Membandingkan dampak tema konten keju versus non-keju terhadap tingkat keramaian, sehingga memberi dasar keputusan kreatif berbasis data.
3. Membandingkan performa algoritma Random Forest dengan Decision Tree
4. Memberikan kontribusi terhadap strategi penyusunan konten kuliner di media sosial dengan mengungkap faktor-faktor konten yang paling berpengaruh terhadap interaksi pengguna, serta menjadi referensi dalam analisis konten digital berbasis fitur engagement dan sentimen.

Hasil dari tahap Business Understanding ini menjadi dasar dari tahapan berikutnya dalam CRISP-DM

3.2 Data Understanding

Tahap Data Preparation bertujuan untuk menyiapkan data agar siap digunakan dalam proses pemodelan. Tahap ini mencakup seleksi fitur dan pembagian dataset menjadi data latih dan data uji, sebagaimana dijelaskan oleh Wirth dan Hipp (2000). Fitur yang digunakan dalam penelitian ini adalah fitur numerik yang berkaitan langsung dengan engagement, yaitu likes, comments, saves, shares, serta sentimen komentar. Dataset dibagi menjadi data latih dan data uji untuk menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Pembagian data ini merupakan praktik umum dalam pengembangan model machine learning untuk menghindari overfitting (Géron, 2019).

Setelah data terkumpul, dilakukan deskripsi dan eksplorasi awal untuk memahami karakteristik distribusi tiap fitur, misalnya sebaran jumlah likes

dan shares, proporsi video keju dibandingkan non-keju, serta kecenderungan sentimen komentar pada masing-masing kategori. Proses eksplorasi ini dapat mencakup visualisasi sederhana, pengecekan nilai ekstrim (outlier), identifikasi nilai hilang, dan pengamatan korelasi awal antara fitur engagement dengan label keramaian.

Hasil pemahaman data di fase ini menjadi dasar perancangan strategi Data Preparation, termasuk pemilihan fitur yang relevan, teknik penanganan ketidakseimbangan kelas, serta pendekatan pra-pemrosesan teks untuk analisis sentimen komentar TikTok.

3.3 Data Preparation

Pada tahap Data Preparation, data keramaian video kuliner TikTok diproses menjadi dataset akhir yang siap digunakan untuk pemodelan dengan cara menyeleksi hanya video yang memenuhi kriteria (tema keju/non-keju, memiliki engagement dan komentar yang memadai), membersihkan data dari duplikasi, nilai hilang, serta ketidakkonsistenan, kemudian menormalkan fitur numerik seperti likes, comments, shares, dan views agar berada pada skala yang sebanding.

Data komentar teks melalui serangkaian pra-pemrosesan seperti pembersihan karakter tidak penting, pengubahan huruf menjadi bentuk seragam, pemenggalan kata, penghapusan stopword, dan stemming untuk menghasilkan representasi teks yang lebih terstruktur sebelum diubah menjadi fitur numerik, sementara itu, jika terdapat ketidakseimbangan kelas pada label keramaian maupun tema konten, dapat diterapkan teknik penanganan seperti penyesuaian bobot kelas atau sampling agar model klasifikasi nantinya tidak bias ketika mempelajari pola keramaian video kuliner

Tabel 1. Cuplikan Dataset Video Kuliner TikTok

Atrib	V1	V2	V3	V4
-------	----	----	----	----

ut				
ID	V001 _DB	V002 _DK	V003 _CK	V004 _CB
Likes	97.50 0	614.2 00	306.8 00	249. 500
Komen	613	2.268	1.258	1.01 7
Saves	2.148	12.20 0	6.639	7.49 5
Shares	8.040	27.10 0	9.915	2.80 3
K_Positive	340	1.035	522	478
K_Negatif	143	316	203	199
K_Neutral	130	917	533	331

3.4 Modeling

Pada tahap **Modeling**, data yang telah melalui proses persiapan digunakan untuk membangun model klasifikasi tingkat keramaian video kuliner TikTok. Pemodelan dilakukan dengan memanfaatkan fitur **engagement** yang meliputi *likes*, *comments_total*, *saves*, *shares*, serta fitur **sentimen komentar** (*positif*, *negatif*, dan *netral*) dan atribut tema konten yang direpresentasikan melalui variabel **is_keju**.

Pada tahap Modeling, data yang telah melalui proses persiapan digunakan untuk membangun model klasifikasi tingkat keramaian video kuliner TikTok. Pemodelan dilakukan dengan memanfaatkan fitur engagement yang meliputi *likes*, *comments_total*, *saves*, dan *shares*, serta fitur sentimen komentar (*positif*, *negatif*, dan *netral*) dan atribut tema konten yang direpresentasikan melalui variabel *is_keju*.

Pada tahap ini digunakan dua algoritma klasifikasi, yaitu **Decision Tree** dan **Random Forest**. Algoritma Decision Tree dipilih karena mampu membentuk aturan keputusan yang mudah diinterpretasikan berdasarkan fitur input (Quinlan, 1993). Sementara itu, Random Forest digunakan sebagai metode

ensemble yang mengombinasikan banyak pohon keputusan untuk meningkatkan stabilitas dan kinerja model klasifikasi (Breiman, 2001).

Dataset dibagi menjadi data latih (*training set*) dan data uji (*testing set*) untuk melatih model dan menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Pembagian data ini merupakan praktik umum dalam pengembangan model machine learning untuk menghindari overfitting dan menilai performa model secara objektif (Géron, 2019).

Proses pelatihan model dilakukan menggunakan data latih, sedangkan data uji digunakan untuk menghasilkan prediksi kelas *rame* dan *tidak rame*. Dari proses ini diperoleh nilai akurasi awal sebagai ukuran kinerja masing-masing model dalam mengklasifikasikan tingkat keramaian video kuliner TikTok.

Khusus pada model Decision Tree, tahap Modeling juga menghasilkan struktur pohon keputusan dan aturan klasifikasi (*rules*) yang merepresentasikan hubungan antar fitur dengan kelas target. Aturan keputusan ini digunakan untuk mengidentifikasi fitur-fitur yang paling berpengaruh dalam menentukan tingkat keramaian video kuliner TikTok, sehingga model tidak hanya bersifat prediktif, tetapi juga mudah diinterpretasikan (Quinlan, 1993).

3.5 Evaluation

Pada tahap Evaluation, model Decision Tree dan Random Forest yang telah dibangun dievaluasi untuk menilai kemampuannya dalam mengklasifikasikan tingkat keramaian video kuliner TikTok. Proses evaluasi dilakukan menggunakan data uji dengan menghitung metrik kinerja berupa akurasi, precision, recall, dan f1-score, yang umum digunakan dalam penelitian klasifikasi untuk menilai performa model secara menyeluruh (Powers, 2011). Evaluasi ini bertujuan untuk memastikan

bahwa model tidak hanya menghasilkan nilai akurasi yang tinggi, tetapi juga mampu mengenali masing-masing kelas secara seimbang.

Hasil evaluasi menunjukkan bahwa model Decision Tree memperoleh akurasi sebesar 75%, sedangkan model Random Forest memperoleh akurasi sebesar 50%. Perbandingan akurasi kedua model tersebut disajikan pada Tabel 2. Selain itu, gambaran struktur dan proses pengambilan keputusan pada model Decision Tree ditunjukkan pada Gambar 2, yang memperlihatkan bagaimana fitur-fitur engagement digunakan oleh model dalam menentukan klasifikasi tingkat keramaian video.

Tabel 2. Perbandingan Akurasi Model Klasifikasi

Model Klasifikasi	Akurasi (%)
Decision Tree	75x
Random Forest	50

Berdasarkan hasil evaluasi tersebut, Decision Tree dipilih sebagai model terbaik dalam memprediksi tingkat keramaian video kuliner TikTok karena menunjukkan performa yang lebih baik dibandingkan Random Forest. Selain keunggulan dari sisi akurasi, model Decision Tree juga memiliki kelebihan dari segi interpretabilitas. Aturan keputusan (*rules*) yang dihasilkan menunjukkan bahwa jumlah komentar netral (*Netral_comments*) menjadi salah satu fitur yang berperan penting dalam menentukan apakah sebuah video diklasifikasikan sebagai *rame* atau tidak *rame*. Hal ini mengindikasikan bahwa respons audiens yang bersifat netral tetap memiliki kontribusi signifikan terhadap tingkat *engagement* konten kuliner.

Pada analisis clustering menggunakan metode K-Means, evaluasi dilakukan dengan menggunakan Silhouette Score untuk menentukan jumlah cluster yang paling optimal. Nilai Silhouette Score untuk masing-masing jumlah cluster disajikan pada Tabel 3. Hasil evaluasi menunjukkan bahwa jumlah cluster dua

(K = 2) menghasilkan nilai Silhouette Score tertinggi dibandingkan jumlah cluster lainnya, sehingga dianggap paling optimal dalam merepresentasikan pola *engagement* video kuliner TikTok berdasarkan fitur numerik yang digunakan.

Tabel 3. Nilai Silhouette Score Clustering K-Means

Jumlah Cluster (K)	Silhouette Score
2	0,385
3	0,100
4	Tidak terdefinisi (NaN)

3.6 Deployment

Pada tahap Deployment, model terbaik yang diperoleh dari proses evaluasi (misalnya Random Forest) dimanfaatkan untuk menghasilkan keluaran yang relevan bagi kebutuhan praktis penyusunan konten kuliner di TikTok. Model ini dapat digunakan untuk memprediksi kategori keramaian (*rame/tidak rame*) pada video baru berdasarkan fitur engagement awal, karakteristik konten, dan tema keju atau non-keju, sehingga kreator maupun pelaku bisnis bisa menilai potensi performa sebuah konten sebelum atau sesaat setelah diunggah.

4. HASIL DAN PEMBAHASAN

4.1 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini berasal dari file Excel yang berisi hasil scraping data video kuliner TikTok dengan tema keju dan non-keju. Setiap baris merepresentasikan satu video dengan kolom ID_Video sebagai identitas unik dan Tema yang membedakan kategori konten seperti Dimsum Biasa, Dimsum Keju, Cireng Keju, dan Cireng Biasa.

Selain itu, terdapat fitur engagement berupa Likes, Comments_total, Saves, dan Shares yang menggambarkan tingkat interaksi pengguna terhadap setiap video, serta tiga kolom sentimen komentar, yaitu Positif_comments, Negatif_comments, dan

Netral_comments, yang menyatakan jumlah komentar pada masing-masing kategori sentimen. Secara keseluruhan, struktur dataset ini dirancang untuk memungkinkan analisis hubungan antara tema konten, tingkat engagement, dan pola sentimen dalam rangka mengklasifikasikan keramaian video kuliner di TikTok.

ID_Video	Tema	Likes	comments_total	saves	shares	Positif_comments	Negatif_comments	Netral_comments
0 V001 Dimsum Biasa	Dimsum Biasa	67500	613	2148	6040	340	143	130
1 V002 Dimsum Keju	Dimsum Keju	614200	2268	12200	27100	1035	316	917
2 V003 Cireng Keju	Cireng Keju	306800	1258	6630	9915	522	203	533
3 V004 Cireng Biasa	Cireng Biasa	249500	1017	7405	2803	478	199	331

Gambar 2. Tampilan Awal Dataset Video Kuliner TikTok

4.2 Statistik Deskriptif

Secara umum, keempat video kuliner pada dataset menunjukkan tingkat interaksi yang cukup tinggi. Rata-rata jumlah likes mencapai sekitar 317.000 dengan rentang 97.500 hingga 614.200, sedangkan komentar total memiliki rata-rata 1.289 dengan nilai minimum 613 dan maksimum 2.268. Angka ini menggambarkan bahwa meskipun seluruh video memiliki engagement yang baik, terdapat satu atau dua video yang jauh lebih menonjol dibanding yang lain.

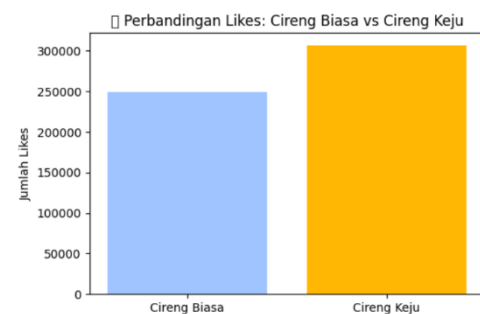
Dari sisi metrik penyimpanan dan penyebaran, nilai saves rata-rata sekitar 7.120 dengan sebaran 2.148 hingga 12.200, sedangkan shares memiliki rata-rata 11.964 dengan rentang 2.803 sampai 27.100. Perbedaan yang cukup besar antara nilai minimum dan maksimum pada kedua variabel ini mengindikasikan adanya variasi yang kuat dalam kecenderungan pengguna untuk menyimpan dan membagikan video kuliner yang dianalisis.

Pada aspek sentimen, komentar positif memiliki rata-rata sekitar 594, komentar negatif sekitar 215, dan komentar netral sekitar 478. Pola ini menunjukkan bahwa respon penonton terhadap konten kuliner dalam dataset cenderung didominasi oleh sentimen positif dan netral, sementara komentar negatif relatif lebih sedikit. Kondisi tersebut mengisyaratkan bahwa secara umum konten yang diteliti

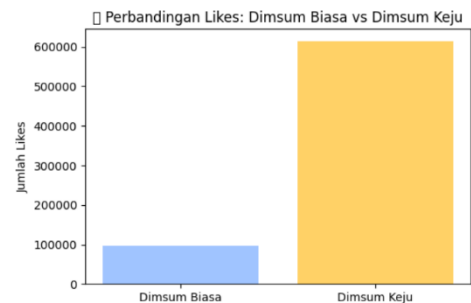
diterima dengan baik oleh audiens.

4.3 Visualisasi Data

Visualisasi data dilakukan menggunakan diagram batang untuk membandingkan jumlah likes antara konten biasa dan konten keju pada dua jenis makanan, yaitu cireng dan dimsum. Pada grafik pertama, terlihat bahwa video Cireng Keju memperoleh jumlah likes yang lebih tinggi dibanding Cireng Biasa, yang mengindikasikan bahwa penambahan keju pada menu cireng mampu meningkatkan daya tarik penonton.



Gambar 3. Perbandingan Jumlah Likes Konten Cireng Biasa dan Cireng Keju



Gambar 4. Perbandingan Jumlah Likes Konten Dimsum Biasa dan Dimsum Keju

Grafik kedua menunjukkan pola yang sama pada kategori dimsum, di mana Dimsum Keju mendapatkan likes jauh lebih banyak daripada Dimsum Biasa. Perbedaan yang cukup mencolok ini memperkuat indikasi bahwa konten kuliner bertema keju cenderung menghasilkan engagement lebih tinggi dibanding varian tanpa keju, setidaknya pada aspek jumlah likes yang diamati dalam dataset ini.

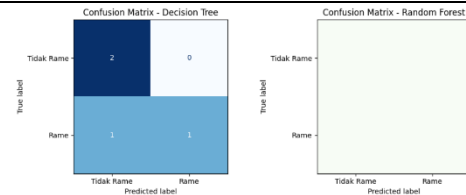
4.4 Temuan Awal

Temuan awal dari analisis deskriptif menunjukkan bahwa konten kuliner bertema keju cenderung memperoleh interaksi lebih tinggi dibandingkan varian biasa (non-keju), khususnya pada metrik likes. Hal ini tampak dari visualisasi jumlah likes, di mana Cireng Keju dan Dimsum Keju secara konsisten melampaui Cireng Biasa dan Dimsum Biasa.

Selain itu, statistik deskriptif engagement mengindikasikan adanya variasi yang cukup besar antarvideo, dengan satu atau dua video bertema keju yang menonjol sebagai outlier positif dari sisi likes, saves, dan shares. Kondisi ini mengisyaratkan bahwa kehadiran keju bukan hanya menambah variasi menu, tetapi juga berpotensi menjadi faktor pendorong utama keramaian konten kuliner di TikTok.

4.5 Hasil Modeling

Hasil pemodelan divisualisasikan menggunakan confusion matrix untuk membandingkan kinerja algoritma Decision Tree dan Random Forest dalam mengklasifikasikan kelas rame dan tidak rame. Pada model Decision Tree, dua video tidak rame berhasil diprediksi dengan benar sebagai tidak rame, satu video rame juga terklasifikasi benar, namun satu video rame lainnya masih salah diprediksi sebagai tidak rame sehingga masih terdapat satu kesalahan klasifikasi. Sementara itu, confusion matrix untuk Random Forest menunjukkan bahwa seluruh data uji ditempatkan ke dalam satu kelas yang sama, sehingga model tidak mampu membedakan video rame dan tidak rame pada skenario data yang sangat kecil ini; kondisi ini mengindikasikan bahwa, meskipun secara teoritis Random Forest lebih kuat, pada ukuran sampel yang terbatas Decision Tree justru memberikan pemisahan kelas yang relatif lebih informatif



Gambar 5. Confusion Matrix Model Decision Tree dan Random Forest

4.6 Interpretasi Model

Interpretasi model secara ringkas dapat dijelaskan sebagai berikut. Model Decision Tree berhasil mengklasifikasikan tiga dari empat video dengan benar sehingga memperoleh akurasi sebesar 75,00%, yang menunjukkan kemampuan cukup baik dalam membedakan kelas rame dan tidak rame meskipun masih terdapat satu kesalahan prediksi.

Sebaliknya, model Random Forest hanya mencapai akurasi 50,00%, yang berarti kinerjanya setara dengan tebakan acak untuk dua kelas. Hal ini sejalan dengan visualisasi confusion matrix yang memperlihatkan bahwa model cenderung mengelompokkan seluruh data uji ke satu kelas, sehingga tidak mampu memisahkan video rame dan tidak rame dengan baik pada ukuran sampel yang sangat kecil.

Dari hasil tersebut, dapat diinterpretasikan bahwa dalam konteks dataset yang terbatas, Decision Tree lebih sesuai digunakan sebagai model utama karena memberikan akurasi lebih tinggi dan pola klasifikasi yang lebih informatif, sementara Random Forest berpotensi membutuhkan jumlah data yang lebih besar agar keunggulan sebagai metode ensemble benar-benar muncul.

5. KESIMPULAN

Kesimpulan penelitian ini menunjukkan bahwa konten kuliner berbahan keju cenderung menghasilkan tingkat keramaian lebih tinggi dibandingkan varian biasa, yang tercermin dari nilai likes, saves, dan shares yang lebih besar pada video Dimsum Keju dan Cireng Keju. Temuan ini menguatkan bahwa

pemilihan bahan atau tema visual yang “menggugah selera”, seperti keju, berperan penting dalam mendorong engagement pengguna TikTok.

Dari sisi pemodelan, algoritma Decision Tree memberikan kinerja terbaik dengan akurasi 75,00%, sedangkan Random Forest hanya mencapai 50,00%, sehingga pada kondisi dataset yang kecil dan sederhana ini Decision Tree lebih reliabel untuk mengklasifikasikan video ke dalam kelas rame dan tidak rame. Secara keseluruhan, penelitian ini membuktikan bahwa kombinasi fitur engagement dan sentimen komentar dapat dimanfaatkan untuk memahami pola keramaian konten kuliner, sekaligus menjadi dasar rekomendasi bagi kreator dan pelaku bisnis dalam merancang strategi konten yang berpotensi lebih viral di TikTok.

DAFTAR PUSTAKA

- Aggarwal, C. C. (2015). *Data mining: The textbook*. Springer.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cvijikj, I. P., & Michahelles, F. (2013). Online engagement factors on Facebook brand pages. *Social Network Analysis and Mining*, 3(4), 843–861.
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
- Hu, Y., Manikonda, L., & Kambhampati, S. (2023). What makes content popular on social media? An analysis of visual and textual features. *Proceedings of the International AAAI Conference on Web and Social Media*, 17, 385–395.
- Kaye, D. B. V., Chen, X., & Zeng, J. (2022). The co-evolution of TikTok and music: Platformization, virality, and music consumption. *New Media & Society*, 24(1), 3–25.
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
- Montag, C., Yang, H., & Elhai, J. D. (2021). On the psychology of TikTok use: A first glimpse from empirical findings. *Frontiers in Public Health*, 9, 641673.
- Omar, B., & Dequan, W. (2020). Watch, share or create: The influence of personality traits and user motivation on TikTok mobile video usage. *International Journal of Interactive Mobile Technologies*, 14(4), 121–137.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining* (pp. 29–39).
- Zeng, J., & Abidin, C. (2021). #OkBoomer: Viral meme, youth subcultures, and generational politics on TikTok. *Information, Communication & Society*, 24(15), 2451–2468.
- Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.