

IMPLEMENTASI MACHINE LEARNING PADA ANALISIS TREN TIKTOK UNTUK MENDUKUNG STRATEGI PROMOSI UMKM

Arina Mana Sikana Navya Putri; Dr. Endah Sudarmilah, S.T., M.Eng.
Program Studi Teknik Informatika, Fakultas Komunikasi dan Informatika,
Universitas Muhammadiyah Surakarta

Abstrak

Media sosial TikTok telah menjadi salah satu platform utama dalam strategi pemasaran digital bagi UMKM. Namun, banyak UMKM masih kesulitan memahami tren konten dan menentukan strategi promosi yang efektif. Penelitian ini bertujuan mengimplementasikan algoritma K-Means untuk mengelompokkan konten TikTok berdasarkan kategori UMKM, serta algoritma Prophet untuk memprediksi waktu unggah terbaik. Metode penelitian meliputi pengumpulan data tren TikTok dari *Kaggle* dan *Apify TikTok scraper API*, pra-pemrosesan data, analisis menggunakan machine learning, serta visualisasi. Hasil penelitian menunjukkan bahwa menggunakan algoritma K-Means menunjukkan jumlah cluster optimal sebanyak 5 cluster dengan nilai Silhouette Score 0.49 sementara model Prophet mampu memprediksi waktu unggah terbaik yang ditunjukkan oleh nilai RMSE (Root Mean Squared Error) sebesar 0.032 setelah proses normalisasi data. Seluruh hasil analisis diimplementasikan ke dalam sebuah webdashboard berbasis Flask yang interaktif, sehingga mendukung pengambilan keputusan promosi digital berbasis data bagi UMKM.

Kata Kunci : TikTok, Machine Learning, K-Means, Prophet, Flask, Analisis Tren.

Abstract

Social media platform TikTok has become one of the main platforms in digital marketing strategies for SMEs. However, many SMEs still find it difficult to understand content trends and determine effective promotional strategies. This study aims to implement the K-Means algorithm to cluster TikTok content based on SME categories, as well as the Prophet algorithm to predict the best upload times. The research methods include collecting TikTok trend data from Kaggle and the Apify TikTok scraper API, pre-processing the data, analyzing it using machine learning, and visualizing the results. The results show that using the K-Means algorithm indicates an optimal number of 5 clusters with a Silhouette Score of 0.55, while the Prophet model is able to predict the best upload time, as indicated by an RMSE (Root Mean Squared Error) value of 0.032 after data normalization. All analysis results are implemented into an interactive Flask-based web dashboard, supporting data-driven digital promotion decision-making for SMEs.

Keywords : Tiktok, Machine Learning, K-Means, Prophet, Flask, Trend Analysis

1. PENDAHULUAN

1.1 LATAR BELAKANG

Media sosial TikTok telah mengalami pertumbuhan signifikan yang mengubah ranah pemasaran digital secara fundamental, *platform* ini telah mencapai lebih dari 1,9 miliar pengguna aktif global pada tahun 2025 (Metricool, 2025). Data menunjukkan bahwa 29% pengguna TikTok membuka aplikasi setidaknya sekali perhari, dengan rata-rata waktu penggunaan mencapai 45,3 menit perhari (Social, 2025). Hal ini menciptakan peluang emas bagi UMKM Indonesia untuk menjangkau *audiens* yang lebih luas, mengingat 74% pengguna Gen Z lebih cenderung mencari informasi lebih lanjut setelah melihat produk yang diiklankan di *platform* ini (Augustinah et al., 2024). Studi terbaru menunjukkan bahwa penggunaan konten video yang menarik di TikTok berdampak dengan peningkatan jumlah penjualan UMKM, dari program pelatihan TikTok untuk UMKM melaporkan peningkatan pendapatan antara 10-50% pada peserta yang menerapkan strategi *digital marketing* melalui *platform* ini (Mahanani et al., 2024).

Meskipun TikTok menawarkan potensi promosi yang signifikan, terdapat sejumlah tantangan dan risiko yang harus dihadapi UMKM dalam memanfaatkan *platform* ini secara optimal. Sisi positif menunjukkan bahwa TikTok Shop telah memberikan akses pasar baru melalui fitur *social commerce* yang interaktif, memungkinkan UMKM menjangkau pelanggan lebih luas dibandingkan kanal pemasaran tradisional (Vera & Raudohtul, 2024). Namun, sisi negatif permasalahan kompleks seperti risiko kebijakan yang dapat berdampak pada model bisnis UMKM, persaingan ketat dengan *platform e-commerce* besar seperti Shopee dan Lazada, serta kekhawatiran terkait keaslian produk yang dapat memengaruhi kepercayaan konsumen (Afkar & Yusmaneli, 2023). Lebih lanjut, ketergantungan pada tren viral yang tidak stabil menciptakan ketidakpastian bagi strategi pemasaran UMKM, di mana keberhasilan seringkali bergantung pada kemampuan menangkap dan merespons tren yang cepat berubah (Bryan & Alvin, 2023).

Permasalahan utama yang dihadapi UMKM dalam mengoptimalkan strategi promosi di TikTok adalah keterbatasan kemampuan analitik untuk memahami dan merespons tren secara efektif. Penelitian menunjukkan bahwa mayoritas UMKM belum memanfaatkan analitik atau tidak memiliki kapasitas untuk melakukan trend analisis, sehingga sulit menentukan waktu *posting* optimal, segmen target yang tepat, dan efektivitas konten yang diproduksi (Mesran et al., 2024). Tantangan ini diperparah oleh kompleksitas algoritma TikTok dan kebutuhan pengelolaan konten yang konsisten, yang seringkali tidak dimiliki UMKM (Susanti & Cory, 2024). Keterbatasan sumber daya untuk penganggaran iklan, optimasi kampanye, dan pengaturan *logistik delivery* juga menjadi hambatan operasional yang signifikan dalam mengoperasikan TikTok *Commerce* secara berkelanjutan (Siswanto, 2024).

1.2 TINJAUAN PUSTAKA

Media sosial telah menjadi wadah strategis bagi UMKM di Indonesia dalam Meningkatkan jangkauan pasar dan memperkuat *brand awareness* (Zakiah et al., 2024). Integrasi media sosial dengan *e-commerce* dapat meningkatkan kesadaran merek UMKM hingga 70%, menunjukkan bahwa *digital marketing* berbasis media sosial memberikan dampak nyata terhadap daya saing bisnis kecil (Samsudin, 2023). Dari dua penelitian tersebut dapat disimpulkan media sosial efektif untuk promosi. kajian akademik mengenai TikTok sebagai media utama promosi produk UMKM masih terbatas, terutama dalam konteks analisis tren berbasis data. Meskipun demikian, sebagian besar kajian lokal masih berfokus pada pemanfaatan media sosial secara deskriptif, dan belum banyak menerapkan metode *machine learning* untuk analisis tren konten yang dapat mengoptimalkan promosi UMKM.

Sejumlah penelitian menunjukkan potensi *machine learning* untuk memahami dan memprediksi tren media sosial. (Arai et al., 2024) mengembangkan metode *predictive trend analytics* dengan memanfaatkan kata kunci, *heat map*, dan sinyal interaksi media sosial (termasuk TikTok), yang terbukti mampu memperkirakan tren pasar secara *real-time*. (Ling et al., 2022) menganalisis faktor-faktor yang membuat video TikTok viral, menemukan bahwa panjang video, *framing visual*, dan interaksi pengguna berperan besar dalam popularitas konten, dan mereka berhasil membangun model klasifikasi untuk membedakan video viral dan tidak viral. Sementara itu, (Boeker & Urman, 2022) melakukan audit algoritma TikTok dan menyimpulkan bahwa fitur *follow*, *like*, serta lama menonton menjadi penentu kuat rekomendasi konten dalam *platform* tersebut. Ketiga penelitian ini menunjukkan bahwa analisis berbasis *machine learning* dapat memberikan wawasan penting mengenai pola tren media sosial, meski sebagian besar belum berfokus pada konteks UMKM di Indonesia.

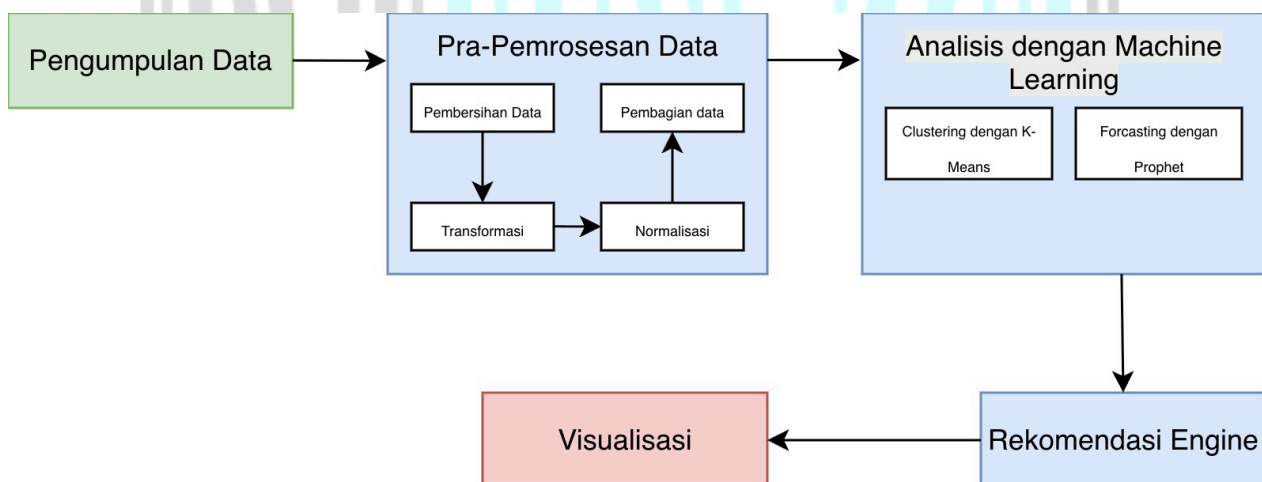
Prophet dan K-Means banyak digunakan untuk memprediksi tren dan mengelompokkan data sosial. Prophet merupakan metode *time series forecasting* yang fleksibel untuk menangani tren dengan pola musiman, kejadian khusus, serta data yang hilang. Studi perbandingan Prophet, ARIMA, dan LSTM menunjukkan bahwa Prophet mampu memberikan hasil prediksi yang cukup akurat dan interpretatif untuk *forecasting* penjualan produk, sehingga potensial untuk diterapkan dalam memprediksi popularitas hashtag atau lagu di TikTok (Suryawan et al., 2024). Sementara itu, penelitian Setiaji et al (2025) menggunakan K-Means untuk mengelompokkan konten TikTok berdasarkan jumlah *view* dan *like*, dan berhasil membedakan pola popularitas konten dengan lebih terstruktur. Hal ini memperkuat bahwa algoritma *clustering* dapat membantu memetakan kategori tren sesuai kebutuhan UMKM.

Meski sudah ada penelitian yang membahas Prophet untuk prediksi tren dan K-Means untuk

analisis konten TikTok, belum ada studi yang secara khusus menggabungkan keduanya dalam konteks promosi UMKM melalui analisis tren TikTok. Identifikasi meliputi belum adanya penerapan Prophet untuk memprediksi tren popularitas hashtag atau lagu TikTok secara *real-time*, lalu minimnya penelitian *clustering* tren TikTok dalam kaitannya dengan kategori produk UMKM, dan kurangnya pendekatan integratif yang menghubungkan hasil prediksi tren dengan rekomendasi strategi promosi. Oleh karena itu, penelitian ini akan mengimplementasikan Prophet untuk memprediksi dinamika popularitas hashtag/lagu, serta K-Means untuk mengelompokkan tren sesuai kategori UMKM. Kombinasi kedua algoritma ini diharapkan mampu memberikan kontribusi praktis berupa sistem analisis tren TikTok yang adaptif dan aplikatif bagi UMKM dalam merumuskan strategi promosi digital.

2. METODE

Diagram alir pada gambar 1. adalah tahapan yang dilakukan untuk alur kerja penelitian, mulai dari pengumpulan data hingga penyajian hasil analisis.



Gambar 1. Diagram Alir untuk Metode Penelitian

2.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan data tren konten TikTok yang diperoleh dari dua sumber utama yaitu dataset pihak ketiga dari *platform* Kaggle serta tambahan yang dikumpulkan secara mandiri menggunakan *Apify TikTok Scraper API*. Dataset Kaggle berjudul “*Video (YouTube/TikTok) Trends – EDA*” dan menggunakan dataset csv *youtube_shorts_tiktok_trends_2025.csv* disusun oleh (Tarek Masryo) dan dapat diakses secara publik. Deskripsi dataset pada platform Kaggle menyebutkan jumlah data sekitar 48k, yang merupakan nilai pendekatan. Setelah dataset diunduh dan dimuat ke dalam Google Colab, jumlah data yang terbaca secara aktual adalah 48.079 baris data. Perbedaan ini disebabkan oleh penggunaan pembulatan pada deskripsi dataset, sedangkan proses pembacaan CSV pada Colab menampilkan jumlah data secara presisi berdasarkan isi file. Oleh karena itu, jumlah data yang digunakan dalam

penelitian ini adalah 48.079 observasi. Pengambilan data dilakukan menggunakan metode pihak ketiga dan *Application Programming Interface (API)*. Dataset Apify digunakan untuk memperoleh data tren terbaru sebanyak 700 data. Periode pengambilan data pada dataset Kaggle mencakup rentang waktu tahun 2024 hingga akhir 2025, sementara data dari Apify dikumpulkan secara periodik pada saat penelitian dilakukan untuk melengkapi informasi tren terkini.

Proses validasi awal juga dijalankan agar menjamin kualitas data dalam hal akurasi dan konsistensi (Irawan et al., 2025). Pada tabel 1 merupakan nama kolom apa saja yang terdapat pada kedua dataset dan jumlahnya sebelum memasuki tahap pra-pemrosesan data.

Jenis Dataset	Jumlah Kolom	Jumlah Data	Nama Kolom
Kaggle	58	48079	platform, country, region, language, category, hashtag, title_keywords, author_handle, sound_type, music_track, week_of_year, duration_sec, views, likes, comments, shares, saves, engagement_rate, trend_label, source_hint, notes, device_type, upload_hour, genre, trend_duration_days, trend_type, engagement_velocity, dislikes, comment_ratio, share_rate,save_rate, like_dislike_ratio, publish_dayofweek, publish_period, event_season, tags, sample_comments, creator_avg_views,creator_tier,season,publish_date_approx, year_month,title,title_length,has_emoji,avg_watch_time_sec, completion_rate, device_brand, traffic_source, is_weekend, row_id,engagement_total,like_rate,dislike_rate, engagement_per_1k,engagement_like_rate, engagement_comment_rate, engagement_share_rate.
Apify	14	700	authorMeta.avatar, authorMeta.name, text, diggCount, shareCount, playCount, commentCount, collectCount, videoMeta.duration,musicMeta.musicName, musicMeta.musicAuthor,musicMeta.musicOriginal, createTimeISO, webVideoUrl.

Tabel 1. Tabel Dataset mentah dari Kaggle dan Apify

2.2 Pra-Pemrosesan Data

Pra-Pemrosesan Data dibutuhkan untuk mengurangi penyebab keterlambatan, ketidakefisienan, serta kesalahan dalam pengolahan informasi, sehingga mudah di olah untuk model machine learning (Utomo & Mustaqim, 2025). Tahapannya mencakup pembersihan data, transformasi ,

normalisasi, dan pembagian data.

2.2.a Pembersihan Data

Pembersihan data adalah proses untuk membersihkan data mulai dari menghapus data yang duplikat hingga menghapus karakter dan kolom yang tidak relevan (Dirgantara & Sudarmilah, 2024). Proses ini dilakukan untuk memastikan data yang digunakan bebas dari duplikasi, kesalahan, dan nilai yang hilang. Data yang dibersihkan berupa beberapa kolom yang tidak relevan dan tidak digunakan untuk pengolahan data. Pada gambar 3. terdapat dataset Kaggle yang digunakan terdapat total 58 kolom dan akan di ambil 11 kolom yang akan digunakan, sedangkan untuk gambar 4. menunjukan Apify punya total 14 kolom dan akan digunakan sebanyak 8 kolom.

```
# PILIH KOLOM YANG DIGUNAKAN
df_kaggle_clean = df_kaggle[[
    'hashtag', 'music_track', 'views', 'likes', 'comments',
    'shares', 'engagement_rate', 'upload_hour',
    'publish_dayofweek', 'year_month', 'week_of_year'
]].copy()
```

Gambar 3. Kode collab untuk kolom Dataset Kaggle yang akan digunakan

```
df_apify = df_apify[[
    'hashtag',
    'music_track',
    'views',
    'likes',
    'comments',
    'shares',
    'upload_time',
    'engagement_rate'
]]
```

Gambar 4. Kode collab untuk kolom yang akan digunakan untuk Apify

Kolom yang tidak digunakan pada Dataset Kaggle berjumlah 47 kolom dari total 58 atribut yang tersedia pada dataset, penelitian ini hanya menggunakan 11 atribut yang relevan dengan analisis tren dan keterlibatan pengguna. Atribut lainnya tidak digunakan karena bersifat metadata, redundan, memiliki korelasi rendah terhadap tujuan penelitian, atau tidak berpengaruh langsung terhadap proses clustering dan peramalan waktu.

Dataset apify yang berjumlah 14 kolom menghapus 6 kolom yaitu authorMeta.avatar, authorMeta.name, collectCount, videoMeta.duration, musicMeta.musicOriginal, webVideoUrl. Sama halnya dengan penghapusan dataset Kaggle. Atribut dihapus karena bersifat kurang berkontribusi untuk tahap pengolahan data. Beberapa kolom apify juga di ubah nama kolomnya agar sama dengan dataset Kaggle yang akan mempermudah proses pengolahan pada algoritma, seperti kolom text yang berisi tidak hanya hashtag tetapi juga memuat caption video yang digunakan jadi harus di bersihkan terlebih dahulu agar menghasilkan hashtag yang

digunakan saja.

2.2.b Transformasi

Transformasi data dilakukan untuk mengubah dalam bentuk yang sesuai dalam proses data mining agar lebih mudah dianalisis oleh algoritma (Wirawan & Fatmawati, 2022). Pada beberapa kolom akan diubah kedalam format yang sudah ditentukan dan diperlukan. Sebagai contohnya kolom *upload time* yang berisi waktu unggah akan diubah dalam format waktu yang dapat diolah secara komputasional. kode transformasi yang terdapat pada gambar 5. merupakan dataset Kaggle yang belum mempunyai datetime seperti Apify oleh karena itu untuk mempermudah pada machine learning beberapa tabel yang ada pada Kaggle di transformasi menjadi 1 tabel yang berformat datetime (YYYY-MM-DD HH:mm:ss). Dataset Kaggle yang semula berjumlah 11 menjadi hanya 8 kolom saja sama seperti dataset Apify. Hasil pada gambar 6(a). adalah hasil kolom *upload_time* Kaggle yang berhasil di ubah sedangkan untuk gambar 6(b). merupakan contoh hasil dataset Apify yang sudah melewati proses pembersihan data.

```
# 5. BUAT DATETIME UPLOAD_TIME
day_map = {
    "Monday": 1, "Tuesday": 2, "Wednesday": 3,
    "Thursday": 4, "Friday": 5, "Saturday": 6, "Sunday": 7
}

df_kaggle_clean["iso_day"] = df_kaggle_clean["publish_dayofweek"].map(day_map)
df_kaggle_clean["year"] = df_kaggle_clean["year_month"].str[:4].astype(int)

df_kaggle_clean["hour_fix"] = df_kaggle_clean["upload_hour"].astype(int)
df_kaggle_clean["add_day"] = (df_kaggle_clean["hour_fix"] == 24).astype(int)
df_kaggle_clean.loc[df_kaggle_clean["hour_fix"] == 24, "hour_fix"] = 0

df_kaggle_clean["date_from_week"] = (
    df_kaggle_clean["year"].astype(str)
    + "-"
    + df_kaggle_clean["week_of_year"].astype(str).str.zfill(2)
    + "-"
    + df_kaggle_clean["iso_day"].astype(str)
)

df_kaggle_clean["date_from_week"] = pd.to_datetime(
    df_kaggle_clean["date_from_week"], format="%Y-%W-%u"
)

df_kaggle_clean["upload_time"] = (
    df_kaggle_clean["date_from_week"]
    + pd.to_timedelta(df_kaggle_clean["hour_fix"], unit="h")
    + pd.to_timedelta(df_kaggle_clean["add_day"], unit="D")
)
```

Gambar 5. Kode transformasi datetime untuk dataset kaggle

hashtag	music_track	views	likes	comments	shares	upload_time	engagement_rate
#viral	8bit loop	252497	21755	3170	1977	2025-01-04 15:00:00	0.120069
#umkm	Street vibe	270580	23080	2124	2108	2025-01-01 03:00:00	0.113005
#workout	Gallery pad	7385	363	33	38	2025-01-05 15:00:00	0.068111
#esports	Neutral piano	72377	6243	811	383	2025-01-03 15:00:00	0.108156
#viral	Soft glam loop	16174	832	68	67	2025-01-04 18:00:00	0.072400

Gambar 6(a). Hasil Transformasi dataset Kaggle

hashtag	music_track	views	likes	comments	shares	upload_time	engagement_rate
#kuliner	jamskuy cari makan enak	402200	6084	95	1028	2025-11-18 12:03:54	0.017919
#chatime	kuliner	23500	126	1	10	2025-10-18 05:40:02	0.005830
#kulinerKlaten	food vlogger	13900	306	6	155	2025-11-24 11:47:22	0.033597
#kulinerbojongsede	tukang visit marjan	3214	146	9	37	2025-11-24 09:57:51	0.059739
#fyp	saramacca ingi n'a yesi	25200	606	0	112	2025-11-22 00:51:02	0.028492

Gambar 6(b). Hasil Transformasi dataset Apify

2.2.c Normalisasi

Normalisasi khusus untuk algoritma k-means adalah proses penskalaan nilai atribut dari data untuk mengubah nilai ke dalam rentang tertentu. Beberapa atribut memiliki skala nilai yang sangat berbeda, seperti view, likes, comments, dan shares dengan numerik puluhan, ratusan, ribuan berbeda dengan engagement_rate nya berupa numerik nol koma, yang dapat menyebabkan atribut dengan nilai besar mendominasi proses perhitungan jarak. jika tidak dinormalisasi model *machine learning* dapat bias terhadap fitur dengan nilai yang besar.

Proses normalisasi menggunakan metode Min-Max *Scaling*, yang mengubah semua nilai ke dalam rentang (0,1) sehingga setiap fitur memiliki kontribusi yang seimbang. Min-Max dipilih karena efektif untuk data media sosial dengan rentang nilai besar.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

hashtag	music_track	views	likes	comments	shares	engagement_rate	upload_time
#clarice	klitik kecantikan sidoarjo	0.000236	0.001425	0.000026	0.002808	1.000000	2025-06-20 18:00:24
#fashionfriends	Tor Monitor Ketua	0.001242	0.004678	0.001368	0.001420	0.394792	2025-11-19 07:02:15
#2000s	originalljud	0.006016	0.021621	0.001576	0.004294	0.361888	2025-11-18 01:09:04
#ootd	sticky fingers	0.000113	0.000428	0.000277	0.000005	0.359033	2025-11-24 09:49:50
#fyp	nissan altima	0.000883	0.003241	0.001437	0.000129	0.350171	2025-11-22 04:53:48

Gambar 7. Hasil dataset setelah proses normalisasi

Setelah proses normalisasi dilakukan akan menghasilkan dataset seperti pada gambar 7. kolom yang berisi nilai numerik seperti *view, likes, comment, shares* akan berganti nilainya menjadi rentang (0,1) sama halnya seperti *engagement_rate*.

2.2.d Pembagian Data

Pembagian data adalah pembagian menjadi 2 bagian data yaitu, data latih sebesar 80% dan data uji sebesar 20% dengan rasio 39.023 data untuk latih dan 9.756 data untuk uji . Fungsi dari pembagian data ini untuk mengukur model *machine learning* khususnya prophet agar mampu mempelajari pola dari data dan dapat memprediksi data baru yang belum pernah diolah sebelumnya (Kusuma & Widayat, 2024). Rasio ini juga yang umum digunakan dalam penelitian machine learning. Sedangkan untuk algoritma K-means tidak memerlukan pembagian data karena bersifat *Unsupervised Learning* hanya fokus pada seluruh data dan evaluasi internal (Shillhoutte).

2.3 Analisis dengan Machine Learning

Tahapan analisis merupakan inti dari penelitian dimana dataset yang telah diolah melalui proses sebelumnya selanjutnya dianalisis menggunakan algoritma *machine learning* untuk memperoleh informasi yang dapat mendukung strategi promosi UMKM di platform TikTok. Penelitian ini menggunakan dua pendekatan algoritma yaitu *clustering* dengan menggunakan algoritma K-Means dan *Forecasting*(peramalan) menggunakan model Prophet.

2.3.a Clustering dengan K-Means

K-means merupakan algoritma yang bersifat *unsupervised learning*. K-means sendiri

mempunyai fungsi untuk mengelompokkan data kedalam data *cluster* (Alhaq et al., 2025). Tujuan dari tahap ini untuk mengelompokkan video tiktok berdasarkan karakteristik dan tingkat interaksi pengguna seperti penggunaan hashtag dan lagu dihitung berdasarkan ratio jumlah engagement rate yang sudah dihitung. Hasil pengelompokan dapat menunjukkan pola tertentu, seperti menyesuaikan jenis konten promosi dengan tren yang sedang berkembang. Algoritma k-means berfungsi untuk membagi sekumpulan data ke dalam k kelompok (*cluster*) berdasarkan kemiripan nilai antar data. Proses pengelompokan dilakukan dengan mencari pusat kelompok (*centroid*) yang meminimalkan jarak antar data dikelompok yang sama menggunakan rumus matematis pada persamaan 2.

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (2)$$

Keterangan :

J : total jarak kuadrat antar data dengan centroid

k : jumlah cluster

C_i : kumpulan data dalam cluster ke- i

μ_i : pusat (centroid) dari cluster ke- i

2.3.b Forecasting dengan Prophet

Prophet merupakan salah satu algoritma *machine learning* yang digunakan untuk peramalan data deret waktu (*time-series forecasting*) dan dikembangkan oleh Facebook (Arslan, 2022). Prophet menggunakan beberapa metode khusus untuk memprediksi data deret waktu sehingga fleksibel dalam menangani data dengan pola musiman serta perubahan tren mendadak yang umum terjadi pada *platform* media sosial. Dataset prophet dibentuk menggunakan dua tabel utama yaitu ds menggunakan kolom `upload_time` dan `y` menggunakan kolom `engagement_rate`. Inisialisasi model prophet menggunakan *seasonalitas* yaitu *daily* (pola jam upload), *weekly* (pola hari dalam seminggu), *yearly* tidak digunakan karena data pendek Sebagai algoritma peramalan berbasis *additive model* yang memecah data deret waktu menjadi tiga komponen utama yang ada pada persamaan 3.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (3)$$

Keterangan :

$g(t)$: tren jangka Panjang(linear atau logistik)

$s(t)$: komponen musiman (harian, mingguan, tahunan)

$h(t)$: efek hari libur atau event khusus

ϵ_t : error atau noise

2.3.c Evaluasi Model Machine Learning

Evaluasi model merupakan tahapan penting pada penelitian yang bertujuan untuk menilai kualitas hasil terhadap data yang telah melalui seluruh tahapan sebelumnya (Abdulpatah et al., 2025). Proses evaluasi pada penelitian ini dilakukan untuk dua model utama, yaitu K-Means sebagai algoritma *clustering* dan Prophet untuk model peramalan deret waktu. Masing – masing model dievaluasi sesuai dengan karakteristik dan tujuan analisis.

2.3.c.1 Evaluasi model K-Means

Pada penelitian ini menggunakan metrik evaluasi *Silhouette Score* yang berfungsi untuk mengukur seberapa mirip suatu objek dalam clusternya sendiri dibandingkan dengan cluster lain. Skor memiliki rentang dari -1 hingga 1. Semakin mendekati 1, maka data point dianggap telah terklasterisasi dengan baik dan terpisah dengan jauh dari data dari cluster lain. Skor di atas 0,5 umumnya menunjukkan bahwa kluster terdefinisi dengan cukup baik (Mafela et al., 2025).

2.3.c.2 Evaluasi model Prophet

Akurasi merujuk pada seberapa baik atau ketepatan suatu model digunakan dalam peramalan. Akurasi pada hasil prediksi menggunakan RMSE (*Mean Absolute Percent Error*) digunakan untuk menghitung rata-rata selisih kuadrat nilai prediksi dan nilai aktual seperti pada persamaan 4 (Simamora et al., 2024).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

Dengan :

n : jumlah data pengamatan

y_i : nilai aktual pada data ke-iii

$\hat{y}_i$: nilai hasil prediksi pada data ke-iii

$(y_i - \hat{y}_i)^2$: kuadrat selisih antara nilai aktual dan prediksi

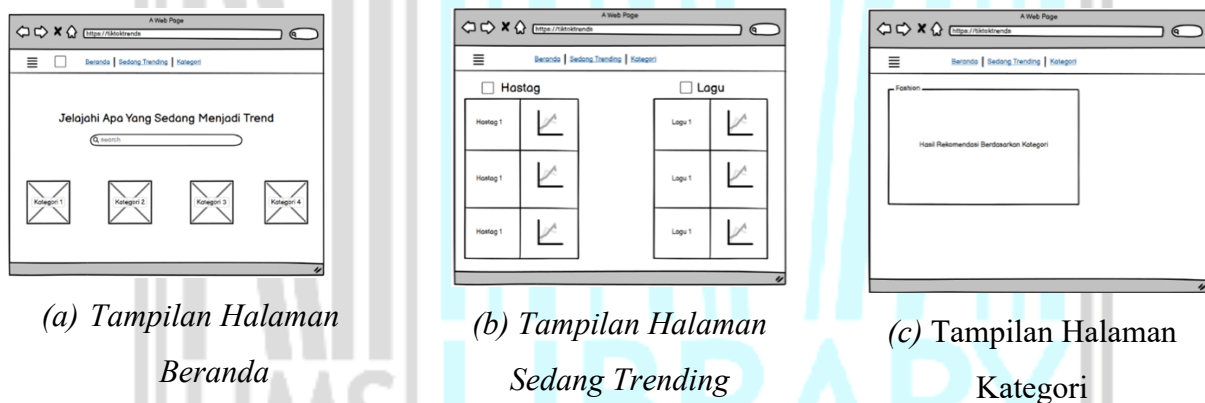
2.4 Rekomendasi Engine

Tahapan ini bertujuan untuk membangun sistem yang dapat memberikan rekomendasi hastag, lagu, dan waktu unggah video tiktok berdasarkan kategori UMKM. Sistem ini dikembangkan sebagai hasil akhir dari proses analisis *machine learning* yang sebelumnya telah dilakukan menggunakan algoritma K-Means dan Prophet. Dengan adanya sistem rekomendasi ini, pelaku UMKM dapat memperoleh saran strategi promosi digital yang terpersonalisasi, sesuai dengan pola tren konten tiktok pada masing-masing kategori usaha.

2.5 Visualisasi

Bagian ini merupakan perancangan antarmuka yang divisualisasikan dalam bentuk wireframe.

Pengembangan system analisis berbasis web *dashboard* interaktif yang dilengkapi dengan proses analisis data dapat mengubah data mentah menjadi informasi yang mudah dipahami dan digunakan sebagai dasar evaluasi serta pengambilan Keputusan (Rakhmadi & Ariyanto, 2021). *Dashboard* ini dirancang menggunakan *framework Flask (Python)* sebagai *backend* dan HTML/CSS untuk visualisasi grafik. Tujuannya agar hasil analisis seperti pengelompokan tren, prediksi tren, dan rekomendasi konten dapat tersaji secara jelas dan menarik. Visualisasi *wireframe* di Gambar 2(a). Berupa tampilan beranda yaitu halaman utama yang memuat fitur *searching* untuk langsung mencari kategori yang diperlukan secara cepat. Pada halaman ini juga memuat beberapa kategori yang sedang menjadi tren. Halaman sedang trending berisi grafik menurut hastag dan lagu yang sedang banyak digunakan dalam kurun waktu tertentu seperti yang di visualisasikan pada Gambar 2(b). Halaman yang memuat rekomendasi *engine* terdapat pada halaman kategori pada gambar 2(c). yang akan menampilkan hastag yang disarankan, lagu yang populer, waktu unggah yang terbaik, berdasarkan kategori yang dipilih.



Gambar 8. Wireframe Dashboard Website

3. HASIL DAN PEMBAHASAN

Analisis data sudah berhasil dilakukan dengan menggunakan dua algoritma *machine learning*, yaitu *k-means* dan *prophet* yang akan dibahas pada tahap ini. Hasil analisis di implementasikan ke dalam sebuah *webdashboard* interaktif yang disesuaikan dengan kebutuhan pengguna.

3.1 Hasil dan Evaluasi Clustering dengan Algoritma K-Means

Penentuan jumlah cluster dilakukan dengan mengevaluasi nilai *Silhouette Score* pada rentang $k = 2$ hingga $k = 8$. Berdasarkan hasil evaluasi, diperoleh nilai *Silhouette Score* tertinggi sebesar 0.55 pada $k = 2$, tetapi dalam penelitian ini memilih jumlah cluster $k = 5$ meskipun skor yang dihasilkan sedikit yaitu 0.49 lebih rendah dibandingkan $k = 2$. Pemilihan ini dilakukan karena 5 cluster mampu memberikan segmentasi performa konten yang lebih detak dan mudah diinterpretasikan,

sehingga lebih relevan untuk kebutuhan analisis tren dan pengambilan keputusan. Oleh karena itu, penelitian ini menggunakan *cluster* 5 sebagai jumlah *cluster* optimal yang ditunjukkan oleh Tabel 1. Nilai ini menandakan bahwa struktur *cluster* telah terbentuk dengan cukup baik dan pemisahan antar *cluster* relatif jelas.

Tabel 1. Hasil Perhitungan *SilhouetteScore* berdasarkan Jumlah *Cluster*

Jumlah <i>Cluster</i>	Hasil <i>Silhouette Score</i>
2	0.548259
3	0.492531
4	0.497583
5	0.499941
6	0.474336
7	0.455832
8	0.435582

hashtag	music_track	views	likes	comments	shares	engagement_rate	upload_time
#clarice	klirik kecantikan sidoarjo	26500	12400	3	9548	0.828340	2025-06-20 18:00:24
#fashionfriends	Tor Monitor Ketua	139700	40700	158	4827	0.327022	2025-11-19 07:02:15
#2000s	originalljud	676800	188100	182	14600	0.299767	2025-11-18 01:09:04
#ootd	sticky fingers	12700	3727	32	18	0.297402	2025-11-24 09:49:50
#fyp	nissan altima	99300	28200	166	437	0.290060	2025-11-22 04:53:48

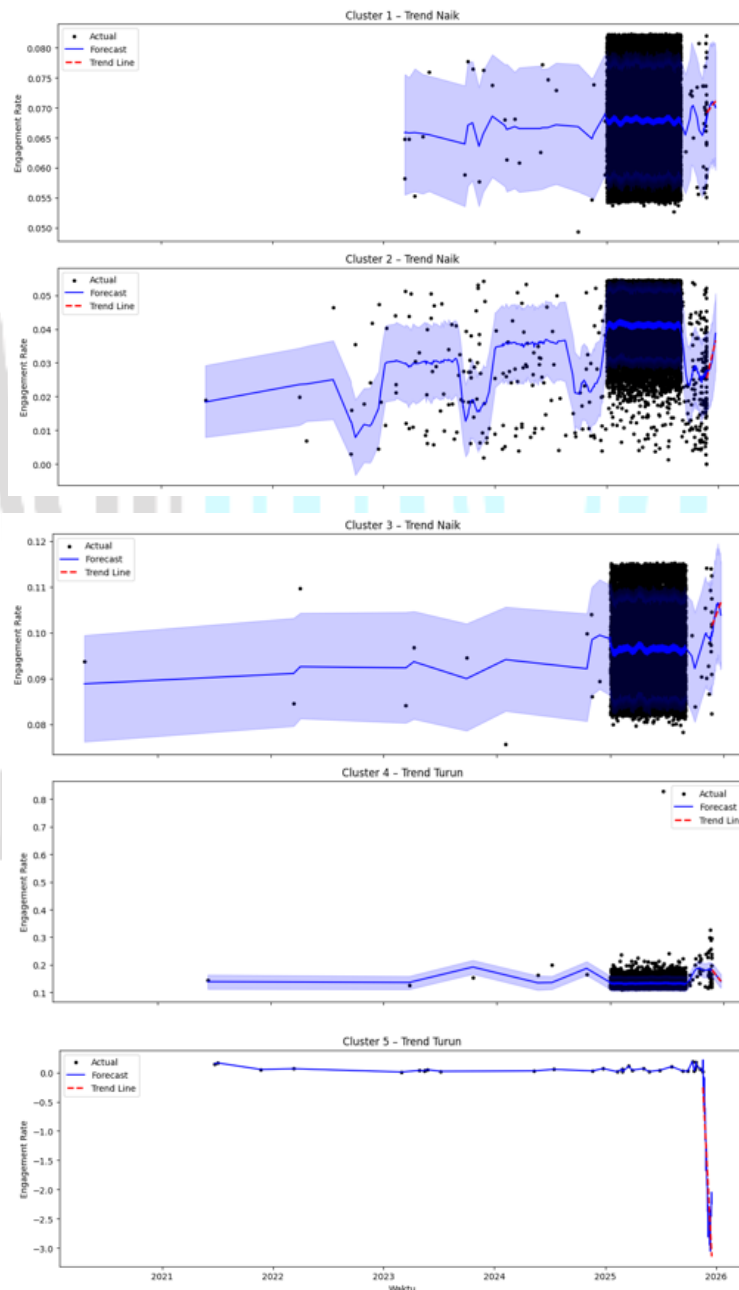
Gambar 9. Hasil dari pembagian data setiap *clusternya* dari terbesar ke terkecil

Sedangkan untuk penentuan kelompok pada *cluster* yang sudah terbentuk ditentukan dengan mempertimbangkan karakteristik data dan hasil evaluasi kualitas *clustering*. Pada Tabel 9 menunjukkan pembagian setiap *cluster* berdasarkan *cluster* tertinggi ke terendah menurut dari jumlah *engagement_rate* nya. Setiap *cluster* menunjukkan pola tingkat interaksi pengguna yang berbeda-beda, sehingga dapat merepresentasikan tingkat popularitas konten TikTok secara lebih terstruktur. Nilai *engagement_rate* yang tinggi meskipun jumlah penayangan (*views*) relatif kecil disebabkan oleh sifat metrik *engagement_rate* yang berbentuk rasio. *Engagement rate* mengukur proporsi interaksi pengguna terhadap jumlah penayangan, sehingga konten dengan jangkauan terbatas namun tingkat interaksi yang tinggi dapat menghasilkan nilai *engagement_rate* yang besar. Kondisi ini menunjukkan bahwa konten tersebut memiliki daya tarik yang kuat bagi audiens yang melihatnya, meskipun belum menjangkau pengguna dalam jumlah besar.

3.2 Hasil dan Evaluasi Prediksi dengan Algoritma Prophet

Berdasarkan hasil evaluasi kinerja model, diperoleh nilai *Mean Absolute Error (MAE)* sebesar 0,025 dan *Root Mean Squared Error (RMSE)* sebesar 0,032 Pada Gambar 3. Mengingat data yang digunakan telah melalui proses normalisasi, nilai error yang berada di bawah 0,05 menunjukkan

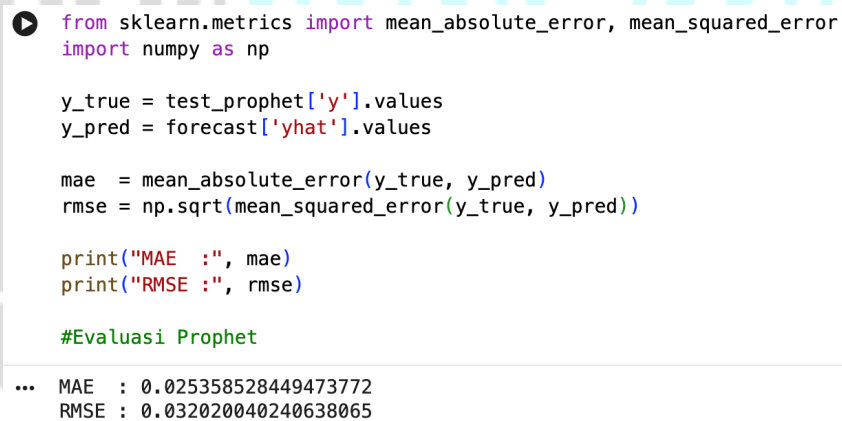
bahwa model memiliki tingkat kesalahan yang sangat rendah dan mampu menghasilkan prediksi yang akurat. Grafik hasil peramalan menggunakan algoritma Prophet yang ada pada gambar 10. menampilkan pola perubahan engagement rate berdasarkan lima cluster hasil pengelompokan K-Means. Pada setiap grafik, titik hitam merepresentasikan data aktual, garis biru menunjukkan hasil prediksi Prophet, sedangkan garis merah putus-putus menunjukkan arah tren secara umum.



Gambar 10. Hasil grafik prediksi ke-5 cluster menggunakan prophet

Berdasarkan hasil peramalan menggunakan algoritma Prophet pada lima cluster, terlihat bahwa setiap cluster memiliki karakteristik tren engagement rate yang berbeda. Cluster 1 menunjukkan tren kenaikan yang stabil dengan fluktuasi rendah, menandakan konten yang konsisten dan cocok digunakan sebagai strategi promosi jangka panjang. Cluster 2 juga memperlihatkan tren kenaikan, namun dengan fluktuasi yang lebih tinggi, yang mengindikasikan potensi viral namun belum

sepenuhnya stabil sehingga sesuai untuk eksperimen variasi konten. Cluster 3 memiliki tingkat engagement yang relatif tinggi sejak awal dan tetap menunjukkan tren meningkat, sehingga dapat dikategorikan sebagai konten berkinerja tinggi yang sangat direkomendasikan untuk promosi utama UMKM. Sebaliknya, Cluster 4 menunjukkan tren penurunan engagement rate meskipun sempat mengalami kenaikan pada beberapa periode, yang mengindikasikan penurunan minat audiens dan perlunya evaluasi terhadap konsep konten. Cluster 5 mengalami penurunan engagement yang paling signifikan dan tidak stabil, menandakan performa konten yang rendah sehingga tidak direkomendasikan untuk dijadikan acuan strategi promosi. Secara keseluruhan, hasil peramalan Prophet memberikan gambaran arah tren yang dapat dimanfaatkan oleh UMKM untuk menentukan apakah suatu pola konten perlu dipertahankan, dioptimalkan, atau dievaluasi kembali dalam strategi pemasaran digital.



```

from sklearn.metrics import mean_absolute_error, mean_squared_error
import numpy as np

y_true = test_prophet['y'].values
y_pred = forecast['yhat'].values

mae = mean_absolute_error(y_true, y_pred)
rmse = np.sqrt(mean_squared_error(y_true, y_pred))

print("MAE :", mae)
print("RMSE :", rmse)

#Evaluasi Prophet
... MAE : 0.025358528449473772
    RMSE : 0.032020040240638065

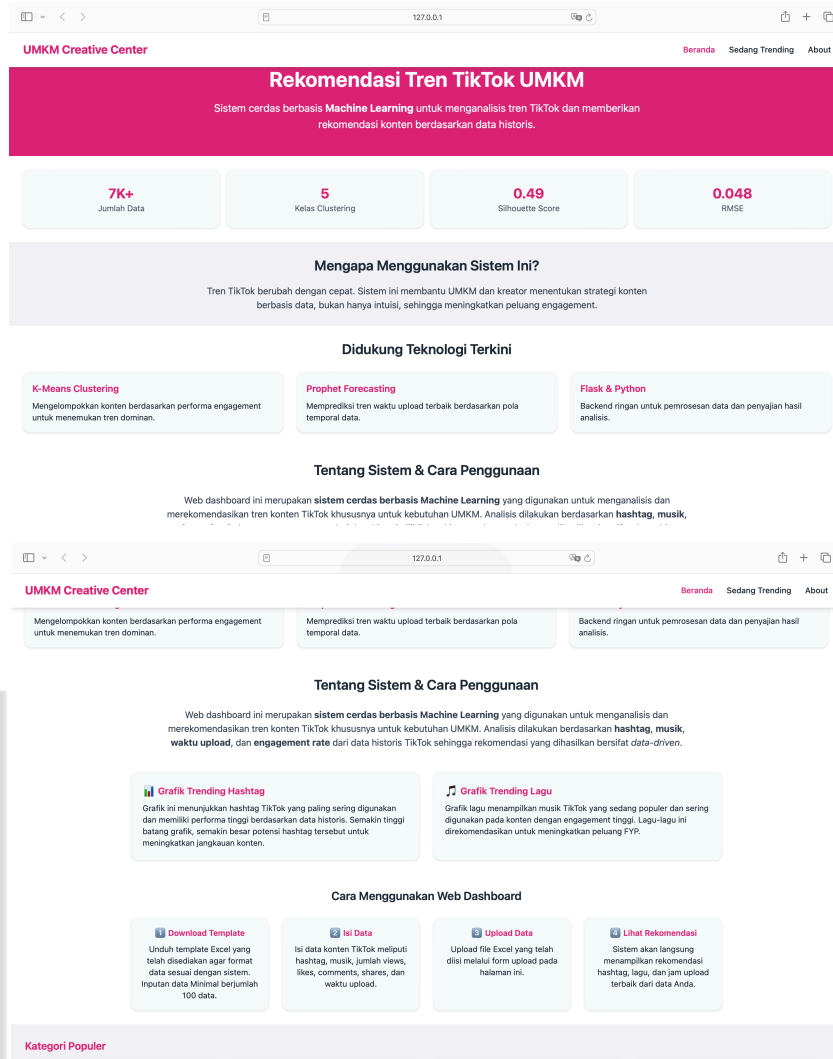
```

Gambar 11. Evaluasi Model Prophet dengan MAE dan RMSE

Berdasarkan hasil evaluasi kinerja model, diperoleh nilai *Mean Absolute Error (MAE)* sebesar 0,025 dan *Root Mean Squared Error (RMSE)* sebesar 0,032 Pada Gambar 3. Mengingat data yang digunakan telah melalui proses normalisasi, nilai error yang berada di bawah 0,05 menunjukkan bahwa model memiliki tingkat kesalahan yang sangat rendah dan mampu menghasilkan prediksi yang akurat.

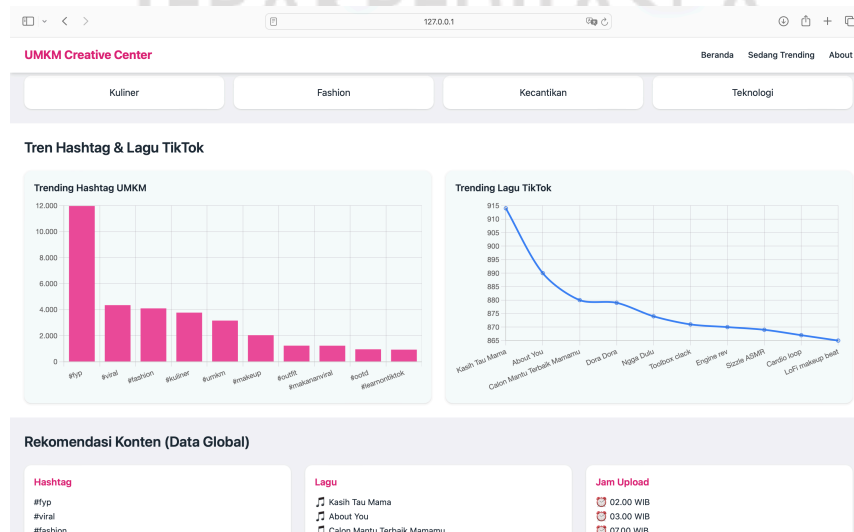
3.2 HASIL VISUALISASI WEBDASHBOARD

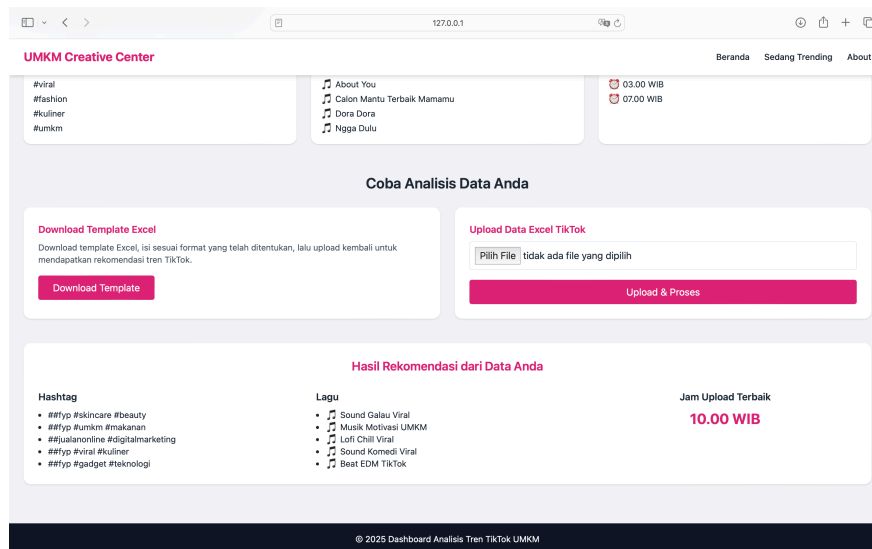
Hasil proses *clustering* dan *forecasting* digunakan untuk mengembangkan halaman website untuk membantu mengambil Keputusan strategi promosi digital pada platform TikTok. *Webdashboard* dirancang untuk menyajikan informasi hasil analisis *machine learning* secara ringkas, informatif, dan mudah dipahami.



Gambar 5. Tampilan Antarmuka Webdashboard

Halaman beranda yang ada pada gambar 5. menunjukkan antarmuka *webdashboard* yang berisi penjelasan dan pengenalan singkat tentang sistemnya serta tata cara penggunaan *webdashboard*nya. Tujuan pada pemberian deskripsi agar memudahkan pengguna untuk menggunakan *webdashboard*.





Gambar 6. Halaman Rekomendasi Konten dari Data yang Sudah Diolah

Dalam penelitian ini, *dashboard* yang disajikan pada gambar 6 digunakan untuk memfasilitasi pengguna untuk analisis peramalan dan rekomendasi hashtag, lagu, dan jam unggah dengan penggunaan algoritma *k-means* dan *prophet*. *Dashboard* dirancang dengan cermat untuk memberikan gambaran secara global untuk *hashtag*, lagu dan jam unggah terbaik yang memungkinkan membantu pengguna dalam bentuk sistem pendukung Keputusan yang aplikatif.

4. PENUTUP

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma *K-Means* dan *Prophet* mampu memberikan kontribusi yang signifikan dalam analisis tren TikTok untuk mendukung strategi promosi UMKM. Algoritma *K-Means* berhasil mengelompokkan konten TikTok berdasarkan tingkat interaksi pengguna sehingga mampu membedakan konten dengan potensi viral dan non-viral secara terstruktur. Hasil evaluasi menggunakan *Silhouette Score* menunjukkan bahwa model *clustering* memiliki kualitas pemisahan *cluster* yang cukup baik. Sementara itu, algoritma *Prophet* mampu menghasilkan prediksi tren popularitas *hashtag* dan lagu TikTok dengan tingkat akurasi yang tinggi. Nilai MAE dan RMSE yang rendah, khususnya setelah data melalui proses normalisasi, menunjukkan bahwa model peramalan yang dibangun memiliki tingkat kesalahan yang kecil dan layak digunakan sebagai dasar pengambilan keputusan. Hasil peramalan ini memungkinkan UMKM untuk melakukan perencanaan konten secara lebih proaktif dan berbasis data. Seluruh hasil analisis kemudian diimplementasikan ke dalam *webdashboard* interaktif yang dirancang untuk memvisualisasikan hasil *clustering*, *forecasting*, serta rekomendasi *hashtag*, lagu, dan waktu unggah konten. *Dashboard* ini diharapkan dapat membantu pelaku UMKM dalam merumuskan strategi promosi digital yang lebih efektif, adaptif terhadap tren, dan sesuai dengan karakteristik algoritma TikTok.

DAFTAR PUSTAKA

- Abdulpatah, T., Sari, B. N., & Susilawati. (2025). *Algoritma K-Means dan Agglomerative Hierarchical Clustering Untuk Pengelompokan Daerah Penghasil Padi*. 13(3).
- Afkar, M. A., & Yusmaneli. (2023). Model Bisnis TikTok Shop dan Implikasinya terhadap Ekonomi Digital di Indonesia. *Journal of Law and Economics*, 2(1), 41–51. <https://doi.org/10.56347/jle.v2i1.180>
- Alhaq, H., Yanto, W., Ichsandi, Sari, R. S., SHolid, R. G., & Septiana, W. A. (2025). *Analisis Dan Perancangan Clustering Siswa Baru Menggunakan Metode K-Means Pada SMK Negeri 1 Sorolangun*. 4(2).
- Arai, K., Fujikawa, I., Nakagawa, Y., & Ogawa, S. (2024). Method for Predictive Trend Analytics with SNS Information for Marketing. *International Journal of Advanced Computer Science and Applications*, 15(2), 419–425. <https://doi.org/10.14569/IJACSA.2024.0150245>
- Arslan, S. (2022). A hybrid forecasting model using LSTM and Prophet for energy consumption with decomposition of time series data. *PeerJ Computer Science*, 8(2020). <https://doi.org/10.7717/PEERJ-CS.1001>
- Augustinah, F., Herawati, A., Oktaviana, S., Listyawati, L., & Sri Lestari, D. (2024). The Use of Video Content on the Tiktok Platform as a Digital Promotion Medium to Increase Sales Volume for Msmes in Surabaya City. *International Journal of Economics, Business and Management Research*, 08(11), 248–268. <https://doi.org/10.51505/ijebmr.2024.81117>
- Boeker, M., & Urman, A. (2022). An Empirical Investigation of Personalization Factors on TikTok. In *WWW 2022 - Proceedings of the ACM Web Conference 2022* (Vol. 1, Issue 1). Association for Computing Machinery. <https://doi.org/10.1145/3485447.3512102>
- Bryan, N., & Alvin, S. (2023). Harnessing TikTok's Marketing Potential: A Case Study of @KUCA.OFFICIAL's Communication Strategy in Inspiring Consumer Purchase Intentions. *Indonesian Journal of Business Analytics*, 3(5), 1443–1456. <https://doi.org/10.55927/ijba.v3i5.5329>
- Dirgantara, I., & Sudarmilah, E. (2024). *Analisis Sentimen Terhadap Isu Islamfobia Pada Platform Twitter Menggunakan Metode Klasifikasi Naive Bayes*. 1–13.
- Irawan, M. A., Juwita, A. R., Awal, E. E., Rohana, T., Informatika, T., Komputer, F. I., Buana, U., & Karawang, P. (2025). *Perbandingan Algoritma Generalized Linear Model Dan Linear Regression Untuk Prediksi Hujan Berbasis Data Kaggle Indonesia Comparison of Generalized Linear Model Algorithms and Linear Regression for Rain Prediction*. 5(9), 2478–2489.
- Kusuma, Z. A., & Widayat, W. (2024). Analisis Sentimen Pengguna Twitter Terhadap Program Makan Siang Gratis Menggunakan Metode Naive Bayes. *Educacao e Sociedade*, 1(1), 15.
- Ling, C., Blackburn, J., De Cristofaro, E., & Stringhini, G. (2022). Slapping Cats, Bopping Heads, and Oreo Shakes: Understanding Indicators of Virality in TikTok Short Videos. *ACM International Conference Proceeding Series*, 164–173. <https://doi.org/10.1145/3501247.3531551>
- Mafela, G., Sujak, M., Rofiq, H. N., & Tawakal, F. I. (2025). *Implementation of K-Means Clustering for Optimizing Non-Communicable Disease Budgets Implementasi K-Means Clustering untuk Optimalisasi Anggaran Penyakit Tidak Menular*. 5(January), 67–74.
- Mahanani, E., Sari, B., & Thantawi, A. M. (2024). Workshop TIKTOK Senjata Rahasia Tingkatkan Omset UMKM Binaan GEMMA Indonesia Raya, Depok. *Ikra-Ith Abdimas*, 8(1), 283–293. <https://doi.org/10.37817/ikra-ithabdimas.v8i1.3203>

- Mesran, M., Suginam, S., Joli Afriany, & Dwika Assrani. (2024). Optimalisasi Peningkatan Penjualan Menggunakan Media Sosial Tiktok. *JPM: Jurnal Pengabdian Masyarakat*, 4(4), 395–401. <https://doi.org/10.47065/jpm.v4i4.1531>
- Metricool. (2025). *50 TikTok Statistics in 2025 for Social Media Marketing*.
- Rakhmadi, A., & Ariyanto, R. (2021). *Measurement Motoric System of Cerebral Palsy Disability using Gross Motor Function Measure (GMFM)*. 32–37.
- Samsudin. (2023). *The Role of Social Media and E-Commerce in Accelerating the Growth of Indonesian MSMEs*. 3(1).
- Setiaji, G. G., Gunata, K. P., & Setiarso, G. (2025). Optimasi Clustering K-Means Menggunakan Algoritma Genetika Dengan Data View Dan Like Di Tiktok. *Jurnal Transformatika*, 22(2), 115–120. <https://doi.org/10.26623/y2tedy77>
- Simamora, J. P., Martha, S., Andani, W., Statistika, P. S., & Tanjungpura, U. (2024). *Penerapan Algoritma Prophet Pada Peramalan Harga Saham*. 3(2), 94–101.
- Siswanto. (2024). Berinovasi dengan TikTok : UMKM Lokal dan Langkah-Langkah Menghadapi Dinamika E-Commerce di Tahun 2024 Siswanto tekanan baru bagi UMKM lokal untuk tetap relevan dan bersaing di pasar yang semakin Ekonomi digital hingga saat ini sangat berkembang meluas di. *Seminar Nasional Prosiding Ilmu Manajemen Kewirausahaan Dan Bisnis*, 1(1).
- Social, S. (2025). *28 TikTok Statistics Marketers Need To Know in 2025*.
- Suryawan, I. G. T., Putra, I. K. N., Meliana, P. M., & Sudipa, I. G. I. (2024). Performance Comparison of ARIMA, LSTM, and Prophet Methods in Sales Forecasting. *Sinkron*, 8(4), 2410–2421. <https://doi.org/10.33395/sinkron.v8i4.14057>
- Susanti, & Cory, V. (2024). *Strategi Pemasaran UMKM di Era Digital: Studi Kasus Gamis Hijab Terbaru Riskirich Menggunakan Aplikasi TikTok Shop*. 3(1), 4785–4797.
- Utomo, I. C., & Mustaqim, A. (2025). *Rancang Bangun Sistem Laporan Keuangan Berbasis Web pada UMKM Jago Sore di Boyolali*.
- Vera, M., & Raudohatul, J. (2024). Analisis Peran Tiktok Shop dalam Mendorong Ekonomi UMKM di Indonesia. *Sammajiva: Jurnal Penelitian Bisnis Dan Manajemen*, 2(2), 138–146. <https://doi.org/10.47861/sammajiva.v2i2.992>
- Wirawan, R., & Fatmawati, S. T. A. (2022). *Analisa Terhadap Tingkat Keaktifan Siswa pada Kegiatan Pembelajaran Sekolah dengan Metode Naive Bayes*.
- Zakiyah, A. R., Rahmanto, A. N., & Sudarmo, S. (2024). Trends in using Social Media as UMKM Marketing. *Proceeding ISETH (International Summit on Science, Technology, and Humanity)*, 1942–1949. <https://doi.org/10.23917/iseth.4611>