

BAB II TINJAUAN PUSTAKA

A. Penelitian Terkait

Tabel 1 Penelitian Terkait

No	Nama (Tahun)	Judul
(1)	(2)	(3)
1	B. Nugroho & Z. Ma'arif(2022)[3]	Penerapan Data Mining Metode Klasifikasi Untuk Menganalisa Penyalahgunaan Sosial Media.
2	D. Sinambela & H. Naparin(2023)[4]	Implementasi Algoritma Decision Tree dan Random Forest dalam Prediksi Perdarahan Pascasalin
3	O. Pratama & J. Haerul(2023)[5]	Penerapan Data Mining Menggunakan Algoritma Apriori Untuk Mengetahui Kebiasaan Konsumen Dan Prediksi Stok Produk
4	M. Qishtiano & P. Prayesy(2023)[7]	Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa
5	B. Wicaksono & N. Cahyono(2024)[8]	Analisis Sentimen Komentar Instagram Pada Program Kampus Merdeka Dengan Algoritma Naive Bayes Dan Decision Tree
6	I, Alfitrah, D. Hardiyanti, & A. Meiriza (2026)[9]	Klasifikasi Keputusan Pembelian Skincare Di TikTok Shop Berdasarkan Faktor Influencer Menggunakan Algoritma Decision Tree
7	T. Alifah, D. Wahdiyati, & N. Rahman (2025)[10]	Analisis Engagement Rate Pada Konten Video di Akun TikTok Grup Idol Virtual Plave @plave official
8	B. kristianto (2024)[11]	Analisis Sentimen Terhadap TikTok Shop dengan K-Nearest Neighbor, Decision Tree, dan Naive Bayes.

9	N, Sapina, A. Nanda, & M, Arifin (2025)[12]	Analisis Faktor – Faktor Yang Mempengaruhi <i>Engagement</i> Video di Platform TikTok Menggunakan <i>Multiple Linear Regression</i>
10	N, Sapina & A. Nanda (2025)[12]	Analisis Faktor-Faktor yang Mempengaruhi <i>Engagement</i> Video di Platform TikTok Menggunakan Multiple Linear Regression.

Berdasarkan tinjauan terhadap sejumlah penelitian sebelumnya yang telah disajikan dalam bentuk tabel, dapat disimpulkan bahwa belum ditemukan kajian yang secara spesifik menerapkan algoritma *decision tree* untuk keperluan regresi atau prediksi pada platform media sosial TikTok. Melihat adanya celah penelitian tersebut, peneliti terdorong untuk melakukan studi lebih lanjut mengenai pemanfaatan algoritma *decision tree* dalam konteks regresi, yaitu untuk memprediksi nilai numerik tertentu. Dalam penelitian ini, fokus utama adalah pada prediksi popularitas video affiliate yang diunggah ke TikTok. Popularitas tersebut diukur menggunakan indikator numerik seperti jumlah tayangan (*views*), jumlah *like*, komentar, atau interaksi lainnya. Dengan kata lain, penelitian ini bertujuan mengeksplorasi bagaimana algoritma *decision tree* dapat digunakan untuk membangun model prediksi yang mampu memperkirakan tingkat popularitas sebuah video berdasarkan atribut-atribut tertentu dari konten tersebut, seperti waktu unggah, durasi video, dan sebagainya.

B. Landasan Teori

Penelitian ini memerlukan dasar teori yang kuat untuk mendukung proses analisis, perancangan, dan implementasi sistem prediksi popularitas video pada platform TikTok menggunakan teknik data mining. Landasan teori berfungsi sebagai acuan konseptual dan teknis dalam menyusun kerangka berpikir serta dalam menjelaskan proses-proses yang terlibat di setiap tahapan penelitian.

1. Konsep Popularitas Konten

Popularitas merupakan ukuran yang menunjukkan tingkat penerimaan, pengakuan, atau ketertarikan yang diterima oleh suatu objek dari masyarakat atau *audiens* luas[13]. Dalam konteks digital, khususnya konten berbasis media, popularitas mencerminkan jumlah perhatian dan keterlibatan yang diterima oleh suatu konten dalam jangka waktu tertentu. Popularitas bukan hanya soal banyaknya yang melihat, tetapi juga tentang seberapa tinggi interaksi atau reaksi yang diberikan pengguna terhadap konten tersebut[14]. Secara umum, popularitas konten digital dapat dipahami melalui dua pendekatan utama: jangkauan (*reach*) dan keterlibatan (*engagement*)[13]. *Reach* mengukur seberapa banyak konten tersebut dilihat atau tersebar, sedangkan *engagement* mengukur interaksi yang terjadi antara pengguna

dengan konten, seperti memberi tanggapan, berbagi, atau memberikan komentar. Kedua aspek ini saling melengkapi dalam memberikan gambaran tentang seberapa menarik atau relevan suatu konten bagi audiensnya[13].

Faktor-faktor yang memengaruhi popularitas konten mencakup kualitas isi, konteks sosial, waktu publikasi, serta teknik penyajian konten seperti penggunaan judul menarik, media visual, dan narasi yang persuasif[14].

Sementara itu, pengukuran popularitas yang bersifat kuantitatif dapat dilakukan melalui penghitungan nilai *engagement rate*, yaitu rasio antara total interaksi pengguna dengan jumlah jangkauan konten tersebut. Pendekatan ini menjadikan popularitas tidak hanya sebagai sesuatu yang bersifat viral, tetapi sebagai metrik yang bisa diukur, dianalisis, dan diprediksi[14].

Dalam konteks penelitian, konsep popularitas menjadi penting karena digunakan sebagai acuan dalam menyusun model prediktif. Kemampuan sistem untuk mengestimasi tingkat popularitas dari suatu konten secara kuantitatif dapat membantu dalam pengambilan keputusan, optimasi strategi distribusi konten, hingga perencanaan kampanye pemasaran digital. Dengan kata lain, popularitas bukan hanya sekadar konsekuensi dari interaksi sosial di dunia maya, melainkan juga dapat menjadi objek analisis ilmiah berbasis data[15].

2. Platform TikTok

TikTok adalah platform media sosial berbasis video pendek yang memungkinkan pengguna untuk membuat, membagikan, dan menemukan konten video dengan durasi singkat [16].

Menurut situs resmi TikTok, misi utama platform ini adalah “*to inspire creativity and bring joy*”, dengan memberikan ruang bagi pengguna untuk mengekspresikan ide, hiburan, dan opini dalam bentuk video[16]. Pada awal tahun 2018, TikTok telah menjadi salah satu aplikasi yang paling banyak diunduh di dunia[16].

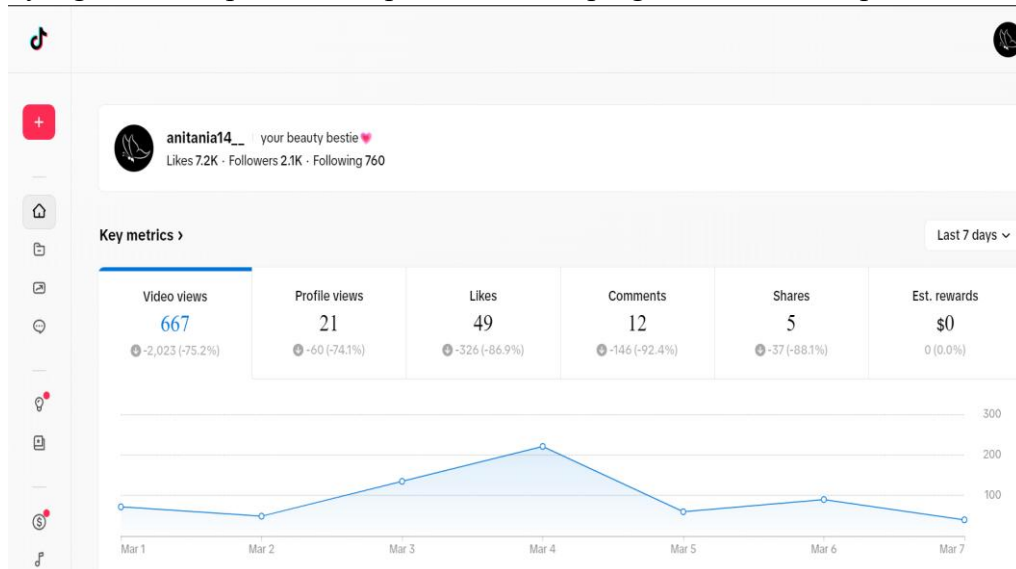
Salah satu kekuatan utama TikTok adalah algoritma rekomendasinya yang secara otomatis menyuguhkan video pada pengguna berdasarkan preferensi dan interaksi sebelumnya. Hal ini membuat kemunculan konten pada halaman “*For You Page*” (FYP) menjadi target utama banyak kreator[17]. Keberhasilan sebuah video diukur melalui beberapa metrik performa yang tersedia dalam TikTok *Creator Tools*, yaitu[17]:

- a. *Views* : menunjukkan jumlah total tayangan yang diterima video.
- b. *Likes* : menunjukkan jumlah pengguna yang menyukai video tersebut.
- c. *Comments* : menunjukkan jumlah komentar yang diberikan oleh pengguna lain.
- d. *Shares* : menunjukkan seberapa sering video tersebut dibagikan ke pengguna lain atau platform eksternal.

Selain itu, kreator juga memiliki akses ke data tambahan seperti durasi rata-rata tonton, serta jam tayang, yang semuanya dapat dianalisis untuk memahami

performa konten secara lebih mendalam. Data-data inilah yang menjadi acuan utama dalam penelitian ini untuk membangun model prediksi popularitas konten berdasarkan perilaku historis video.

Dalam penelitian ini, konten video affiliate yang digunakan sebagai objek penelitian bersumber dari akun TikTok pribadi peneliti yaitu @anitanial4_, yang aktif mempromosikan produk melalui program TikTok Shop Affiliate.



Gambar 1 Akun TikTok Penelitian

3. *Affiliate*

Affiliate marketing merupakan strategi pemasaran di mana sebuah merek (*merchant*) memberikan komisi kepada pihak ketiga (*affiliate marketer*) yang berhasil mempromosikan dan menjual produk mereka. Pada platform TikTok, program TikTok Shop Affiliate memungkinkan kreator membuat “video konten affiliate” berupa video pendek yang menyajikan ulasan produk, menyematkan tautan afiliasi atau kode diskon, serta mengarahkan penonton untuk melakukan pembelian melalui link tersebut. Mekanisme kerjanya meliputi penyediaan produk oleh *merchant*, distribusi katalog melalui *affiliate network* (apabila digunakan), pembuatan konten video oleh kreator dengan penambahan parameter pelacakan unik, hingga pembelian oleh konsumen yang terhubung kembali ke kreator melalui tautan atau kode yang telah disisipkan. Setiap transaksi yang valid ditelusuri dan dikonfirmasi oleh merchant atau jaringan afiliasi, kemudian komisi dibayarkan sesuai ketentuan yang berlaku. Karakteristik utama video konten affiliate antara lain kreativitas dan keotentikan penyajian produk, durasi singkat sesuai format TikTok, serta call-to-action yang eksplisit untuk mendorong konversi pembelian[18].

4. Decision Tree

Decision Tree atau pohon keputusan merupakan salah satu algoritma dalam pembelajaran mesin (*machine learning*) yang digunakan untuk menyelesaikan masalah klasifikasi maupun regresi. Algoritma ini membentuk model prediktif

dalam bentuk struktur pohon, di mana setiap cabang (branch) merepresentasikan hasil dari suatu kondisi atau keputusan atas fitur tertentu, dan setiap daun (leaf node) mewakili hasil akhir atau prediksi[19].

Dalam struktur algoritma Decision Tree, terdapat beberapa komponen penting yang membentuk arsitektur dasar pohon keputusan. Komponen utama yang pertama adalah root node, yaitu simpul awal dari pohon keputusan yang menjadi titik permulaan dalam proses pengambilan keputusan. Simpul ini memuat fitur atau atribut pertama yang dianggap paling informatif berdasarkan proses pemilihan atribut, seperti menggunakan kriteria gain atau impurity[19]. Selanjutnya terdapat internal node, yaitu simpul-simpul penghubung yang merepresentasikan hasil pengujian terhadap atribut-atribut tertentu. Setiap internal node bercabang ke node lainnya sesuai dengan hasil evaluasi terhadap nilai atribut pada data. Simpul ini menjadi titik keputusan yang menentukan arah jalur pencarian berdasarkan kondisi tertentu[19].

Terakhir adalah leaf node atau simpul daun, yang merupakan simpul terminal pada pohon keputusan. Node ini tidak memiliki cabang lebih lanjut dan berfungsi untuk menyajikan hasil akhir dari proses klasifikasi atau prediksi. Pada pohon klasifikasi, leaf node berisi label kelas, sedangkan pada pohon regresi, leaf node memuat nilai prediksi hasil perhitungan dari input yang ditelusuri sepanjang jalur dari root hingga ke leaf node[19].

Pada setiap simpul, algoritma akan memilih fitur terbaik untuk membagi data berdasarkan kriteria tertentu. Dalam kasus klasifikasi, kriteria ini biasanya didasarkan pada pengukuran impurity seperti Gini Index atau Entropy, sedangkan dalam kasus regresi, digunakan metrik seperti Mean Squared Error (MSE) untuk menentukan pembagian data yang paling optimal[19].

Untuk regresi, Decision Tree bekerja dengan membagi dataset ke dalam kelompok-kelompok berdasarkan fitur input hingga kelompok tersebut cukup homogen dalam nilai targetnya. Ketika data baru diberikan, algoritma akan menelusuri jalur pada pohon berdasarkan nilai dari fitur input, dan hasil prediksi akan diambil dari nilai rata-rata target pada daun yang sesuai[19].

5. Python

Python adalah bahasa pemrograman tingkat tinggi (high-level programming language) yang pertama kali dikembangkan oleh Guido van Rossum dan dirilis secara publik pada tahun 1991. Python mendukung berbagai paradigma pemrograman, termasuk prosedural, berorientasi objek, dan fungsional. Bahasa ini banyak digunakan dalam pengembangan perangkat lunak, analisis data, kecerdasan buatan, dan ilmu data (data science)[20].

Python memiliki ekosistem pustaka yang luas dan terus berkembang, yang memfasilitasi berbagai aktivitas pengolahan data, pembelajaran mesin, dan pembangunan aplikasi[20]. Dalam konteks penelitian ini, beberapa pustaka yang digunakan antara lain:

a. Pandas

pustaka sumber terbuka (*open-source*) dalam Python yang menyediakan struktur data dan fungsi analitis untuk memproses data berformat tabel, terutama melalui objek utama berupa *DataFrame*. Pandas memungkinkan pengorganisasian data dalam bentuk dua dimensi, yang mendukung operasi baris dan kolom secara efisien[21].

b. Scikit-learn

pustaka Python yang menyediakan berbagai algoritma dan alat bantu untuk pembelajaran mesin (*machine learning*), termasuk metode klasifikasi, regresi, klasterisasi, dan pengolahan data. Pustaka ini mendukung pemodelan statistik, evaluasi performa, serta integrasi dengan pustaka lainnya seperti NumPy dan SciPy[22].

c. Flask

Flask merupakan pustaka atau kerangka kerja (*framework*) berbasis bahasa pemrograman Python yang dikategorikan sebagai *micro web framework*. Flask dirancang untuk membangun aplikasi web dengan struktur yang ringan, modular, dan fleksibel. Meskipun bersifat minimalis, Flask menyediakan fitur-fitur dasar yang diperlukan dalam pengembangan aplikasi web maupun layanan antarmuka pemrograman aplikasi, atau yang dikenal dengan API (*Application Programming Interface*)[23].

API adalah antarmuka yang memungkinkan suatu sistem atau aplikasi untuk berkomunikasi dan bertukar data dengan sistem lain. Dalam konteks pengembangan perangkat lunak modern, API umumnya digunakan untuk membangun layanan web (*web service*) yang memungkinkan client (misalnya: browser, aplikasi web, atau mobile) mengakses dan memproses data yang disediakan oleh server. API yang berbasis web disebut juga sebagai Web API, dan biasanya diimplementasikan dalam bentuk RESTful API (*Representational State Transfer*)[23].

RESTful API beroperasi di atas protokol HTTP (*HyperText Transfer Protocol*), yaitu protokol komunikasi yang digunakan di internet untuk mengatur pengiriman data antar perangkat. HTTP mendefinisikan berbagai metode atau verbs yang digunakan untuk menunjukkan aksi terhadap sumber daya (*resources*) dalam server[23].

6. Data Mining

Data mining merupakan proses untuk menemukan pola-pola tersembunyi, hubungan, atau pengetahuan baru dari kumpulan data dalam jumlah besar dengan memanfaatkan teknik statistik, matematika, dan pembelajaran mesin. Data mining tidak hanya berfokus pada penggalian informasi dari data, tetapi juga pada transformasi data menjadi pengetahuan yang berguna untuk mendukung pengambilan keputusan[24].

Menurut Han, Kamber, dan Pei, data mining merupakan salah satu tahap dalam proses Knowledge Discovery in Databases (KDD), yang mencakup keseluruhan tahapan mulai dari pengumpulan, pembersihan, transformasi, analisis data, hingga interpretasi hasil. Proses KDD secara umum terdiri atas seleksi data, pembersihan data, transformasi data, data mining, evaluasi dan interpretasi[25]:

Dalam praktiknya, data mining terbagi menjadi beberapa tugas utama, antara lain:

- a. Klasifikasi (*classification*): mengelompokkan data ke dalam kategori berdasarkan label tertentu.
- b. Klasterisasi (*clustering*): mengelompokkan data tanpa label ke dalam kelompok berdasarkan kesamaan.
- c. Asosiasi (*association rule mining*): mencari pola hubungan antar item dalam suatu dataset.
- d. Regresi (*regression*): memprediksi nilai numerik berdasarkan variabel input.