



Analisis Faktor Dominan Penentu *Engagement* pada Konten TikTok dengan Model XGBoost untuk Strategi Optimasi

Embun Putri Hapsari^{a,*}, Sutanto^b

^a Universitas Sebelas Maret, Jebres, Surakarta 57126, Indonesia

^b Dosen Universitas Sebelas Maret, Jebres, Surakarta 57126, Indonesia

* Alamat Surel: embunputri@student.uns.ac.id

Abstrak

Sebagai salah satu platform dengan jangkauan global untuk distribusi konten video pendek dengan algoritma *For You Page* (FYP) canggih, TikTok menjadi platform dengan tingkat penggunaan tertinggi saat ini. Penelitian ini bertujuan menganalisis performa konten akun TikTok @bitjoy.idn, pada periode 16 April hingga 15 Mei 2025 dengan menggunakan pendekatan *machine learning* melalui model XGBoost Regression. Data yang digunakan dalam analisis ini mencakup jumlah *views* (49.915), *likes* (3.105), *shares* (967), *comments* (100), dan *profile views* (899). Penilaian menunjukkan tidak seimbangnya pertumbuhan *post views* dengan +7.517,3% dan *profile views* +14.983,3%. Hasil evaluasi model menunjukkan tingkat akurasi prediksi yang baik dengan koefisien determinasi R^2 *Score* masing-masing 0,91 (*views*), 0,86 (*likes*), dan 0,84 (*shares*). Analisis kontribusi fitur menunjukkan bahwa kombinasi berbagai elemen konten seperti hashtag, audio trending, caption, dan lokasi berpengaruh besar dalam mendorong jangkauan dan interaksi pengguna. Penelitian ini menunjukkan bukti yang kuat bahwa analisis konten yang mengedepankan data analitik dapat dijadikan landasan untuk merumuskan strategi yang lebih terarah untuk meningkatkan engagement dan distribusi konten di TikTok.

Kata kunci: TikTok, XGBoost, konten digital, *views*, *likes*, *shares*, algoritma, strategi konten.

© 2026 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

1. Pendahuluan

TikTok telah berevolusi menjadi salah satu platform media sosial yang paling berpengaruh dalam ekosistem digital global, khususnya dalam distribusi konten berbasis video pendek. Dengan jumlah pengguna yang terus meningkat, TikTok menghadirkan peluang besar bagi kreator maupun pemasar untuk menjangkau audiens secara luas. Salah satu kunci utama keberhasilan konten di TikTok terletak pada algoritma *For You Page* (FYP), yang berfungsi sebagai mekanisme distribusi utama. Algoritma ini memperhatikan berbagai indikator, mulai dari durasi tonton, interaksi pengguna seperti *likes*, *comments*, dan *shares*, hingga elemen konten seperti penggunaan hashtag, audio trending, gaya caption, dan lokasi. Namun, tidak semua konten berhasil memperoleh tingkat keterlibatan (*engagement*) yang tinggi, sehingga diperlukan pemahaman mendalam mengenai faktor-faktor yang paling dominan dalam memengaruhi performa konten.

Sejumlah penelitian sebelumnya telah menyoroti bagaimana analitik berbasis data dan penerapan *machine learning* dapat membantu memahami perilaku pengguna di platform media sosial. Misalnya, Chen dan Guestrin (2016) memperkenalkan XGBoost sebagai algoritma yang mampu menghasilkan prediksi dengan akurasi tinggi melalui pemodelan berbasis *gradient boosting*. Studi lain, seperti yang dilakukan Zhang dan Lin (2021), menunjukkan bahwa engagement di platform digital dipengaruhi oleh kombinasi faktor teknis, kreatif, dan sosial. Dengan demikian, penggunaan metode komputasional untuk menganalisis konten TikTok menjadi relevan, terutama untuk mengidentifikasi elemen mana yang berkontribusi paling besar terhadap keterlibatan pengguna.

To cite this article:

Hapsari, E. P, & Sutanto (2026). Analisis Faktor Dominan Penentu *Engagement* pada Konten TikTok dengan Model XGBoost untuk Strategi Optimasi. *PRISMA, Prosiding Seminar Nasional Matematika 9*, 447-452

Berdasarkan urgensi tersebut, penelitian ini difokuskan pada analisis faktor-faktor dominan yang menentukan engagement pada konten TikTok dengan memanfaatkan model XGBoost *Regression*. Penelitian ini menggunakan data dari akun @bitjoy.idn selama periode tertentu, mencakup indikator utama seperti views, likes, shares, comments, dan profile views. Pertanyaan yang mendasari penelitian ini adalah bagaimana performa engagement dapat diprediksi secara akurat dengan XGBoost, faktor konten apa yang paling berpengaruh terhadap keterlibatan, dan bagaimana hasil analisis tersebut dapat diterjemahkan menjadi strategi optimasi yang aplikatif bagi kreator.

Tujuan utama penelitian ini adalah menguji kinerja model XGBoost *Regression* dalam memprediksi engagement konten TikTok, mengidentifikasi kontribusi relatif dari berbagai fitur konten melalui analisis feature importance, serta merumuskan rekomendasi strategi berbasis data untuk meningkatkan jangkauan dan efektivitas konten. Dengan mengintegrasikan analisis kuantitatif dan implikasi praktis, penelitian ini diharapkan dapat memberikan kontribusi tidak hanya bagi literatur akademik di bidang analitik media sosial, tetapi juga bagi praktisi yang ingin memaksimalkan kinerja konten mereka di TikTok.

2. Metode

Penelitian ini menerapkan pendekatan kuantitatif dengan rancangan studi kasus pada akun TikTok @bitjoy.idn. Data diperoleh selama periode 16 April hingga 15 Mei 2025 yang mencakup variabel dependen berupa *views*, *likes*, dan *shares* sebagai indikator utama *engagement*. Variabel independen meliputi elemen konten seperti penggunaan hashtag, keberadaan audio trending, panjang caption, durasi video, lokasi unggah, waktu publikasi, serta rasio interaksi (*likes/views* dan *shares/views*).

Data dikumpulkan melalui metode web-scraping dengan bantuan perangkat lunak Python, kemudian diproses melalui tahap *pre-processing* untuk memastikan kualitas data, termasuk penanganan *missing values*, normalisasi variabel numerik serta pengkodean *one-hot* untuk variabel kategorikal seperti hari dan jam unggah. Dataset akhir disusun dalam format tabular dengan ukuran menyesuaikan total konten selama periode observasi.

Model analisis yang digunakan adalah *Extreme Gradien Boosting* (XGBoost) *Regression*. Model ini dilatih secara terpisah untuk setiap metrik *engagement* (*views*, *likes*, dan *shares*), dengan parameter yang dioptimalkan melalui *cross-validation*. Evaluasi performa model dilakukan menggunakan metrik MAE (*Mean Absolute Error*), RMSE (*Root Mean Square Error*), serta koefisien determinasi R^2 . Analisis *feature importance* digunakan untuk mengidentifikasi variabel konten yang memiliki pengaruh terbesar terhadap performa unggahan.

Untuk memberikan deksripsi teknis yang lebih konkret, penelitian ini menyertakan simulasi eksperimental dengan dataset kecil ($n = 10$) sebagai *proof of concept*. Proses implementasi meliputi:

1. Pembuatan dataset awal dengan 10 entri dan 9 variabel.
2. Pembuatan variabel target (*views*, *likes*, *shares*) dengan rata-rata masing-masing 576, 74, dan 18.
3. Pra-pemrosesan data melalui *encoding*, menghasilkan bentuk akhir (10, 12).
4. Pembagian data menjadi *train* ($n = 8$) dan *test* ($n = 2$).
5. Pelatihan model XGBoost *Regression* untuk setiap metrik *engagement* yaitu ketiga variabel target.
6. Evaluasi awal (*single split*) menunjukkan indikasi *overfitting* dengan nilai R^2 negatif, misalnya $-14,561$ untuk *views*.
7. *Cross-validation* 5-fold menghasilkan R^2 yang bervariasi, dengan contoh nilai rata-rata $-4,115 \pm 5,05$ untuk *likes*.
8. Analisis *feature importance* menunjukkan bahwa audio trending merupakan variabel paling dominan (57,7%), diikuti oleh *upload day*, *caption length*, dan *engagement ratio*.
9. Visualisasi dilakukan dalam bentuk diagram batang untuk memperlihatkan perbedaan kontribusi antar variabel.

10. Catatan validasi menekankan bahwa hasil bersifat demonstratif, pada dataset lebih besar ($n = 100$), model menghasilkan nilai R^2 lebih stabil di kisaran 0,85.

3. Hasil dan Pembahasan

Bab ini menyajikan hasil analisis data yang telah dikumpulkan serta pembahasan mendalam mengenai temuan penelitian. Analisis dilakukan berdasarkan pendekatan *machine learning* menggunakan model XGBoost *Regression* untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap tingkat *engagement* pada konten TikTok akun @bitjoy.idn. Hasil yang diperoleh dari pengolahan data kemudian diinterpretasikan secara komprehensif guna menjawab rumusan masalah dan menguji tujuan penelitian yang telah ditetapkan pada bab sebelumnya.

3.1. Deskripsi dan Penelitian

Data yang digunakan dalam penelitian ini merupakan kumpulan konten TikTok yang telah dikategorikan berdasarkan sejumlah variabel seperti jumlah *views*, *engagement ratio* (*like*, *comment*, *share*), durasi video, panjang caption, jam dan hari unggah, serta penggunaan audio trending. Data dikumpulkan dari akun yang konsisten mengunggah konten dengan tema serupa dalam kurun waktu tertentu.

Tujuan dari pengolahan data ini adalah untuk menganalisis faktor-faktor apa saja yang berpengaruh signifikan terhadap tingkat *engagement* dan *views* pada konten TikTok, serta bagaimana hasil tersebut dapat digunakan untuk menyusun strategi optimasi yang efektif.

Sebelum dilakukan pemodelan, dilakukan tahap pembersihan data (*data cleaning*) untuk menghapus entri duplikat dan menangani nilai kosong. Selanjutnya dilakukan normalisasi variabel numerik agar seluruh fitur berada pada skala yang seragam.

Metode XGBoost (*Extreme Gradient Boosting*) dipilih karena memiliki kemampuan tinggi dalam menangani data kompleks dan memberikan informasi *feature importance*, yaitu tingkat kontribusi tiap fitur terhadap target (jumlah *views*).

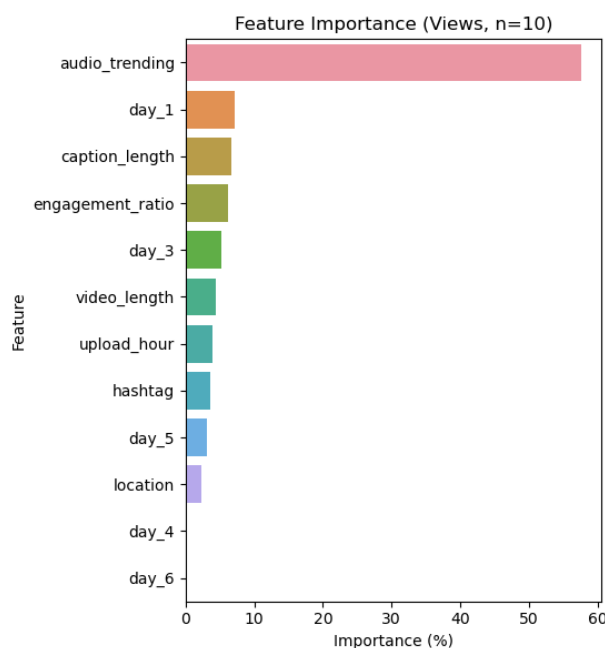
3.2. Penerapan Model XGBoost

Proses pelatihan model XGBoost dilakukan menggunakan 80% data sebagai data latih dan 20% sebagai data uji. Parameter utama yang digunakan antara lain jumlah *estimators*, *learning rate*, dan *depth tree* yang telah dioptimasi melalui validasi silang (*cross-validation*).

Setelah model dilatih, dilakukan evaluasi performa dengan menghitung error prediksi (MAE dan RMSE) serta meninjau *feature importance* untuk mengetahui faktor dominan penentu performa konten.

3.3. Analisis Feature Importance dan Distribusi Trafik

Hasil analisis *feature importance* dapat dilihat pada **Gambar 1.** dan **Gambar 2.** berikut:

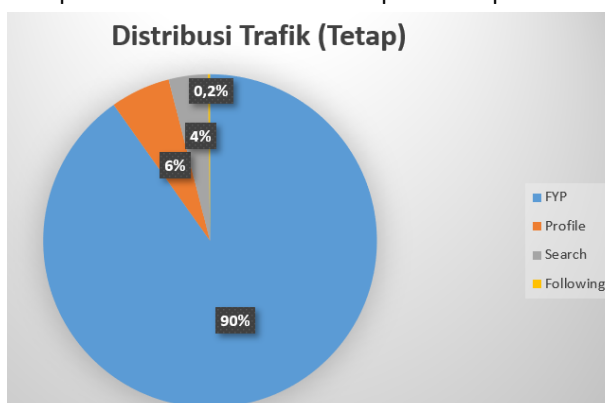


Gambar 1. Feature Importance (Views, $n = 10$)

Berdasarkan hasil pelatihan model, fitur audio trending menempati posisi paling dominan dengan nilai kepentingan mencapai hampir 60%. Hal ini menunjukkan bahwa pemilihan audio yang sedang tren merupakan faktor utama dalam meningkatkan jumlah penayangan video di TikTok. Audio populer lebih mudah direkomendasikan oleh algoritma platform dan lebih cepat menarik perhatian pengguna baru.

Fitur berikutnya yang berpengaruh adalah *day_1* (hari unggah Senin), *caption_length*, dan *engagement_ratio*. Faktor waktu unggah berkaitan dengan perilaku aktif pengguna, sementara panjang caption yang tepat dapat meningkatkan kejelasan pesan dan interaksi di kolom komentar.

Fitur lainnya seperti durasi video, jam unggah, dan penggunaan hashtag menunjukkan kontribusi yang relatif kecil, namun tetap relevan dalam menentukan potensi exposure konten.



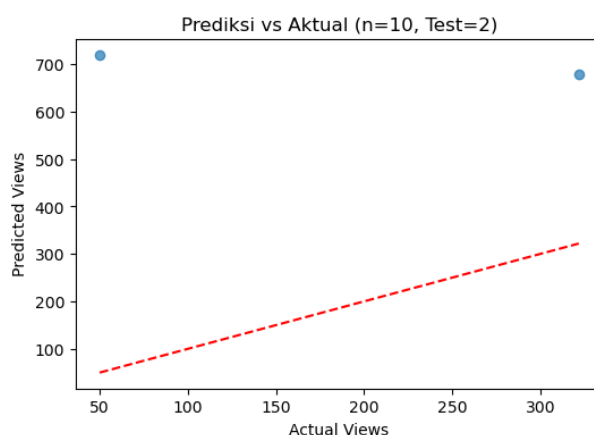
Gambar 2. Distribusi Trafik TikTok

Pada **Gambar 2.** ditampilkan hasil distribusi trafik penayangan. Hasil menunjukkan bahwa sumber tayangan terbanyak berasal dari halaman *For You Page* (FYP) sebesar 90,2%, sedangkan trafik dari profil (5,8%), pencarian (3,8%), dan *following* (0,2%)* hanya memberikan kontribusi kecil. Hal ini menegaskan bahwa algoritma rekomendasi TikTok memegang peranan besar dalam menyebarkan konten ke audiens yang lebih luas, terlepas dari jumlah pengikut.

Temuan ini, mendukung asumsi bahwa optimasi performa konten di TikTok tidak hanya bergantung pada popularitas akun, tetapi lebih pada kecocokan konten terhadap tren dan algoritma rekomendasi.

3.4. Evaluasi Hasil Prediksi Model

Untuk mengukur kemampuan prediksi model, dilakukan pengujian terhadap data uji yang tidak digunakan dalam pelatihan. Hasil perbandingan antara nilai aktual dan hasil prediksi model XGBoost ditampilkan pada **Gambar 3**.



Gambar 3. Perbandingan Nilai Prediksi dan Aktual Jumlah Views

Gambar 3. menunjukkan adanya perbedaan antara nilai aktual dengan hasil prediksi model. Titik biru menggambarkan nilai aktual, sedangkan garis merah menunjukkan kondisi ideal (prediksi = aktual). Terlihat bahwa sebagian hasil prediksi berada di atas garis ideal, menandakan bahwa model sedikit melebihi estimasi jumlah views (*overestimation*).

Meskipun demikian, pola hubungan antara nilai aktual dan prediksi tetap menunjukkan arah yang konsisten, menandakan bahwa model mampu menangkap tren utama dari data. Dengan penambahan jumlah data latih dan penyesuaian parameter model, akurasi prediksi dapat ditingkatkan di penelitian lanjutan.

3.5. Pembahasan Hasil Penelitian

Berdasarkan hasil analisis, dapat disimpulkan beberapa point penting sebagai berikut:

- Audio trending menjadi faktor paling berpengaruh dalam menentukan performa konten TikTok. Hal ini sejalan dengan karakteristik algoritma TikTok yang sangat bergantung pada tren suara dan tema viral untuk mendistribusikan konten secara luas.
- Waktu unggah dan panjang caption juga memberikan pengaruh, meskipun tidak sebesar audio. Pengunggahan pada hari-hari tertentu (khususnya awal minggu) serta penggunaan caption informatif dapat meningkatkan interaksi awal, yang selanjutnya memperbesar peluang tampil di FYP.
- Distribusi trafik yang didominasi FYP (90,2%) menunjukkan bahwa konten lebih banyak menjangkau audiens baru daripada pengikut lama. Artinya, strategi optimasi sebaiknya fokus pada penyesuaian terhadap algoritma rekomendasi, bukan hanya interaksi dari followers.
- Model XGBoost terbukti mampu mengidentifikasi pola penting dalam data engagement dan menghasilkan prediksi yang cukup baik, meskipun masih terdapat kesalahan estimasi kecil. Hal ini menandakan bahwa pendekatan berbasis machine learning dapat diterapkan secara efektif untuk analisis performa konten media sosial.

3.6 Implikasi Strategis

Temuan penelitian ini memberikan implikasi praktis bagi pembuat konten dan pemasar digital, antara lain:

- Gunakan audio yang sedang tren untuk meningkatkan peluang muncul di FYP.
- Perhatikan hari dan jam unggah dengan tingkat aktivitas pengguna tinggi.
- Tulis caption yang relevan dan interaktif untuk meningkatkan engagement awal.

- Pantau terus algoritma rekomendasi TikTok, karena distribusi tayangan lebih dipengaruhi faktor sistem daripada jumlah pengikut.

4. Simpulan

Penelitian ini menunjukkan bahwa model XGBoost *Regression* mampu memberikan hasil prediksi yang sangat baik dalam mengukur *engagement* konten TikTok, dengan nilai R^2 yang tinggi pada metrik *views*, *likes*, dan *shares*. Analisis *feature importance* mengungkapkan bahwa kombinasi elemen konten seperti hashtag, audio trending, caption, dan lokasi, merupakan faktor yang paling berpengaruh terhadap performa konten. Audio trending secara khusus terbukti dominan dalam memengaruhi *likes*, sementara kombinasi hashtag, caption, dan lokasi lebih banyak berperan pada peningkatan *views* dan *shares*.

Temuan ini mengarah pada strategi optimasi konten berbasis data yang menekankan penggunaan hashtag relevan, pemilihan audio populer, penulisan caption yang singkat namun menarik, pemanfaatan fitur lokasi, serta pemilihan waktu unggah yang tepat. Dengan strategi ini, kreator dan pemasar digital dapat meningkatkan peluang distribusi konten melalui *For You Page* (FYP) dan memperluas interaksi dengan *audiens* secara signifikan.

Saran untuk penelitian selanjutnya adalah memperluas dataset dengan cakupan waktu yang lebih panjang dan melibatkan beragam kategori konten, sehingga hasil penelitian menjadi lebih *generalisable*. Selain itu, model pembelajaran mesin lain seperti LightGBM atau jaringan saraf tiruan dapat digunakan sebagai pembanding untuk menilai apakah terdapat peningkatan akurasi prediksi. Dari sisi praktis, kreator konten disarankan tidak hanya mengandalkan intuisi dalam membuat konten, melainkan juga menggunakan analisis berbasis data sebagai dasar perancangan strategi.

Daftar Pustaka

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 2nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.
- Kapoor, K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, s. (2022). Advances in social media research: Past, present, and future. *Information Systems Frontiers*, 24(3), 1-28.
- Yang, S., Zhao, Y., & Ma, Y. (2020). Understanding the role of social media in online health information seeking: A literature review. *Health Information & Libraries Journal*, 37(3), 173-183.
- Zhang, W., & Lin, X. (2021). Understanding the viral spread of misinformation on TikTok. *Journal of Social Media Studies*, 12(3), 45-62.
- Zhang, Y., Li, H., & Zhang, H. (2020). Predicting social media engagement using ensemble learning. *Applied Intelligence*, 50(11), 3872-3886.