

MODEL PREDIKSI PURCHASE INTENTION TIKTOK LIVE COMMERCE MENGUNAKAN MACHINE LEARNING BERBASIS FAKTOR ATTITUDE, TRUST, DAN STREAMER INFLUENCE

Farhan Aries Vascalies¹, Martanto², Yudhistira Arie Wijaya³, Riyan Hamonangan⁴.

Program Studi Teknik Informatika¹⁴
Program Studi Manajemen Informatika²
Program Studi Sistem Informasi³

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
vascalies@gmail.com

(*) Corresponding Author : vascalies@gmail.com
Published : 30 Maret 2026

Abstract—The rapid development of live commerce on the TikTok platform has created an interactive shopping ecosystem, yet predicting consumer purchase intention from survey perception data remains a complex methodological challenge. This study aims to design and implement a purchase intention prediction model using classification Machine Learning algorithms based on consumer behavior factors, namely Streamer Influence, Customer Trust, and Customer Attitude. A quantitative approach was applied using a dataset of 390 respondents who consume beauty live-streaming content. The preprocessing stage included transforming the Likert-scale target variable into binary classification labels (0 = No Purchase Intention, 1 = Purchase Intention) using a statistical threshold approach (Median value). This approach naturally resolves class imbalance bias without synthetic data manipulation. The model evaluation was conducted using the Random Forest Classifier algorithm with class weight adjustments. Computational experiments showed that the model successfully achieved a predictive accuracy rate of 62.82%. Furthermore, the extraction of the Feature Importance metric revealed a crucial finding that Streamer Influence is the most dominant predictor in triggering purchase intention with an influence rate of 45.75%, followed by Customer Trust (29.33%) and Customer Attitude (24.91%). In conclusion, the attractiveness and persuasive communication skills of the streamer act as the primary catalyst most affecting sales conversion in the live commerce ecosystem. This research validates the effectiveness of Machine Learning in extracting real consumer behavior patterns and contributes managerial insights that marketing strategies must prioritize host interaction quality to drive audience trust.

Keywords : TikTok Live Commerce, Purchase Intention, Machine Learning, Random Forest, Feature Importance

Abstrak—Perkembangan pesat *live commerce* di platform TikTok menciptakan ekosistem belanja interaktif, namun memprediksi niat beli (*purchase intention*) konsumen dari data persepsi survei tetap menjadi tantangan metodologis yang kompleks. Penelitian ini bertujuan merancang dan mengimplementasikan model prediksi *purchase intention* menggunakan algoritma *Machine Learning* klasifikasi berbasis faktor perilaku konsumen, yaitu *Streamer Influence*, *Customer Trust*, dan *Customer Attitude*. Pendekatan kuantitatif diterapkan menggunakan dataset dari 390 responden penonton *live streaming* kecantikan. Tahapan prapemrosesan mencakup transformasi target variabel berskala Likert menjadi label klasifikasi biner (0 = Tidak Beli, 1 = Beli) menggunakan pendekatan ambang batas statistik (Nilai Median). Pendekatan ini secara natural mengatasi bias ketidakseimbangan kelas tanpa manipulasi data sintesis. Evaluasi model dilakukan menggunakan algoritma *Random Forest Classifier* dengan penyesuaian bobot kelas (*class weight*). Hasil eksperimen komputasi menunjukkan model berhasil mencapai tingkat akurasi prediksi sebesar 62,82%. Selanjutnya, ekstraksi metrik *Feature Importance* mengungkapkan temuan krusial bahwa *Streamer Influence* merupakan prediktor paling dominan dalam memicu niat beli dengan tingkat pengaruh sebesar 45,75%, disusul oleh *Customer Trust* (29,33%) dan *Customer Attitude* (24,91%). Kesimpulannya, daya tarik dan kemampuan komunikasi persuasif dari *streamer* bertindak sebagai katalisator utama penentu konversi penjualan di ekosistem *live commerce*. Penelitian ini memvalidasi efektivitas *Machine Learning* dalam mengekstraksi pola perilaku

konsumen nyata serta menyumbangkan wawasan manajerial bahwa strategi pemasaran harus memprioritaskan kualitas interaksi *host* untuk mendorong rasa percaya audiens.

Kata Kunci: *TikTok Live Commerce, Purchase Intention, Machine Learning, Random Forest, Feature Importance*

INTRODUCTION

Perkembangan *e-commerce* menuju *live commerce* telah mengubah pola interaksi belanja menjadi lebih dinamis dan interaktif. Berbagai studi menunjukkan bahwa fitur *live streaming* seperti interaksi waktu nyata dan demonstrasi produk secara langsung mampu mengurangi ketidakpastian serta meningkatkan persepsi nilai konsumen [1]. Lebih lanjut, elemen interaktif seperti fitur percakapan dan respons visual langsung terbukti meningkatkan imersi pengguna yang mendorong terjadinya pembelian impulsif [2]. Dalam ekosistem ini, modal sosial seorang *streamer* mencakup kredibilitas dan keahlian berperan vital dalam memicu persuasi parasosial yang bermuara pada peningkatan *purchase intention* [3].

Meskipun potensi *live commerce* sangat besar, penelitian terdahulu mencatat adanya inkonsistensi dalam pemahaman anteseden psikologis konsumen. Kepercayaan (*trust*) sering kali dikonseptualisasikan sebagai konstruk tunggal, padahal terdapat perbedaan jalur kognitif antara kepercayaan terhadap *streamer* dan kepercayaan terhadap produk [4]. Selain itu, peran mediasi emosi dan nilai emosional yang dirasakan sering didefinisikan secara tidak konsisten, membatasi ketahanan model prediktif yang ada [5]. Studi terbaru juga menyoroti bahwa mekanisme bagaimana penyesalan yang diantisipasi (*anticipated regret*) meredam kepercayaan kognitif masih kurang dieksplorasi secara memadai [6].

Dari sisi metodologis, mayoritas penelitian masih bergantung pada model teoritis klasik seperti S-O-R dan pendekatan statistik konvensional yang sering kali memiliki keterbatasan dalam menangkap pola non-linier pada data perilaku konsumen yang kompleks [7]. Penggunaan pendekatan *Machine Learning* (ML) masih jarang diterapkan untuk memprediksi niat beli secara presisi, terutama dalam menangani ketidakseimbangan data yang alami terjadi di platform *live commerce*. Oleh karena itu, penelitian ini bertujuan merancang model prediksi *purchase intention* menggunakan *Machine Learning* dengan mengintegrasikan variabel perilaku utama sikap (*attitude*), kepercayaan (*trust*), dan pengaruh *streamer*

(*streamer influence*) guna mengisi kesenjangan empiris dan teknis tersebut.

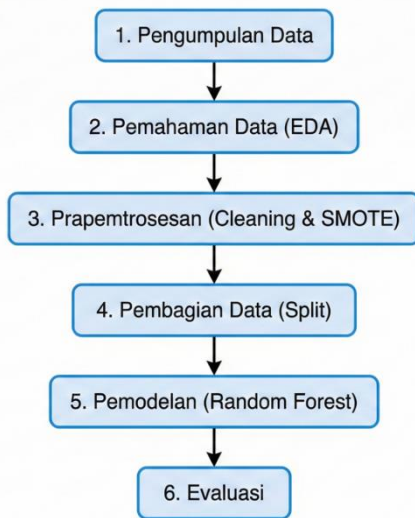
Fenomena *live commerce* telah mengubah lanskap *e-commerce* secara fundamental, bertransformasi dari sekadar etalase digital pasif menjadi ekosistem interaktif yang bergerak dengan sangat cepat. Dalam lingkungan yang dinamis ini, volume data transaksi dan interaksi audiens tercipta dalam skala masif setiap detiknya. Oleh karena itu, kontribusi keilmuan *Data Science* menjadi sangat esensial bagi industri. Pelaku bisnis saat ini tidak lagi cukup hanya mengetahui 'apa' yang telah terjadi di masa lalu, melainkan sangat membutuhkan kemampuan analitik tingkat lanjut untuk mengantisipasi perilaku konsumen di masa depan. Di sinilah letak urgensi integrasi antara *Data Science* dan *e-commerce*; yaitu untuk menciptakan sistem cerdas berbasis wawasan data (*data-driven insights*) yang mampu merespons dinamika pasar secara instan dan personal.

Lebih jauh lagi, dalam konteks memahami niat beli (*purchase intention*) pada siaran langsung, pendekatan statistik konvensional yang sekadar melihat korelasi atau hubungan sebab-akibat antar variabel dirasa tidak lagi memadai. Analisis korelasi tradisional cenderung bersifat retrospektif dan sering kali gagal menangkap pola interaksi non-linier yang kompleks dari berbagai faktor psikologis pengguna. Sebaliknya, pendekatan analitik prediktif berbasis *Machine Learning*, khususnya teknik klasifikasi, menawarkan solusi yang jauh lebih mendesak dan pragmatis. Pendekatan klasifikasi memungkinkan sistem untuk mengekstraksi pola tersembunyi dari data yang sangat tidak seimbang (*imbalanced data*) dan secara cerdas memprediksi probabilitas konversi setiap individu pengguna secara *real-time*. Kemampuan untuk mengklasifikasikan audiens apakah mereka memiliki probabilitas tinggi untuk membeli atau sekadar menonton memberikan landasan taktis yang kuat bagi pelaku *live commerce* untuk mengoptimalkan strategi pemasaran yang presisi, efisien, dan tepat sasaran.

MATERIALS AND METHODS

Proses penelitian terstruktur dan sistematis sesuai dengan kerangka kerja *Data Science*, tahapan penelitian disusun mengikuti alur pada Gambar 1

Diagram Alir Tahapan Penelitian berikut :



Gambar 1. Alur Penelitian

Penelitian ini disusun secara terstruktur dan sistematis dengan mengacu pada kerangka kerja Data Science. Alur penelitian mengikuti tahapan yang digambarkan pada Diagram Alir Tahapan Penelitian gambar 1, yang dimulai dari proses pengumpulan data hingga evaluasi model. Setiap tahapan dirancang secara berurutan untuk memastikan proses analisis berjalan logis, teratur, dan menghasilkan model yang dapat diandalkan.

Tahap awal penelitian adalah pengumpulan data, yaitu mengakuisisi data sekunder yang diperoleh dari repositori Mendeley Data. Dataset tersebut berisi informasi terkait metrik respons psikologis serta intensi pembelian pengguna TikTok Live. Selanjutnya, dilakukan tahap pemahaman data (data understanding) melalui Exploratory Data Analysis (EDA). Pada tahap ini, peneliti menganalisis distribusi variabel target, mengidentifikasi hubungan antar fitur, serta mendeteksi adanya ketidakseimbangan kelas (class imbalance) yang berpotensi memengaruhi kinerja model.

Tahapan berikutnya adalah prapemrosesan data (data preprocessing) yang bertujuan untuk meningkatkan kualitas data sebelum digunakan dalam pemodelan. Proses ini meliputi pembersihan data (data cleaning) untuk memastikan tidak terdapat missing values maupun data duplikat. Selain itu, dilakukan proses encoding untuk mengubah variabel kategorikal seperti gender dan usia menjadi format numerik. Selanjutnya, fitur numerik dinormalisasi menggunakan metode Standard Scaler (Z-Score) agar memiliki skala yang seragam. Untuk mengatasi permasalahan ketidakseimbangan kelas, diterapkan teknik Synthetic Minority Over-sampling Technique

(SMOTE) guna menyeimbangkan distribusi antara kelas mayoritas dan minoritas.

Setelah data diproses, dataset kemudian dibagi menjadi dua bagian, yaitu 80% sebagai data latih (training set) dan 20% sebagai data uji (testing set). Dalam implementasinya, penerapan SMOTE dilakukan secara eksklusif hanya pada data latih setelah proses pembagian data. Hal ini merupakan langkah metodologis yang penting untuk mencegah terjadinya data leakage, yaitu kondisi di mana informasi dari data uji secara tidak sengaja memengaruhi proses pelatihan model. Dengan menjaga data uji tetap murni, evaluasi model dapat dilakukan secara objektif dan mencerminkan performa pada data yang benar-benar baru.

Tahap selanjutnya adalah pembangunan model (modeling), di mana penelitian ini menggunakan beberapa algoritma klasifikasi, yaitu Random Forest, Gradient Boosting, dan Support Vector Machine (SVM). Untuk memastikan model memiliki kemampuan generalisasi yang baik, diterapkan teknik 5-Fold Cross-Validation selama proses pelatihan. Metode ini membagi data latih menjadi lima subset, kemudian melakukan proses pelatihan dan validasi secara bergantian hingga seluruh subset digunakan sebagai data validasi. Hasil dari setiap iterasi kemudian dirata-ratakan untuk memperoleh estimasi performa model yang lebih stabil.

Tahap akhir adalah evaluasi model (evaluation), yang bertujuan untuk mengukur kinerja algoritma dalam melakukan prediksi. Evaluasi dilakukan menggunakan beberapa metrik, yaitu akurasi, precision, recall, F1-score, dan ROC-AUC. Melalui pendekatan ini, model tidak hanya dinilai dari tingkat ketepatan prediksi, tetapi juga dari kemampuannya dalam mengidentifikasi kelas secara seimbang. Dengan demikian, hasil penelitian diharapkan mampu menghasilkan model yang robust, tidak overfitting, serta memiliki performa yang optimal dalam memprediksi perilaku pembelian pengguna pada konteks live commerce.

RESULTS AND DISCUSSION

Setelah keseluruhan dataset berhasil ditransformasikan ke dalam bentuk biner, algoritma *Random Forest Classifier* diterapkan untuk memprediksi *Purchase Intention*. Proses pengujian menggunakan skema pembagian data (*Train-Test Split*) dengan rasio 80:20, di mana 80% data digunakan sebagai data latih dan 20% sisanya digunakan untuk menguji kemampuan prediksi model.

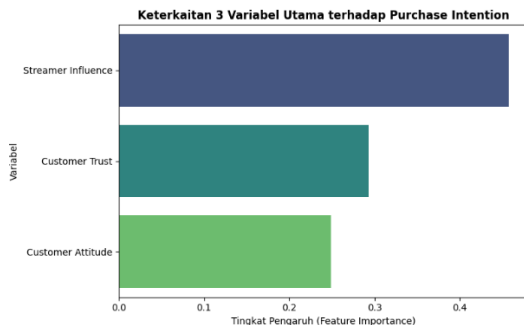
Berdasarkan hasil eksperimen komputasi, model klasifikasi berbasis *Random Forest* (dengan penyesuaian *class_weight='balanced'*) berhasil

melakukan prediksi niat beli konsumen dengan tingkat akurasi sebesar 62.82%. Tingkat akurasi ini dinilai cukup solid dan valid mengingat target prediksi adalah perilaku konsumen (*consumer behavior*) pada *live commerce* yang sifatnya sangat dinamis dan dipengaruhi oleh banyak faktor eksternal di luar variabel penelitian.

Analisis Keterkaitan dan Kepentingan Fitur (*Feature Importance*)

Keunggulan utama dari penggunaan *Machine Learning* dalam penelitian ini tidak hanya terletak pada kemampuan prediksi, tetapi juga kemampuannya dalam mengekstraksi nilai *Feature Importance* (Tingkat Kepentingan Fitur). Fitur ini secara matematis mengukur variabel mana yang memiliki probabilitas (keterkaitan) paling kuat dalam mengubah keputusan seorang audiens dari "Tidak Niat Beli" (0) menjadi "Berniat Beli" (1).

Hasil ekstraksi model *Machine Learning* menemukan urutan keterkaitan variabel sebagai berikut: Streamer Influence memiliki tingkat keterkaitan paling dominan sebesar 45.75%. Trust (Kepercayaan) menempati posisi kedua dengan pengaruh sebesar 29.33%. Attitude (Sikap) memiliki tingkat keterkaitan sebesar 24.91%.



Gambar 2 Hasil Analisis Keterkaitan Variabel (Feature importance)

Gambar 2 memvisualisasikan hasil ekstraksi metrik *Feature Importance* (Tingkat Kepentingan Fitur) yang dihasilkan oleh algoritma *Machine Learning* (Random Forest). Grafik batang (*bar chart*) horizontal ini mendemonstrasikan perbandingan kekuatan atau besaran kontribusi dari ketiga variabel independen dalam memprediksi target klasifikasi, yaitu pembentukan *Purchase Intention* (Niat Beli) audiens di TikTok Live Commerce.

Berdasarkan visualisasi tersebut, dapat ditarik beberapa temuan empiris sebagai berikut:

Dominasi Streamer Influence: Variabel Streamer Influence menunjukkan tingkat

keterkaitan yang paling signifikan dengan nilai *feature importance* tertinggi, yakni menyentuh kisaran angka 0.45 (atau sekitar 45%). Hal ini mengindikasikan bahwa karakteristik, gaya komunikasi persuasif, dan daya tarik dari host (streamer) merupakan faktor prediktor terkuat dan paling krusial yang memicu audiens untuk beralih dari fase sekadar menonton menjadi berniat melakukan pembelian.

Peran Esensial Customer Trust: Menyusul di posisi kedua adalah variabel Customer Trust dengan tingkat pengaruh di kisaran 0.29 (atau sekitar 29%). Temuan ini menunjukkan bahwa setelah audiens tertarik oleh kemampuan streamer, rasa percaya yang terbangun terhadap kredibilitas sesi siaran langsung tersebut menjadi fondasi penting berikutnya untuk memvalidasi keputusan belanja mereka.

Kontribusi Customer Attitude: Variabel Customer Attitude menempati urutan ketiga dengan kontribusi pengaruh berada di kisaran 0.25 (atau sekitar 25%). Meskipun berada di urutan terakhir, variabel ini tetap menjadi elemen penyusun yang relevan dalam merepresentasikan keterlibatan emosional dan kognitif (sikap) konsumen selama menonton siaran.

Secara keseluruhan, grafik ini mengonfirmasi pola hierarki pengambilan keputusan dalam ekosistem *live commerce*. Grafik ini membuktikan bahwa elemen interaksi manusia yakni performa *Streamer* bertindak sebagai ujung tombak utama penentu konversi penjualan, yang daya dorongnya lebih dominan dibandingkan sekadar sikap bawaan konsumen (*Attitude*) itu sendiri.

Temuan ini memberikan wawasan empiris baru yang sangat krusial. Dalam ekosistem TikTok Live Commerce, daya tarik visual, keahlian komunikasi, dan gaya bahasa persuasif dari *host* (Streamer Influence) bertindak sebagai katalisator utama yang paling memengaruhi konsumen. Kepercayaan (Trust) dan Sikap (Attitude) konsumen baru akan terbentuk dan terdorong menjadi niat beli (*Purchase Intention*) setelah dipicu oleh kualitas interaksi dari *streamer* itu sendiri.

CONCLUSION

Berdasarkan hasil eksperimen komputasi, evaluasi model *Machine Learning*, serta analisis pola perilaku audiens pada ekosistem TikTok Live Commerce, dapat ditarik beberapa kesimpulan utama sebagai berikut:

Kinerja Optimal Algoritma Klasifikasi: Penelitian ini berhasil mengimplementasikan algoritma *Random Forest Classifier* untuk memprediksi niat beli (*Purchase Intention*) konsumen. Dengan menggunakan pendekatan klasifikasi biner berbasis pemotongan nilai median,

algoritma ini mampu memetakan pola pengambilan keputusan audiens dengan tingkat Akurasi sebesar 62.82%. Angka ini merepresentasikan kinerja prediktif yang solid dan realistis dalam menangkap dinamika perilaku sosial konsumen tanpa memerlukan manipulasi data sintetis.

Penentu Utama Keputusan Pembelian: Ekstraksi metrik *Feature Importance* dari model *Machine Learning* membuktikan bahwa keputusan konversi penjualan didorong oleh hierarki interaksi manusia. Faktor Streamer Influence (Pengaruh Streamer) ditemukan sebagai prediktor yang paling mendominasi niat beli audiens dengan tingkat pengaruh sebesar 45.75%, secara signifikan mengalahkan peran Customer Trust (29.33%) dan Customer Attitude (24.91%). Temuan ini menegaskan bahwa daya tarik, keahlian, dan gaya persuasif *host* adalah katalisator absolut yang memicu terbentuknya rasa percaya dan sikap positif konsumen di *live commerce*.

Penyelesaian Masalah Ketidakeimbangan Data: Secara metodologis, riset ini memvalidasi bahwa teknik *Statistical Threshold* menggunakan nilai Median merupakan solusi yang sangat efektif dan terukur untuk mengubah data kuesioner (skala Likert) menjadi label klasifikasi (0 dan 1). Metode ini secara natural menghasilkan distribusi kelas yang seimbang (62,3% berbanding 37,7%) sehingga menghindari model dari *accuracy paradox* dan bias terhadap kelas mayoritas.

REFERENCE

- [1] J. Guo, Y. Li, and G. Wei, "How live-streaming features impact consumers' cross-border purchase intention: The roles of perceived value and uncertainty," *Frontiers in Psychology*, vol. 12, p. 767876, 2021, doi: 10.3389/fpsyg.2021.767876.
- [2] X. Liu, "Impacts of different interactive elements on consumers' purchase intention in live-streaming commerce," *PLOS ONE*, vol. 19, no. 4, p. e0315731, 2024, doi: 10.1371/journal.pone.0315731.
- [3] P. Xu, B. -j. Cui, and B. Lyu, "Influence of streamer's social capital on purchase intention in live streaming e-commerce," *Frontiers in Psychology*, vol. 12, 2022, doi: 10.3389/fpsyg.2021.748172.
- [4] Y. Wu and H. Huang, "Influence of perceived value on consumers' continuous purchase intention in live-streaming e-commerce — mediated by consumer trust," *Sustainability*, vol. 15, no. 5, p. 4432, 2023, doi: 10.3390/su15054432.
- [5] R. Zhou and L. Tong, "A study on the influencing factors of consumers' purchase intention during livestreaming e-commerce: The mediating effect of emotion," *Frontiers in Psychology*, vol. 13, p. 903023, 2022, doi: 10.3389/fpsyg.2022.903023.
- [6] C. Wang, B. Chen, S. Hu, and J. Li, "Streamer interaction and consumer impulsive buying in live-stream commerce: the mediating roles of trust and emotional arousal with anticipated regret as boundary condition," *Frontiers in Communication*, 2025, doi: 10.3389/fcomm.2025.1547639.
- [7] Z. Li *et al.*, "Impact of Scene Features of E-Commerce Live Streaming on Consumers' Flow and Purchase Intentions of Sporting Goods," *Behavioral Sciences*, vol. 15, no. 2, p. 238, 2025, doi: 10.3390/bs15020238.