

BAB I

PENDAHULUAN

Bab 1 berfungsi sebagai landasan awal dalam penelitian ini. Bab ini menyajikan gambaran umum mengenai latar belakang permasalahan yang menjadi fokus penelitian, rumusan masalah yang ingin dijawab, tujuan yang ingin dicapai, serta manfaat yang diharapkan dari penelitian. Bab ini juga menjelaskan ruang lingkup penelitian agar pembahasan tetap terfokus dan tidak melebar. Pemahaman terhadap bab ini dapat memberikan pembaca konteks dan arah dari keseluruhan penelitian yang dilakukan.

1.1 Latar Belakang

Virtual YouTuber (VTuber) adalah bentuk baru dari penyajian konten video *online* yang menampilkan kreator dalam bentuk digital yang dapat berupa 2D atau 3D. Konsep VTuber pertama kali dikenalkan pada tahun 2016 dengan kemunculan ‘Kizuna Ai’. Meledaknya popularitas VTuber dimulai ketika pandemi corona di 2020 (Lehtovirta, 2023). *Lockdown* yang terjadi di seluruh dunia memaksa banyak orang untuk mencari hiburan secara *online*. Maraknya budaya *anime* Jepang dan konten video YouTube, menyebabkan budaya VTuber menjadi populer. Semenjak kemunculan ‘Kizuna Ai’, perkembangan VTuber berkembang pesat dengan total 49.500 VTuber aktif di platform Twitch dan YouTube pada tahun 2023 (An, 2023). Di Indonesia, VTuber mulai memasuki pasar lokal. Hal ini ditandai dengan kolaborasi VTuber, Kobo Kanaeru, dengan merek besar seperti Tokopedia, Indomie, dan Honda seperti yang dilaporkan oleh laporan Cover Corp. pada tahun 2024 (Cover, 2024). Tingginya jumlah kreator dan konten yang dipublikasikan setiap hari, persaingan antar VTuber pun semakin ketat, sehingga pemahaman mengenai performa video menjadi aspek penting dalam mempertahankan perhatian penonton.

Performa konten seorang VTuber dapat dilihat melalui tingkat *engagement* dari video yang diunggah. Dalam konteks media digital, *engagement* merujuk pada tingkat interaksi penonton terhadap suatu video yang umumnya diukur melalui metrik seperti *views*, *likes*, *comments*, dan jumlah *subscribers*. Penelitian yang dilakukan oleh Ameira dkk. (2023) dan Munaro dkk. (2021),, menegaskan bahwa metrik-metrik tersebut merupakan indikator utama dalam menentukan tingkat *engagement*. *Views* memiliki peran yang sangat penting terhadap tingkat *engagement*, dikarenakan interaksi seperti *likes* dan *comments* tidak dapat

terjadi tanpa adanya jumlah penonton yang memadai. Metrik-metrik ini juga menjadi indikator penting dalam kesuksesan seorang VTuber di platform seperti YouTube, yang merupakan penyedia layanan video terbesar di dunia dan lebih sering digunakan dibandingkan dengan platform video lainnya (Budzinski dkk., 2021).

Sejumlah penelitian telah dilakukan dalam mempelajari fenomena tentang VTuber. Pada studi yang dilakukan oleh Normah dkk. (2022), dilakukan penelitian yang menganalisis komentar di Twitter yang membahas tentang VTuber di Indonesia. Komentar yang dikumpulkan dianalisis dengan menggunakan algoritma *Support Vector Machine* (SVM), yang menghasilkan akurasi 88,18%. Temuan yang didapatkan menunjukkan 89,09% komentar menunjukkan sentimen positif terhadap fenomena VTuber di Indonesia. Penelitian lain yang dilakukan oleh Ameira dkk. (2023), bertujuan untuk menemukan korelasi antara jumlah *views* dan video terhadap jumlah *subscribers* dengan menggunakan metode SEM-PLS (*Structural Equation Modelling – Partial Least Squares*). Hasil penelitian tersebut memberikan nilai koefisien regresi 0,610 dan 0,100 untuk jumlah *views* dan jumlah video. Hasil ini menunjukkan bahwa jumlah *views* memiliki pengaruh kuat terhadap jumlah *subscribers* dan jumlah video hanya berpengaruh sedang. Penelitian-penelitian ini menunjukkan bahwa aspek performa dan interaksi penonton terhadap VTuber telah menjadi topik penelitian yang relevan, namun fokus dari penelitian masih terbatas pada analisis korelasional atau perilaku penonton.

Selain studi tentang VTuber, berbagai penelitian mengenai prediksi *engagement* telah dilakukan pada konteks yang lebih umum. Dalam melakukan prediksi terkait tingkat *engagement* suatu video, model *Random Forest* menunjukkan performa yang luar biasa bersama model *ensemble* seperti *Stacked Trees* jika dibandingkan dengan model lain seperti SVM, KNN, *Decision Tree* dan *Naive Bayes*. Model *Random Forest* menunjukkan rata-rata akurasi sekitar 90% (Wassermann, dkk. 2019). Hal ini didukung dari penelitian lain yang berfokus pada prediksi tingkat *engagement* user melalui *Click-Through Rates* (CTR). *Random Forest* menjadi model dengan tingkat akurasi tertinggi kedua setelah *Recurrent Neural Networks* (RNNs) mencapai 91% (Kamal & Bablu, 2022). Pada penelitian yang lain, *Random Forest* menunjukkan performa yang sama dengan model *XGBoost* memiliki rentang AUC (*Area Under the Curve*) dari 0,83 – 0,93. Penelitian-penelitian tersebut menunjukkan bahwa *Random Forest* algoritma yang cocok untuk tugas prediksi berbasis interaksi pengguna.

Meskipun penelitian mengenai VTuber dan prediksi *engagement* telah banyak dilakukan, belum terdapat penelitian yang secara khusus membangun model prediksi untuk mengukur tingkat *engagement* video VTuber, khususnya pada wilayah Indonesia. Penelitian terdahulu lebih banyak berfokus pada analisis sentimen, hubungan antar metrik, atau evaluasi performa kreator, tetapi belum mengeksplorasi bagaimana sebuah model dapat mengklasifikasikan tingkat *engagement* berdasarkan karakteristik video VTuber. Selain itu, belum ada penelitian yang membandingkan performa fitur-fitur yang memengaruhi *engagement* melalui pendekatan pemodelan prediktif. Kekosongan inilah yang menjadi dasar pentingnya pengembangan model prediksi *engagement* untuk VTuber.

Berdasarkan pemaparan permasalahan di atas, penting untuk mengembangkan model yang mampu memprediksi tingkat *engagement* video VTuber sebagai dasar analisis performa konten. Oleh karena itu, penulis menyusun skripsi dengan judul “Prediksi Tingkat *Engagement* Video VTuber di Indonesia Menggunakan Algoritma *Random Forest*”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, rumusan masalah yang dapat diambil adalah sebagai berikut :

1. Bagaimana konfigurasi model terbaik pada prediksi tingkat *engagement* video VTuber menggunakan algoritma *Random Forest*.
2. Bagaimana performa model *Random Forest* dalam memprediksi tingkat *engagement* video VTuber?
3. Fitur mana yang memiliki pengaruh paling besar terhadap hasil prediksi *engagement*?

1.3 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut :

1. Mengetahui konfigurasi model terbaik pada prediksi tingkat *engagement* video VTuber menggunakan algoritma *Random Forest*.
2. Menganalisis performa model *Random Forest* dalam memprediksi tingkat *engagement* video VTuber.
3. Menentukan fitur-fitur yang memberikan kontribusi paling signifikan terhadap hasil prediksi model.

1.4 Manfaat

Penelitian ini memiliki manfaat baik secara langsung maupun tidak langsung dalam perkembangan VTuber di Indonesia. Adapun manfaat dari penelitian ini adalah sebagai berikut :

a. Manfaat Teoritis

Penelitian ini memberikan kontribusi bagi pengembangan literatur mengenai analisis performa konten digital, khususnya dalam ekosistem VTuber di Indonesia. Selain itu, penelitian ini memperluas referensi ilmiah mengenai penerapan algoritma *Random Forest* dalam membangun model prediksi tingkat *engagement*, sehingga dapat menjadi dasar bagi studi-studi berikutnya yang berfokus pada prediksi performa konten atau analitik media sosial.

b. Manfaat Praktis

Penelitian ini menyediakan wawasan berbasis data bagi VTuber, agensi, serta pelaku industri konten digital mengenai faktor-faktor yang paling berpengaruh terhadap tingkat *engagement*. Temuan mengenai *feature importance* memberikan indikasi fitur mana yang dapat dioptimalkan untuk meningkatkan performa konten. Selain itu, model prediksi yang dapat dihasilkan dapat dimanfaatkan untuk mendukung perencanaan strategi konten secara lebih terstruktur.

1.5 Ruang Lingkup

Dataset yang digunakan adalah *dataset* video VTuber Indonesia dari tahun 2020 – 2024 di platform YouTube dengan database VTuber Asia sebagai acuannya. Video yang dianalisis diambil dari 50 *channel* VTuber yang diurutkan berdasarkan jumlah *views*. Pengumpulan data dilakukan dengan teknik *web scrapping* dengan menggunakan python yang diambil pada tanggal 3 Juli 2025. Variabel yang dijadikan ukuran adalah jumlah tayangan, jumlah *likes*, jumlah komentar, durasi video, waktu upload, jenis video, apakah *livestream*, dan agensi. Model yang digunakan untuk melakukan prediksi adalah dengan menggunakan algoritma *Random Forest*.

1.6 Sistematika Penulisan

Adapun sistematika penulisan penelitian ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Bab ini merupakan bab pendahuluan yang terdiri dari : Latar belakang, Rumusan masalah, Tujuan, Manfaat, dan Sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Bab ini menjelaskan tentang landasan teori yang berkaitan dengan pembahasan masalah yang diteliti. Bab tinjauan pustaka terdiri dari : Definisi Virtual YouTuber, Penelitian terkait Virtual YouTuber, Definisi YouTube *engagement*, Algoritma *Random Forest* yang terdiri dari Mekanisme *Resampling*, Skema *Splitting*, dan Evaluasi Kinerja, *One-Hot-Encoding*, serta *Interquartile Range*.

BAB III : METODOLOGI PENELITIAN

Bab metodologi penelitian berisikan tentang pengambilan data dengan menggunakan YouTube API, *Data Pre-processing* yang terdiri dari proses *data cleaning*, *data transformation*, *feature selection*, dan *data splitting*, serta implementasi dari algoritma *Random Forest* yang terdiri dari penentuan target variabel, pengaturan *hyperparameter*, pelatihan model, dan evaluasi model.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang deskripsi data, pembangunan model, dan hasil pembahasan implementasi dari algoritma *Random Forest* pada *dataset* VTuber Indonesia.

BAB V : KESIMPULAN DAN SARAN

Bab ini berisikan tentang kesimpulan yang didapatkan dari penelitian serta saran untuk penelitian selanjutnya.