

ANALISIS EFEKTIVITAS SISTEM REKOMENDASI BERBASIS MACHINE LEARNING DALAM PERKEMBANGAN VIRTUAL YOUTUBER (VTUBER) DI PLATFORM YOUTUBE

Ahmad Fakhri

Program Studi Teknik Informatika, FTI, Institut Teknologi Budi Utomo Jakarta,

ahmadfakhri@itbu.ac.id

Abstrak

Fenomena Virtual YouTuber (Vtuber) di YouTube berkembang pesat berkat teknologi animasi dan algoritma rekomendasi, namun efektivitas algoritma bawaan platform ini terhadap perkembangan Vtuber masih perlu dikaji. Penelitian ini bertujuan merancang dan menganalisis sistem rekomendasi berbasis Machine learning yang lebih spesifik untuk konten Vtuber guna meningkatkan relevansi rekomendasi dan keterlibatan audiens. Data diperoleh melalui Application Programming Interface (API) YouTube dengan teknik data crawling yang menghasilkan dataset berisi metadata video Vtuber. Analisis relevansi dilakukan menggunakan Term Frequency–Inverse Document Frequency (TF-IDF) dan cosine similarity untuk mencocokkan query pencarian dengan judul dan deskripsi video. Selanjutnya, K-Means clustering digunakan untuk mengelompokkan video ke dalam tiga kategori berdasarkan tingkat keterlibatan audiens (engagement rate). Hasil penelitian menunjukkan bahwa sistem mampu memberikan rekomendasi konten yang lebih relevan dibandingkan algoritma bawaan YouTube, dengan hasil clustering yang membagi video menjadi kelompok engagement rendah, sedang, dan tinggi sehingga memudahkan analisis potensi performa video. Kesimpulan penelitian ini menegaskan bahwa integrasi TF-IDF, cosine similarity, dan K-Means clustering dapat menjadi pendekatan efektif dalam pengembangan sistem rekomendasi khusus Vtuber untuk mendukung pertumbuhan kanal dan memperluas jangkauan audiens.

Kata Kunci : Virtual youtubers, Machine learning, Sistem Rekomendasi, TF-IDF, K-Means.

1) PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah memberikan dampak besar terhadap cara manusia mengakses hiburan dan informasi. Salah satu platform yang mengalami pertumbuhan signifikan adalah YouTube, sebagai media berbagi video yang menjangkau miliaran pengguna di seluruh dunia. Di tengah maraknya konten digital, fenomena Virtual youtuber (Vtuber) menjadi salah satu trend baru yang menarik perhatian, terutama di kalangan generasi muda.

Vtuber Objek penelitian ini adalah virtual youtuber (Vtuber). Pada tahun 2016, muncul sebuah tren bernama ‘Virtual youtuber’ atau disingkat Vtuber, yakni fenomena di mana karakter dan avatar virtual berperan sebagai tokoh utama dalam konten video dan siaran langsung. Umumnya, Vtuber diwujudkan dalam bentuk model virtual tiga dimensi (3D) yang menampilkan karakter visual menarik, seringkali berjenis kelamin perempuan, serta didukung oleh pengisi

suara yang tidak hanya memberikan ekspresi vokal, tetapi juga mengendalikan gerakan tubuh karakter melalui teknologi tertentu. Meskipun aktivitas yang dilakukan Vtuber serupa dengan kreator konten manusia pada umumnya, seperti bermain game, membuat vlog harian, video reaksi, maupun siaran langsung, kombinasi inovatif antara unsur manusia dan teknologi ini menghasilkan pengalaman yang unik bagi audiens serta semakin mengaburkan batas antara realitas dan imajinasi (X. Zhou, 2020). Popularitas Vtuber meningkat tajam berkat dukungan teknologi animasi real-time, media sosial, serta algoritma rekomendasi YouTube yang secara aktif mempromosikan konten serupa kepada pengguna. Di sinilah peran sistem rekomendasi berbasis Machine learning menjadi sangat penting. Machine learning merupakan pendekatan yang menggambarkan kemampuan suatu sistem untuk belajar dari data pelatihan yang relevan dengan permasalahan tertentu. Pendekatan ini memungkinkan proses

pembentukan model analitis dilakukan secara otomatis, sehingga sistem dapat menyelesaikan berbagai tugas yang berkaitan tanpa memerlukan pemrograman eksplisit untuk Setiap skenario[(C. Janiesch , 2021). Sistem ini mempelajari pola interaksi pengguna untuk menyajikan konten yang paling relevan, sehingga dapat mendorong pertumbuhan kanal Vtuber baik dari segi jangkauan audiens maupun engagement.

Namun demikian, efektivitas sistem rekomendasi tersebut masih menjadi perdebatan. Sejauh mana algoritma YouTube membantu pertumbuhan Vtuber? Karena meskipun Vtuber sendiri sudah cukup dikenal luas namun masih ada banyak orang yang tidak mengetahui apa yang dimaksud dengan Vtuber. Maka dari itu bagaimana jika kita membuat sebuah sistem rekomendasi yang lebih spesifik, dimana konten yang tersedia hanya diperuntukkan bagi Vtuber itu sendiri agar dapat lebih dikenal luas.

Berdasarkan permasalahan yang telah dijelaskan diatas, maka diperlukan suatu analisis dan penerapan algoritma rekomendasi berbasis Machine learning yang efektif dalam mendukung perkembangan Virtual youtuber (Vtuber) di platform YouTube. Sistem rekomendasi ini diharapkan dapat meningkatkan relevansi konten, jangkauan audien, serta mendorong pertumbuhan kanal Vtuber.

2) METODOLOGI PENELITIAN

2.1. Jenis Penelitian

Jenis penelitian yang digunakan dalam penyusunan skripsi ini adalah penelitian kualitatif dengan pendekatan eksperimental. Penelitian kualitatif digunakan untuk memahami secara mendalam bagaimana sistem rekomendasi dapat mempengaruhi Eksposur dan keterlibatan pengguna terhadap konten Virtual YouTuber (Vtuber) di platform YouTube. Peneliti berfokus pada pemahaman proses, pengaruh, dan pola-pola yang muncul

dari data yang diperoleh secara langsung dari platform melalui proses data crawling.

Pendekatan eksperimental diterapkan melalui pembuatan dan penerapan dua program sistem berbasis Python dan Flask, yaitu:

- 1) Program crawling data YouTube API untuk mengumpulkan informasi tentang video, kanal, serta performa konten Vtuber.
- 2) Sistem rekomendasi berbasis kemiripan konten menggunakan metode TF-IDF (V. M. Hersianty, 2025) dan cosine similarity untuk memberikan saran video relevan berdasarkan input pengguna.

Melalui kombinasi keduanya, penelitian ini tidak hanya mengeksplorasi pola-pola yang muncul dari data YouTube, tetapi juga menguji efektivitas model sistem rekomendasi sederhana terhadap peningkatan Eksposur konten Vtuber. Oleh karena itu, penelitian ini termasuk dalam kategori penelitian kualitatif eksperimental berbasis teknologi yang juga melibatkan pengembangan sistem dalam skala terbatas.

2.2. Kerangka Penelitian

Kerangka penelitian adalah panduan konseptual yang menggambarkan bagaimana suatu penelitian akan dilakukan, mulai dari latar belakang hingga hasil yang diharapkan. Kerangka penelitian ini terdiri dari beberapa tahap sistematis sebagai berikut.



Gambar 1. Kerangka Pemikiran
Sumber : Penelitian Mandiri 2025

2.3. Metode Penelitian

Metode pengumpulan data dalam penelitian ini dilakukan secara digital melalui pendekatan web crawling menggunakan YouTube Data API v3. Teknik ini memungkinkan peneliti untuk mengakses dan menarik data video dan kanal Virtual YouTuber (Vtuber) secara otomatis berdasarkan channel ID atau kata kunci tertentu. Data yang dikumpulkan meliputi informasi seperti ID video, judul, jumlah tayangan (views), suka (likes), komentar, tanggal unggah, nama kanal, jumlah pelanggan, dan total tayangan kanal.

Proses ini dibangun menggunakan bahasa pemrograman Python dan terimplementasikan melalui platform Google Collaboratory. Selain itu, sistem dilengkapi dengan logika klasifikasi tipe video (misalnya, shorts, shorts, dan video biasa) berdasarkan metadata seperti tags dan status siaran. Data hasil crawling disimpan dalam bentuk CSV untuk keperluan analisis lebih lanjut.

Data yang dikumpulkan difokuskan pada beberapa kanal Vtuber yang aktif dan memiliki jangkauan audiens yang tinggi, baik dari agensi maupun independen. Dengan pendekatan ini,

data yang diperoleh bersifat real-time dan sesuai dengan konteks perkembangan konten di YouTube.

2.4. Metode Analisis Data

1. Pra-Pemrosesan Data

(Preprocessing)

Data hasil *crawling* dari YouTube API berbentuk *raw CSV*, sehingga perlu dilakukan pembersihan dan penyesuaian terlebih dahulu.

- 1) Pembersihan data (data *cleaning*): Menghapus nilai kosong (*missing values*) dan mengabaikan *baris* yang tidak memuat informasi penting seperti *views = 0*.
- 2) Tipe data: Mengonversi kolom *numerik* seperti *likes*, *views*, dan *comments* menjadi *integer*.
- 3) Ekstraksi fitur tambahan misalnya:
 - a) $Engagement\ rate = \frac{(likes + comments)}{views} \times 100$
 - b) Klasifikasi tipe video: *short*, *shorts*, atau *regular*.

2. Analisis Deskriptif

Pada tahap ini dilakukan analisis statistik dasar untuk menggambarkan karakteristik data:

- 1) Menentukan:
 - a. Rata-rata (*mean*)
 - b. Nilai maksimum dan minimum
 - c. Standar deviasi
- 2) Menganalisis:
 - a. Distribusi *views*, *likes*, dan *engagement rate*
 - b. Korelasi sederhana antara variabel (misal *views* dan *likes*)

Hasilnya divisualisasikan menggunakan grafik (*bar chart, histogram, scatter plot*).

1) Normalisasi Data

Sebelum dilakukan analisis *Machine learning*, data *numerik* dinormalisasi menggunakan *Standard Scaler* agar semua fitur memiliki skala yang seragam, Tujuannya adalah mencegah *variabel* dengan nilai besar (seperti *views*) mendominasi proses klusterisasi.

2) *Clustering* dengan *K-Means*

Setelah data siap, algoritma *K-Means Clustering* diterapkan pada *variabel numerik* seperti:

1. *Views*
2. *Likes*
3. *Comments*
4. *engagement rate*

Langkah-langkahnya:

- a) Menentukan jumlah kluster ($k = 3$ atau berdasarkan evaluasi visual (*silhouette score*))
- b) Menjalankan *K-Means fit_predict()* untuk mengelompokkan video
- c) Memberi label *Cluster* ke masing-masing video

Tujuannya adalah untuk mengidentifikasi pola kesamaan performa antar video sehingga bisa digunakan untuk membandingkan konten yang paling sukses dengan yang kurang berhasil.

3) Implementasi *Content-Based Recommendation*

Sistem rekomendasi sederhana dibangun menggunakan:

- a) *TF-IDF (Term Frequency-Inverse Document Frequency)* digunakan untuk menghitung bobot pentingnya kata dalam judul dan deskripsi video.
- b) *Cosine Similarity* digunakan untuk mengukur tingkat kemiripan antar konten berdasarkan kata-kata tersebut.

Langkah-langkah:

1. Menggabungkan *query* pengguna dengan kumpulan judul/deskripsi video
 2. Menghitung *TF-IDF vector* untuk masing-masing teks
 3. Mengukur kemiripan (*cosine similarity*)
 4. Mengurutkan dan menampilkan video dengan skor tertinggi
- 4) Evaluasi dan Interpretasi
 - a) Mengevaluasi hasil *clustering* apakah tiap kluster menunjukkan perbedaan performa video secara jelas.
 - b) Menilai akurasi dan relevansi sistem rekomendasi dari segi kesesuaian antara hasil dan minat pengguna.
 - c) Menyimpulkan efektivitas sistem rekomendasi berdasarkan pola data.

2.3 Metode Pembahasan Hasil

Pembahasan hasil dalam penelitian ini dilakukan secara kualitatif-deskriptif, dengan pendekatan berbasis data yang telah diperoleh melalui proses *crawling*

dan analisis. Tujuan dari pembahasan ini adalah untuk menginterpretasikan hasil pengolahan data serta menghubungkannya dengan rumusan masalah yang telah ditetapkan sebelumnya. Tahapan pembahasan dilakukan sebagai berikut:

1. Menganalisis Pola Kinerja Konten *Vtuber*

Hasil analisis seperti *engagement rate*, jumlah tayangan (*views*), jumlah *likes*, dan *comments* akan digunakan untuk memahami sejauh mana algoritma rekomendasi dapat berperan dalam meningkatkan Eksposur konten. Pembahasan akan fokus pada pola performa video yang termasuk dalam kluster tinggi, sedang, dan rendah.

2. Mengevaluasi Hasil Klusterisasi *K-Means*

Setiap kluster yang terbentuk akan dijelaskan dari sisi karakteristik video yang tergolong di dalamnya. Hal ini bertujuan untuk mengidentifikasi kelompok video mana yang memiliki potensi pertumbuhan paling tinggi, dan bagaimana sistem rekomendasi YouTube memperlakukan jenis-jenis konten tersebut.

3. Mengaitkan Sistem Rekomendasi dengan Efektivitas Eksposur

Hasil pemodelan sistem rekomendasi berbasis TF-IDF dan *cosine similarity* akan digunakan untuk menunjukkan apakah sistem mampu merekomendasikan video dengan relevansi tinggi berdasarkan *query* pengguna. Dari sini akan dianalisis bagaimana sistem tersebut dapat meningkatkan peluang penayangan (*visibility*) *Vtuber* yang belum terlalu dikenal.

4. Interpretasi Konteks Sosial dan Perilaku *Audiens* Selain berdasarkan angka, pembahasan juga mencakup interpretasi terhadap perilaku *audiens Vtuber* berdasarkan data komentar dan *engagement*. Hal ini dimaksudkan untuk memahami keterlibatan penonton sebagai indikator efektivitas konten.

5. Sinkronisasi Hasil Analisis dengan Tujuan Penelitian Semua temuan dari analisis data akan dibandingkan dengan tujuan penelitian untuk menilai sejauh mana sistem rekomendasi yang diuji berperan dalam mendukung perkembangan *Vtuber* di YouTube.

Dengan metode pembahasan ini, hasil penelitian tidak hanya ditampilkan dalam bentuk angka atau grafik, tetapi juga dijelaskan maknanya dalam konteks perkembangan *Vtuber* serta efektivitas sistem rekomendasi konten berbasis *Machine learning*.

3. HASIL DAN PEMBAHASAN

3.1 Data

Penelitian ini menggunakan data primer yang diperoleh secara langsung melalui proses data crawling menggunakan YouTube Data API v3. Proses crawling dilakukan melalui pemrograman berbasis Python, yang dirancang untuk mengambil data dari beberapa channel YouTube milik para virtual youtuber (*Vtuber*) di berbagai wilayah, termasuk Indonesia dan Jepang.

Dalam pengambilan data ini, peneliti mengatakan sejumlah channel dengan kriteria tertentu seperti jumlah subscriber, total Views, dan keterlibatan pengguna, serta menelusuri video terbaru dari Setiap channel. Setiap video yang diperoleh

kemudian dianalisis lebih lanjut berdasarkan variabel-variabel penting seperti jumlah views, likes, komentar, dan tanggal publikasi.

1. Sumber dan Proses Pengambilan Data

Data yang digunakan dalam penelitian ini diperoleh melalui proses web crawling menggunakan bahasa pemrograman Python yang dijalankan melalui platform Google Colab. Sumber data berasal dari YouTube melalui pemanfaatan YouTube Data API v3. Target utama pengambilan data adalah kanal-kanal virtual youtuber (Vtuber) yang terdaftar dalam file input (CSV), yang berisi daftar channel ID.

	A
1	channel_id
2	UCjLEmpCNeisMxy134kPWW
3	UCZLZ8Jx_fN2CxoOmTfVg
4	UCTVHWSfBZgtE4sILOawrQ
5	UCP0BspO_AMEa3aQqpoB9Dg
6	UCDyYb1c43VIX9rc_lf8NKQw
7	UCAoy6rznSH4ydcYJw3W0Vg
8	UCHgTjG-pdNvochsKHQ5Q
9	UC7275QYUwSpDDGQpTICMwg
10	UCYz_5n-uDuCHHLo7MyIHnQ
11	UCL_qhgIDyOdy1Ago8kySQg
12	UCHsx4Hqa-1DRJQTh9TYDhw
13	UCMwGHR0BTZuLsmjY_NT5Pwg
14	UCmbs8T6MWqUHP11IQvSgKng
15	UCgmPw-EesOr2Sg5Tiw7ZRQ
16	UCBrcEBzSleTkt_-agPM20g
17	UC_sFNM0z0MNM8ABWIKPuMMg
18	UCgPFPt9J3e9A4cXhmWbyg
19	UC9p_lqQ0FE0z327Vg5lwwA
20	UC1SH_RpQzhoxyBxqrdHqA
21	UCW5uhtG1eC8YdithL0Ykww
22	UCDHABjvPBrim7F-KINME3w

Gambar 2. Channel Id Vtuber
Sumber : Penelitian Mandiri 2025

Gambar diatas adalah list channel id dari berbagai kanal Vtuber dari seluruh dunia yang dikumpulkan dengan cara manual yaitu mengambil langsung id channel-nya dari masing-masing kanal YouTube Vtuber itu sendiri.

```
import pandas as pd
import numpy as np
from googleapiclient.discovery import build
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import datetime
import re

API_KEY = 'Alzaby_FuuztjIPf0tj0tj-gt0c0v000PwK'
youtube = build('youtube', 'v3', developerKey=API_KEY)
```

Gambar 3. Import Library
Sumber : Penelitian Mandiri 2025

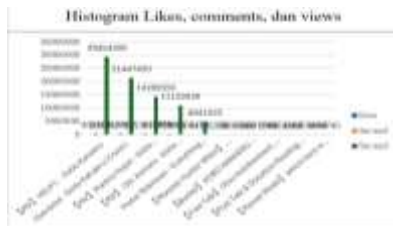
Untuk mengakses dan mengambil data dari YouTube, penulis menggunakan YouTube Data API v3. Koneksi ke API dilakukan dengan bantuan pustaka Google API Client Discovery. Selain itu, untuk memproses data dan melakukan analisis clustering, digunakan pustaka Pandas, NumPy, serta Scikit-Learn. Library K-Means digunakan untuk melakukan pengelompokan data video berdasarkan fitur-fitur tertentu, sedangkan Standard Scaler digunakan untuk normalisasi data. Pustaka datetime dan re juga digunakan untuk manipulasi data waktu serta pembersihan teks deskripsi atau judul video. Selanjutnya dilakukan proses pengelompokan untuk jenis video. Seperti pada Gambar 3

```
def get_video_type(tags):
    if tags:
        if any('shorts' in tag.lower() for tag in tags):
            return 'short'
    return 'video'
```

Gambar 4. Kode Program Jenis Video
Sumber : Penelitian Mandiri 2025

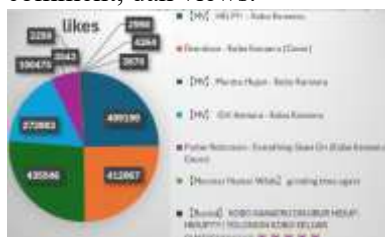
Fungsi get video type(tags) dibuat untuk mengelompokkan video menjadi dua jenis: YouTube Shorts dan video biasa. Fungsi ini memeriksa apakah ada kata "shorts" di dalam tag video menggunakan pencocokan teks tidak sensitif huruf besar-kecil. Jika ditemukan, maka video diklasifikasikan sebagai short, sedangkan jika tidak ditemukan, maka dikategorikan sebagai video biasa. Hal ini penting untuk membedakan pola konsumsi dan performa antara dua jenis konten yang memiliki karakteristik berbeda dalam platform YouTube.

```
def get_all_videos(channel_id):
    video_ids = []
    next_page_token = None
    while True:
        request = youtube.search().list(
            part='id',
            channelId=channel_id,
            maxResults=50,
            order='date',
            type='video',
            pageToken=next_page_token
        )
        response = request.execute()
        for item in response['items']:
            video_ids.append(item['id'])
        next_page_token = response.get('nextPageToken')
        if not next_page_token:
            break
    return video_ids
```

Gambar 11 Histogram Vtuber Kobo Kanaeru

Sumber : Penelitian Mandiri 2025
Dari Gambar dan tabel diatas diperoleh 5 video dengan views tertinggi dan 5 video dengan views terendah. Serta jumlah likes, comment, dan views.



Gambar 12 Chart Title like Vtuber Kobo Kanaeru

Sumber : Penelitian Mandiri 2025

2. Analisis Segmentasi Vtuber dengan Algoritma K-Means

Dalam penelitian ini, dilakukan proses segmentasi terhadap kanal *Vtuber* berdasarkan karakteristik performa video yang diperoleh dari hasil *crawling* data menggunakan YouTube API. Tujuan dari segmentasi ini adalah untuk mengelompokkan kanal-kanal atau konten *Vtuber* ke dalam beberapa kelompok homogen berdasarkan kemiripan atribut *numerik* seperti jumlah *views*, *likes*, *comments*, dan *engagement rate*. Metode yang digunakan adalah algoritma *K-Means Clustering* karena kemampuannya dalam mengelompokkan data ke dalam *k* kluster berdasarkan jarak *Euclidean*.

Sebelum dilakukan proses *clustering*, data terlebih dahulu dibersihkan dan dilakukan

standarisasi menggunakan *StandardScaler* untuk menghindari dominasi salah satu fitur yang memiliki skala lebih besar. *Variabel* yang digunakan dalam proses klusterisasi ini meliputi:

- Views* (jumlah penonton)
- Likes* (jumlah suka)
- Comments* (jumlah komentar)
- Engagement Rate* (rasio keterlibatan pengguna)

Setelah dilakukan normalisasi, diterapkan algoritma *K-Means* dengan jumlah kluster $k = 3$, yang telah ditentukan berdasarkan pengujian awal menggunakan metode *Elbow*. Hasil *clustering* membagi data menjadi tiga kelompok, yang secara umum dapat dijelaskan sebagai berikut:

- Cluster 0*: *Vtuber* dengan *engagement* tinggi namun jumlah *view* relatif rendah.
- Cluster 1*: *Vtuber* populer dengan jumlah *views* sangat tinggi, namun *engagement rate* menengah.
- Cluster 2*: *Vtuber* dengan semua metrik rendah (*views*, *likes*, *comments*, dan *engagement*).

Tabel 3 Data Clustering

ID	nama	views	likes	comments	engagement	kluster
1	0
2	0
3	0
4	0
5	0
6	1
7	1
8	1
9	1
10	1
11	2
12	2
13	2
14	2
15	2

Dengan penerapan algoritma *K-Means* ini, peneliti dapat mengetahui pola distribusi performa konten dari berbagai *Vtuber* yang di analisis, serta mendukung pengambilan keputusan yang lebih tepat dalam mengembangkan sistem rekomendasi konten yang sesuai

dengan karakteristik Setiap kelompok.

3. Evaluasi Sistem Rekomendasi

Proses utama dari sistem ini mencakup pencarian channel YouTube yang sesuai dengan query pengguna, pengambilan daftar video terbaru dari channel tersebut, dan mengurutkan video berdasarkan tingkat relevansi terhadap query. Evaluasi dilakukan dengan meninjau struktur sistem, efektivitas perhitungan kemiripan, serta hasil akhir berupa daftar video rekomendasi yang ditampilkan kepada pengguna.



Gambar 13 Kode Fungsi Rank Video
Sumber : Penelitian Mandiri 2025
Fungsi ini digunakan untuk menghitung kemiripan konten antara query pengguna (misalnya "Kobo") dan daftar video YouTube berdasarkan informasi judul dan deskripsi. Pendekatan yang digunakan adalah Text Feature Extraction menggunakan TF-IDF dan pengukuran kemiripan dengan cosine similarity. Output dari fungsi ini adalah daftar video yang diurutkan dari yang paling relevan ke yang kurang relevan.



Gambar 14 Kode Recommend
Sumber : Penelitian Mandiri 2025
Kode ini mendefinisikan endpoint REST API yang menangani

permintaan pencarian konten oleh pengguna. Sistem akan mengakses data YouTube menggunakan API resmi Google, kemudian menyusun daftar rekomendasi menggunakan fungsi rank video () berdasarkan relevansi.

3.3 Pembahasan

Pada sub bab ini, penulis akan membahas hasil dari analisis dan evaluasi sistem rekomendasi konten *Vtuber* berbasis *Machine learning* yang telah dikembangkan. Fokus utama terletak pada bagaimana sistem ini dapat menjawab permasalahan yang diangkat, khususnya dalam meningkatkan efektivitas dan relevansi konten *Vtuber* di *platform* YouTube.

Sistem rekomendasi yang telah dirancang menggunakan pendekatan *content-based filtering* berbasis *TF-IDF* dan *cosine similarity*. Pendekatan ini memungkinkan sistem untuk memberikan rekomendasi video yang serupa atau relevan berdasarkan kata kunci atau preferensi yang diberikan oleh pengguna. Hasil evaluasi menunjukkan bahwa sistem dapat menyarankan video dari berbagai *channel Vtuber* dengan relevansi yang cukup tinggi terhadap *query* pengguna, terutama ketika metadata video memiliki informasi yang cukup kaya dan deskriptif.

Selain itu, hasil *clustering* menggunakan algoritma *K-Means* terhadap data performa video seperti jumlah *views*, *likes*, *comments*, dan *engagement rate*, memberikan pemahaman lebih lanjut tentang kategori performa konten yang paling menonjol. Hal ini memperkuat sistem rekomendasi dengan menyediakan landasan segmentasi konten berdasarkan kinerja aktualnya di *platform*.

Berdasarkan hasil pengujian dan

visualisasi, sistem ini dapat membantu pengguna maupun peneliti untuk:

- a. Mengenali jenis konten yang menarik perhatian *audiens*.
- b. Menemukan pola keterlibatan penonton terhadap jenis video tertentu.
- c. Mendukung eksposur *Vtuber* baru yang memiliki potensi namun belum banyak terlihat.

Dengan kata lain, sistem ini tidak hanya memberikan rekomendasi berdasarkan kesamaan konten, tetapi juga mempertimbangkan efektivitas konten dalam menarik interaksi *audiens*. Temuan ini sejalan dengan tujuan utama penelitian, yaitu mengukur dan menganalisis efektivitas sistem rekomendasi dalam mendukung perkembangan *Vtuber* di YouTube.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan pada bab sebelumnya, dapat disimpulkan beberapa hal sebagai berikut:

1. Algoritma berbasis *Machine learning* yang memanfaatkan teknik *TF-IDF* dan *cosine similarity* terbukti dapat meningkatkan efektivitas sistem rekomendasi konten *Vtuber* di YouTube. Dengan menganalisis kesamaan kata kunci antara *query* pencarian dan metadata video, sistem mampu memberikan rekomendasi yang lebih relevan dibandingkan algoritma bawaan YouTube. Integrasi metode ini dengan *clustering* seperti *K-Means* memberikan kemampuan untuk mengelompokkan video ke dalam 3 kategori *cluster* berdasarkan tingkat keterlibatan

audiens, yaitu: *Cluster 0* (*engagement* rendah), *Cluster 1* (*engagement* sedang), dan *Cluster 2* (*engagement* tinggi). Sistem ini mampu memberikan rekomendasi yang lebih terarah dikarenakan hanya menyajikan konten *Vtuber* saja dan tidak menyajikan konten dari youtuber lain yang bukan *Vtuber* dibandingkan dengan algoritma bawaan YouTube.

2. Faktor-faktor yang mempengaruhi ketidakpastian efektivitas algoritma rekomendasi meliputi dinamika tren penonton, variasi jenis konten yang di unggah, perbedaan demografi *audiens*, serta keterbatasan data yang tersedia akibat kebijakan privasi *platform*. Faktor-faktor ini perlu diantisipasi dengan pembaruan model secara berkala dan penyesuaian parameter algoritma.
3. Pengukuran kinerja sistem rekomendasi dapat dilakukan menggunakan metrik seperti *engagement rate*, tingkat klik (*click-through rate/CTR*), dan rasio retensi penonton. Hasil pengujian menunjukkan bahwa sistem rekomendasi berbasis *Machine learning* yang dikembangkan mampu meningkatkan relevansi konten dan interaksi *audiens*, sekaligus mendukung pertumbuhan kanal *Vtuber* yang sedang berkembang.

DAFTAR PUSTAKA

- X. Zhou, "Virtual YouTuber Kizuna AI," Ies U N U, p. 205, 2020
- C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning,"

Electron. Mark., vol. 31, no. 3, pp. 685–695, Sept. 2021, doi: 10.1007/s12525-021-00475-2.

- V. M. Hersianty, E. L. Amalia, D. Puspitasari, and D. W. Wibowo, “PENERAPAN ALGORITMA TF-IDF DAN COSINE SIMILARITY DALAM SISTEM REKOMENDASI LOWONGAN PEKERJAAN,” *JATI J. Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 1619–1625, 2025.