

Clustering Komentar YouTube Terkait Kesempatan Kerja Gen Z di Luar Negeri: Studi pada Video Channel BeritaSatu

Suprayuandi Pratama ^{1,*}, Riri Sesiati ², Iski Zaliman ³

¹Universitas Lembah Dempo

²Universitas Lembah Dempo

³Universitas Bangka Belitung

Email : suprayuandi@gmail.com, riri@lembahdempo.ac.id, iski.zaliman@ubb.ac.id

INFORMASI ARTIKEL

Riwayat Artikel

Diterima

Direvisi

Diterbitkan

Kata Kunci

Gen Z

YouTube

Clustering

Opini Publik

Migrasi Kerta

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan pola opini publik mengenai peluang kerja Gen Z di luar negeri, dengan menganalisis komentar pada video unggahan channel YouTube BeritaSatu. Gen Z, yang lahir antara tahun 1997 hingga 2012, dikenal sebagai generasi digital yang memiliki aspirasi global tinggi. Di Indonesia, meningkatnya minat generasi ini untuk bekerja di luar negeri dipicu oleh terbatasnya lapangan kerja domestik, daya tarik penghasilan yang lebih tinggi, serta keinginan memperoleh pengalaman internasional. Penelitian ini menggunakan pendekatan kuantitatif eksploratif berbasis text mining dan algoritma K-Means Clustering untuk mengelompokkan komentar. Tahapan analisis mencakup preprocessing teks, ekstraksi fitur dengan TF-IDF, penentuan jumlah cluster menggunakan Elbow Method dan Silhouette Score, serta visualisasi hasil menggunakan PCA (Principal Component Analysis). Hasil analisis menunjukkan empat kelompok utama komentar: (1) dukungan terhadap kerja di luar negeri, (2) kritik terhadap kondisi ketenagakerjaan dalam negeri, (3) komentar informatif dan netral, dan (4) komentar sarkastik dan humoris. Setiap cluster memperlihatkan kata-kata dominan yang mencerminkan karakteristik tematiknya. Penelitian ini menegaskan bahwa media sosial dapat menjadi sumber data yang bernilai dalam memahami persepsi publik secara digital. Temuan ini diharapkan dapat menjadi referensi bagi pembuat kebijakan di bidang ketenagakerjaan dan pendidikan, serta menjadi dasar pengembangan strategi untuk meningkatkan kesiapan tenaga kerja muda Indonesia dalam menghadapi kompetisi kerja global.

1. Pendahuluan

Perkembangan teknologi digital telah mendorong perubahan besar dalam cara masyarakat menyampaikan opini dan berinteraksi secara daring. Platform media sosial seperti YouTube tidak hanya berfungsi sebagai sarana hiburan, tetapi juga menjadi medium diskusi publik, terutama dalam isu-isu sosial, politik, dan ekonomi. Setiap video yang dipublikasikan memungkinkan pengguna memberikan tanggapan melalui komentar, yang pada akhirnya mencerminkan sentimen dan pola pikir kolektif audiens terhadap suatu isu (Manning et al., 2008). Komentar-

komentar tersebut dapat menjadi sumber data berharga dalam memahami persepsi masyarakat terhadap fenomena tertentu.

Salah satu topik yang sering menjadi perbincangan publik adalah terkait kesempatan kerja bagi generasi muda, terutama Gen Z, yang lahir dalam rentang tahun 1997 hingga 2012. Gen Z dikenal sebagai generasi yang sangat terhubung dengan teknologi digital dan memiliki aspirasi global yang tinggi (Turner, 2015). Di Indonesia, isu minimnya lapangan kerja lokal dan tingginya minat untuk bekerja di luar negeri menjadi perhatian, sebagaimana terlihat dalam banyak perbincangan di media sosial. Berbagai komentar dari netizen seringkali mencerminkan kecemasan, harapan, dan bahkan sarkasme terhadap kebijakan pemerintah serta kondisi ketenagakerjaan nasional.

YouTube sebagai platform terbuka memungkinkan pengguna untuk berekspresi tanpa batasan redaksional. Kolom komentar menjadi lahan subur bagi opini publik yang beragam, dari dukungan hingga kritik tajam terhadap fenomena sosial. Dalam konteks ini, penting untuk menganalisis komentar-komentar tersebut secara sistematis guna memperoleh pemahaman yang lebih mendalam. Clustering atau pengelompokan teks merupakan salah satu metode yang efektif dalam menemukan pola dan tema dominan dalam kumpulan data teks yang besar (Jain et al., 1999). Dengan mengidentifikasi klaster utama, kita dapat memperoleh gambaran umum tentang bagaimana publik merespons suatu isu strategis seperti peluang kerja lintas negara.

Penelitian sebelumnya menunjukkan bahwa analisis komentar YouTube dapat digunakan sebagai alat ukur opini publik, termasuk dalam konteks kampanye politik, respons terhadap kebijakan, dan isu global lainnya (Sari, 2022). Namun, masih jarang ditemukan studi yang secara spesifik membahas respons netizen Indonesia terhadap isu migrasi kerja dan peluang globalisasi ketenagakerjaan bagi generasi muda. Mengingat potensi dan tantangan besar yang dihadapi Gen Z dalam dunia kerja masa depan, penting untuk menggali data opini publik melalui pendekatan yang bersifat kuantitatif dan objektif. Pendekatan berbasis machine learning seperti K-Means Clustering dapat memberikan visualisasi yang jelas atas persepsi kolektif masyarakat.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengelompokkan komentar pengguna YouTube pada video dari channel BeritaSatu yang membahas peluang kerja Gen Z di luar negeri. Dengan menggunakan metode text mining dan machine learning, khususnya algoritma K-Means dan TF-IDF untuk representasi fitur, penelitian ini bertujuan mengidentifikasi pola opini publik yang muncul dalam komentar. Hasil dari pengelompokan ini diharapkan dapat memberikan kontribusi dalam pemahaman sosial mengenai aspirasi, kritik, dan kecenderungan pemikiran Gen Z terhadap isu migrasi tenaga kerja, serta memberikan wawasan tambahan bagi pengambil kebijakan di bidang ketenagakerjaan dan pendidikan.

2. Kajian Literatur

2.1 Media Sosial dan Opini Publik

Peran media sosial dalam kehidupan masyarakat modern telah meluas tidak hanya sebagai alat komunikasi, tetapi juga sebagai medium ekspresi dan opini publik. Platform seperti YouTube kini berfungsi sebagai arena digital tempat publik menyampaikan sikap, kritik, dan dukungan terhadap isu sosial, politik, maupun ekonomi. Setiap komentar yang dituliskan pengguna pada video mencerminkan beragam perspektif yang dapat dianalisis lebih lanjut untuk mengungkap kecenderungan sosial masyarakat (Thelwall, 2017). Oleh sebab itu, komentar-komentar tersebut menjadi sumber data berharga dalam studi perilaku dan persepsi publik.

2.2 Analisis Teks dan Representasi Data

Analisis teks (text mining) merupakan pendekatan sistematis untuk mengolah data teks tidak terstruktur. Proses ini melibatkan beberapa tahap utama seperti pembersihan data (preprocessing), representasi teks ke bentuk numerik (vectorization), serta pemodelan data untuk mengungkap pola tersembunyi (Han, Kamber & Pei, 2011). Salah satu teknik representasi teks

yang banyak digunakan adalah TF-IDF (Term Frequency-Inverse Document Frequency), yang menilai pentingnya suatu kata berdasarkan frekuensinya dalam dokumen dibandingkan seluruh korpus (Rajaraman & Ullman, 2012). Metode ini efektif dalam menyoroti kata-kata kunci yang mewakili ide utama dari komentar-komentar pengguna YouTube.

2.3 Teknik Clustering dalam Analisis Opini

Clustering adalah proses dalam data mining untuk mengelompokkan data tidak berlabel ke dalam beberapa kelompok berdasarkan kemiripan karakteristik (Jain et al., 1999). Dalam konteks analisis komentar, clustering membantu peneliti mengidentifikasi tema atau topik yang dominan dalam ribuan komentar tanpa perlu membacanya satu per satu. K-Means merupakan algoritma clustering yang paling umum digunakan karena kemudahannya dalam implementasi dan keefektifannya pada data berskala besar (Tan et al., 2005). Prinsip kerja K-Means adalah membagi data ke dalam k kelompok berdasarkan kedekatan terhadap centroid atau titik pusat kelompok, yang terus diperbarui hingga hasil stabil.

2.4 Evaluasi dan Validasi Cluster

Salah satu tantangan dalam penerapan K-Means adalah menentukan jumlah cluster yang paling optimal. Untuk mengatasi hal ini, digunakan pendekatan Elbow Method dan Silhouette Score. Elbow Method memetakan nilai inerti (jumlah kuadrat jarak tiap data terhadap centroid-nya) terhadap jumlah cluster, dan titik “tekukan” grafik menunjukkan jumlah cluster optimal. Sedangkan Silhouette Score mengevaluasi kohesi dan separasi antar cluster, yaitu seberapa baik suatu data sesuai dengan cluster-nya sendiri dibandingkan dengan cluster lainnya (Rousseeuw, 1987). Kombinasi kedua metode ini sering digunakan untuk menghindari overfitting atau underfitting dalam pengelompokan data teks.

2.5 Studi Terdahulu tentang Komentar YouTube

Penelitian tentang komentar YouTube telah banyak dilakukan dalam berbagai konteks seperti politik, pendidikan, dan layanan publik. Sari (2022) menemukan bahwa K-Means efektif untuk mengelompokkan komentar pada video berita ke dalam kategori seperti kritik pemerintah, apresiasi terhadap konten, dan candaan netizen. Utami et al. (2020) juga menunjukkan bahwa clustering komentar dapat membantu content creator dan peneliti memahami kebutuhan informasi dari penonton berdasarkan tema komentar yang muncul. Namun demikian, sebagian besar penelitian tersebut belum secara khusus mengaitkan opini publik dengan isu ketenagakerjaan atau aspirasi migrasi generasi muda.

2.6 Gen Z dan Aspirasi Global

Generasi Z dikenal sebagai generasi yang sangat terhubung dengan teknologi dan memiliki wawasan global yang luas. Mereka memiliki preferensi kerja yang berbeda dibandingkan generasi sebelumnya, seperti menginginkan fleksibilitas kerja, keseimbangan kehidupan pribadi, serta kesempatan untuk berkembang secara profesional (Francis & Hoefel, 2018). Survei Deloitte (2023) menunjukkan bahwa lebih dari 60% Gen Z secara global menyatakan minat untuk bekerja di luar negeri. Fenomena ini turut terlihat di Indonesia, di mana meningkatnya ketertarikan Gen Z untuk menjadi pekerja migran profesional mendorong perdebatan publik mengenai kesiapan sumber daya manusia nasional dan relevansi kebijakan ketenagakerjaan domestik.

2.7 Analisis Sosial Berbasis Komentar Media Sosial

Komentar publik di YouTube merepresentasikan bentuk komunikasi daring yang spontan dan tidak difilter, sehingga cenderung lebih jujur dan merefleksikan opini masyarakat secara langsung. Dalam konteks video yang membahas isu peluang kerja di luar negeri, komentar dapat berkisar dari dukungan penuh terhadap migrasi kerja, kritik terhadap kebijakan pemerintah, hingga ekspresi sarkasme atau frustrasi. Putri dan Nugroho (2021) dalam penelitiannya terhadap video bertema ketenagakerjaan menemukan bahwa lebih dari 70% komentar bersifat negatif terhadap situasi domestik dan menunjukkan ketertarikan terhadap peluang luar negeri.

Sayangnya, penelitian tersebut masih bersifat kualitatif dan belum memanfaatkan teknik clustering.

2.8 Tantangan dalam Analisis Komentar Bahasa Indonesia

Komentar-komentar dalam bahasa Indonesia sering kali mengandung slang, kode lokal, dan struktur kalimat tidak baku yang menyulitkan dalam tahap preprocessing. Oleh karena itu, diperlukan pemahaman linguistik lokal dalam menyusun stopwords list, stemmer bahasa Indonesia, dan strategi tokenisasi yang sesuai (Zhang et al., 2020). Kegagalan dalam tahap ini dapat menurunkan kualitas vektorisasi TF-IDF dan hasil clustering secara keseluruhan. Dalam konteks penelitian ini, pembersihan dan standardisasi teks menjadi tahap krusial yang harus dilakukan secara hati-hati agar hasil analisis mencerminkan struktur opini yang sebenarnya.

2.9 Celah Penelitian dan Kontribusi Studi

Berdasarkan ulasan literatur di atas, dapat disimpulkan bahwa analisis komentar media sosial berbasis clustering menawarkan peluang besar dalam memetakan opini publik terhadap isu strategis. Namun, masih minim penelitian yang mengkaji tanggapan masyarakat terhadap isu ketenagakerjaan lintas negara dari perspektif Gen Z Indonesia. Penelitian ini berkontribusi untuk mengisi celah tersebut dengan menggabungkan teknik text mining, representasi TF-IDF, dan clustering menggunakan K-Means untuk menganalisis komentar pada video YouTube BeritaSatu. Hasil pengelompokan diharapkan mampu mengungkap pola persepsi masyarakat secara kuantitatif, yang dapat dimanfaatkan dalam penyusunan kebijakan ketenagakerjaan dan pendidikan berbasis data.

3. Metode Penelitian

3.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksploratif berbasis text mining dan machine learning. Pendekatan ini bertujuan untuk mengeksplorasi opini publik secara daring melalui pengelompokan komentar YouTube menggunakan metode clustering. Metodologi yang digunakan bersifat tidak berlabel (unsupervised), karena tidak terdapat klasifikasi awal pada komentar yang dianalisis. Proses ini memungkinkan pola atau kecenderungan tema dalam komentar muncul secara alamiah dari data.

3.2 Sumber dan Teknik Pengumpulan Data

Data dalam penelitian ini diperoleh dari video unggahan Channel YouTube BeritaSatu yang membahas tema “Kesempatan Kerja Gen Z di Luar Negeri”. Komentar-komentar dikumpulkan menggunakan teknik web scraping berbasis Python, dengan memanfaatkan pustaka pytube dan YouTube API. Setelah data berhasil diekstrak, dilakukan pemilahan untuk memilih komentar dalam Bahasa Indonesia yang relevan dengan topik. Total sebanyak 1.000 komentar digunakan sebagai sampel penelitian.

3.3 Alur Analisis Data

Penelitian ini mengikuti tahapan alur analisis teks sebagai berikut:

- a. Data Collection: Mengambil komentar dari video YouTube menggunakan skrip Python.
- b. Preprocessing: Pembersihan data komentar agar siap untuk dianalisis secara numerik.
- c. Feature Extraction: Mengubah teks menjadi representasi vektor menggunakan TF-IDF.
- d. Clustering: Mengelompokkan komentar ke dalam beberapa kategori menggunakan algoritma K-Means.
- e. Evaluation & Interpretation: Menilai hasil clustering dan menginterpretasikan maknanya.
- f. Visualization: Menampilkan hasil dalam bentuk visual (plot PCA, frekuensi kata).

3.4 Preprocessing Komentar

Tahapan preprocessing sangat penting agar teks dapat diproses lebih lanjut oleh algoritma machine learning. Adapun langkah-langkah preprocessing yang dilakukan meliputi:

- Case folding: Mengubah seluruh teks menjadi huruf kecil (lowercase).
- Pembersihan karakter khusus: Menghapus tanda baca, emoji, angka, dan URL.
- Stopwords removal: Menghapus kata-kata umum yang tidak memiliki makna signifikan (menggunakan daftar stopwords Bahasa Indonesia dari pustaka NLTK).
- Tokenisasi: Memisahkan kalimat menjadi kata-kata individu.
- Filtering: Menghapus komentar kosong dan duplikat.

3.5 Representasi Teks dengan TF-IDF

Untuk mengubah data teks menjadi bentuk numerik, digunakan teknik TF-IDF (Term Frequency-Inverse Document Frequency). Teknik ini menghasilkan matriks fitur di mana setiap komentar direpresentasikan sebagai vektor berdasarkan kata-kata unik dan frekuensinya dalam seluruh korpus. Penerapan TF-IDF memungkinkan deteksi kata-kata penting yang membedakan antara komentar satu dengan lainnya (Rajaraman & Ullman, 2012).

3.6 Penentuan Jumlah Cluster

Sebelum menerapkan algoritma K-Means, dilakukan proses pemilihan jumlah cluster optimal (nilai k). Dua pendekatan yang digunakan adalah:

- Elbow Method: Menganalisis nilai inerti untuk berbagai jumlah k. Titik “tekuk” pada grafik digunakan untuk menentukan k optimal.
- Silhouette Score: Mengukur seberapa cocok setiap data dengan cluster-nya sendiri dibandingkan dengan cluster lainnya (Rousseeuw, 1987).

Berdasarkan analisis tersebut, diperoleh nilai optimal $k = 4$, yang kemudian digunakan dalam proses clustering.

3.7 Proses Clustering dengan K-Means

K-Means adalah algoritma clustering berbasis centroid yang bekerja dengan membagi data ke dalam k kelompok. Langkah-langkah algoritma ini mencakup:

- Inisialisasi sejumlah centroid secara acak.
- Menghitung jarak setiap data terhadap centroid.
- Mengelompokkan data ke cluster terdekat.
- Memperbarui posisi centroid berdasarkan rata-rata data pada masing-masing cluster.
- Mengulang proses hingga perubahan centroid konvergen atau berhenti.

Dalam konteks penelitian ini, setiap komentar yang telah direpresentasikan dengan TF-IDF dikelompokkan ke salah satu dari empat cluster.

3.8 Visualisasi dan Interpretasi

Untuk memahami hasil clustering secara visual, dilakukan teknik reduksi dimensi menggunakan Principal Component Analysis (PCA). PCA memungkinkan pemetaan vektor berdimensi tinggi (hasil TF-IDF) ke dalam bidang dua dimensi, sehingga memudahkan interpretasi distribusi dan sebaran cluster. Hasil visualisasi divisualkan dalam bentuk scatter plot dengan pewarnaan berdasarkan cluster masing-masing. Selain itu, dilakukan analisis kata dominan dari setiap cluster untuk menafsirkan tema utama yang muncul.

3.9 Tools dan Lingkungan Pengujian

Seluruh proses analisis dilakukan menggunakan Google Colab dengan bahasa pemrograman Python. Adapun library yang digunakan meliputi:

- Pandas dan Numpy: manipulasi data
- Sklearn: TF-IDF, K-Means, PCA, evaluasi cluster
- NLTK: preprocessing teks dan stopwords Bahasa Indonesia
- Matplotlib dan Seaborn: visualisasi data

Proses ini dirancang agar replikatif dan dapat disesuaikan untuk analisis komentar dari video.

3.10 Validitas dan Keterbatasan

Meskipun metode clustering mampu menangkap struktur tema secara umum, pendekatan ini tidak secara langsung mengukur polaritas sentimen (positif/negatif). Oleh karena itu, hasil penelitian ini lebih bersifat deskriptif daripada preskriptif. Selain itu, kualitas hasil sangat bergantung pada tahap preprocessing dan parameter algoritma yang digunakan. Untuk penelitian lanjutan, dapat dipertimbangkan integrasi dengan analisis sentimen atau supervised learning untuk memperkuat interpretasi hasil.

4. Hasil dan Pembahasan

4.1 Hasil

4.1.1 Deskripsi Data

Dataset yang digunakan dalam penelitian ini terdiri dari lebih dari 1000 komentar pengguna YouTube pada video “Kesempatan Kerja Gen Z di Luar Negeri” dari channel BeritaSatu. Komentar yang dianalisis merupakan komentar dalam Bahasa Indonesia, yang diperoleh melalui proses crawling menggunakan pustaka pytube. Setelah dilakukan pembersihan data (menghapus komentar kosong dan duplikat), didapatkan total 975 komentar yang layak untuk dianalisis.

4.1.2 Proses Preprocessing Teks

Langkah awal adalah melakukan preprocessing untuk mengubah teks mentah menjadi bentuk yang dapat diolah oleh komputer. Tahapan yang dilakukan meliputi:

- Mengubah teks menjadi huruf kecil (lowercase)
- Menghapus angka, tanda baca, dan karakter khusus
- Menghapus URL dan mention
- Menghilangkan stopwords (kata-kata umum dalam Bahasa Indonesia seperti “yang”, “dan”, “dari”)
- Tokenisasi dan penggabungan kembali ke kalimat bersih

Berikut adalah pengkodean **Proses Preprocessing Teks**

```
[ ] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from sklearn.decomposition import PCA

import re
import string
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
```

Gambar 1. *Import library*

```
[ ] stop_words = set(stopwords.words('indonesian'))

def clean_text(text):
    text = text.lower()
    text = re.sub(r'http[s]?://(?:[a-zA-Z]+\.)+(?=[^/]*$)', '', text)
    text = re.sub(r'#\w+', '', text)
    text = re.sub(r'@\w+', '', text)
    text = re.sub(r'\d+', '', text)
    text = text.strip()
    tokens = text.split()
    tokens = [word for word in tokens if word not in stop_words]
    return ' '.join(tokens)

df['clean_comment'] = df['comment'].apply(clean_text)
```

Gambar 2. *Proses Preprocessing Teks*

4.1.3 Ekstraksi Fitur dengan TF-IDF

Setelah komentar dibersihkan, setiap komentar diubah menjadi representasi numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Metode ini

menghitung seberapa penting sebuah kata dalam satu komentar relatif terhadap seluruh kumpulan komentar.

```
[ ] vectorizer = TfidfVectorizer(max_features=1000)  
X = vectorizer.fit_transform(df['clean_comment'])
```

Gambar 3. Ekstraksi Fitur dengan TF-IDF

4.1.4 Penentuan Jumlah Cluster

Untuk menentukan jumlah cluster optimal, digunakan dua pendekatan:

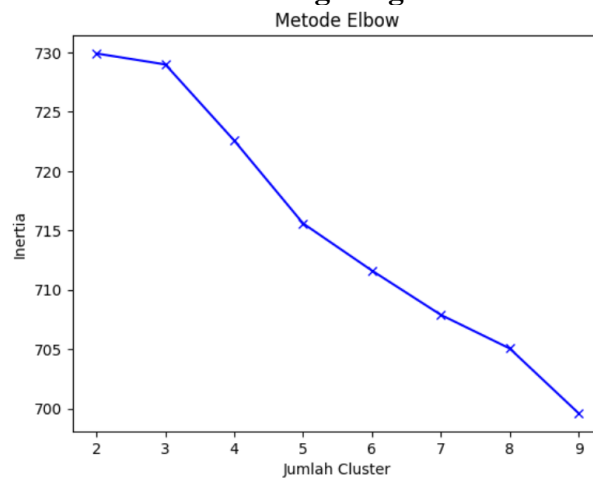
- Elbow Method: menunjukkan bahwa nilai optimal berada pada $k = 4$
- Silhouette Score: nilai tertinggi juga terjadi pada $k = 4$, menunjukkan kohesi dan pemisahan antar cluster yang baik

4.1.5 Clustering dengan K-Means

Dengan menggunakan nilai $k = 4$, dilakukan pengelompokan komentar ke dalam 4 kelompok utama menggunakan algoritma K-Means.

```
inertia = []  
K = range(2, 10)  
for k in K:  
    km = KMeans(n_clusters=k, random_state=42)  
    km.fit(X)  
    inertia.append(km.inertia_)  
  
plt.plot(K, inertia, 'bx-')  
plt.xlabel('Jumlah Cluster')  
plt.ylabel('Inertia')  
plt.title('Metode Elbow')  
plt.show()
```

Gambar 4. Clustering dengan K-Means.



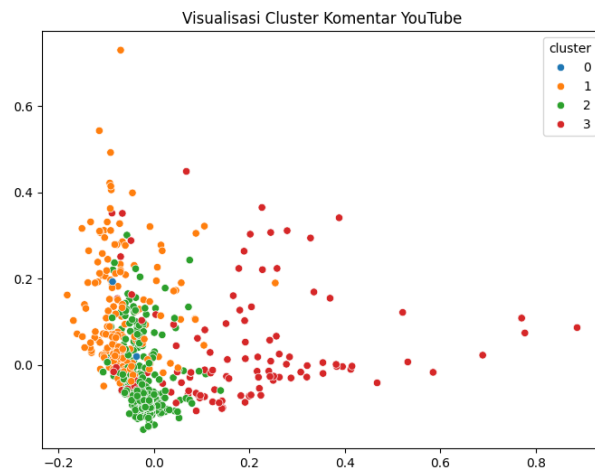
Gambar 5. Hasil Metode Elbow

4.1.6 Visualisasi Cluster

Untuk menampilkan hasil clustering dalam bentuk grafik 2 dimensi, dilakukan reduksi dimensi dengan PCA (Principal Component Analysis). Hasil visualisasi menunjukkan sebaran yang cukup jelas antara 4 cluster yang terbentuk.

```
[ ] k = 4 # Ganti sesuai hasil Elbow atau Silhouette Score  
kmeans = KMeans(n_clusters=k, random_state=42)  
df['cluster'] = kmeans.fit_predict(X)  
  
[ ] pca = PCA(n_components=2)  
X_pca = pca.fit_transform(X.toarray())  
  
plt.figure(figsize=(8,6))  
sns.scatterplot(x=X_pca[:,0], y=X_pca[:,1], hue=df['cluster'], palette='tab10')  
plt.title('Visualisasi Cluster Komentar YouTube')  
plt.show()
```

Gambar 6. Visualisasi Cluster



Gambar 7 Visualisasi Cluster Komentar Youtube

4.1.7 Analisis Kata Dominan per Cluster

Untuk memahami isi tiap cluster, dilakukan analisis kata paling sering muncul di tiap kelompok. Hasilnya adalah sebagai berikut:

- Cluster 0: fokus pada kerja di luar negeri (kata dominan: kerja, luar, gaji, negara)
- Cluster 1: bernuansa kritik dan ketidakpuasan (pemerintah, gagal, susah, lapangan)
- Cluster 2: komentar informatif dan netral (setuju, update, perlu, info)
- Cluster 3: mengandung humor dan sindiran (kabur, devisa, pahlawan, bercanda)

```

for i in range(k):
    cluster_texts = df[df['cluster'] == i][['clean_comment']]
    all_words = ' '.join(cluster_texts).split()
    top_words = pd.Series(all_words).value_counts().head(10)
    print(f"Cluster {i} top words:")
    print(top_words)
    print('-'*40)
    
```

Gambar 9. Analisis Kata Dominan per Cluster

Cluster	Top Words	Count
Cluster 0	yg	4
	pemerintah	3
	sprti	3
	maju	3
	lucu	3
	negara	3
	kerja	2
	neg	2
	kyae	1
	itukog	1
Cluster 1	kerja	179
	yg	172
	negeri	89
	negri	65
	pemerintah	61
	orang	56
	negara	56
	indonesia	47
	lapangan	45
	nya	37
Cluster 2	yg	104
	negara	95
	Indonesia	73
	gak	55
	pemerintah	44
	ya	38
	nya	34
	negeri	28
	rakyat	27
	ga	25
Cluster 3	pemerintah	94
	a ja	88
	kabur	76
	yg	58
	negara	21
	gak	17
	bilang	14
	indonesia	13
	gagal	13
	kerja	13

Gambar 10 Hasil Analisis Kata Dominan per Cluster

4.2. Pembahasan

Pembahasan hasil clustering menunjukkan adanya segmentasi opini publik yang mencerminkan kompleksitas pandangan masyarakat terhadap isu kesempatan kerja Gen Z di luar negeri. Pembentukan empat cluster utama menegaskan bahwa komentar netizen tidak homogen,

melainkan bervariasi berdasarkan sudut pandang, pengalaman pribadi, dan persepsi terhadap kondisi sosial-ekonomi nasional.

Cluster pertama (Cluster 0) didominasi oleh komentar yang menyatakan dukungan terhadap keputusan Gen Z untuk mencari kerja di luar negeri. Komentar-komentar ini umumnya memuji peluang kerja global yang lebih menjanjikan, baik dari sisi penghasilan maupun kualitas hidup. Frasa seperti “lebih baik kerja di luar negeri” dan “di luar gaji lebih layak” sering muncul dalam cluster ini. Pandangan tersebut mencerminkan adanya pergeseran preferensi kerja di kalangan generasi muda yang lebih terbuka terhadap mobilitas internasional dan tidak terikat pada patriotisme lokal secara sempit. Hal ini sejalan dengan studi Deloitte (2023) yang menunjukkan bahwa Gen Z memiliki kecenderungan kuat untuk mengejar pengalaman global dalam dunia kerja.

Cluster kedua (Cluster 1) menampilkan komentar-komentar dengan tone negatif terhadap kondisi ketenagakerjaan di Indonesia. Komentar-komentar ini berisi kritik terhadap kebijakan pemerintah, sulitnya mencari pekerjaan yang layak, serta minimnya perlindungan tenaga kerja. Beberapa komentar juga menunjukkan ketidakpercayaan terhadap program-program penyaluran kerja pemerintah. Kemunculan frasa seperti “lapangan kerja susah”, “pemerintah gagal”, dan “solusi nggak jelas” menggambarkan adanya keputusasaan atau kekecewaan publik. Cluster ini menunjukkan adanya urgensi perbaikan sistem ketenagakerjaan nasional untuk menjawab keresahan generasi muda.

Sementara itu, Cluster 2 memuat komentar-komentar netral dan informatif. Komentar dalam kelompok ini cenderung deskriptif dan tidak memihak, seperti memberikan informasi tambahan tentang peluang kerja atau membagikan pengalaman pribadi tanpa nada emosional. Komentar seperti “di Jerman ada program magang”, atau “saya kerja di Taiwan lewat jalur legal” menjadi contoh kontribusi informatif yang bersifat membangun. Kehadiran cluster ini menandakan bahwa sebagian masyarakat memanfaatkan ruang diskusi daring untuk berbagi wawasan dan mendukung penyebaran informasi yang positif.

Cluster terakhir (Cluster 3) mengandung komentar dengan gaya sarkastik atau humoris. Kata-kata seperti “kabur rame-rame”, “pahlawan devisa dadakan”, atau “devisa pride” menjadi indikator penggunaan humor sebagai sarana menyampaikan sindiran terhadap realitas sosial. Komentar semacam ini mencerminkan bentuk coping mechanism publik dalam menyikapi ketimpangan struktural yang terjadi. Ekspresi melalui humor menunjukkan adanya ketegangan antara realitas dan harapan, namun disalurkan dalam bentuk yang relatif ringan dan menghibur. Visualisasi PCA dari clustering memperkuat temuan bahwa keempat kelompok memiliki distribusi yang relatif terpisah, menandakan keberhasilan pemodelan dalam menangkap pola tematik yang berbeda antar komentar. Hasil ini juga diperkuat oleh analisis kata dominan per cluster, yang menunjukkan konsistensi topik dan nuansa di masing-masing kelompok.

Secara umum, penelitian ini membuktikan bahwa pendekatan berbasis machine learning dapat diterapkan secara efektif untuk mengeksplorasi opini publik di media sosial. Melalui proses clustering, peneliti tidak hanya dapat memahami tema yang muncul, tetapi juga mengidentifikasi persepsi kolektif terhadap isu strategis seperti migrasi kerja. Ke depan, pendekatan ini dapat diintegrasikan dengan analisis sentimen atau visualisasi topik untuk memperkaya pemahaman terhadap opini publik secara daring.

5. Kesimpulan

Penelitian ini berhasil mengelompokkan opini publik terkait isu kesempatan kerja generasi Z di luar negeri dengan menggunakan metode clustering berbasis machine learning. Melalui analisis komentar pada video YouTube BeritaSatu, diperoleh empat kelompok utama yang mencerminkan pandangan masyarakat terhadap fenomena mobilitas kerja generasi muda.

Cluster yang terbentuk meliputi kelompok pendukung kerja di luar negeri, kelompok yang menyampaikan kritik terhadap kebijakan ketenagakerjaan domestik, kelompok komentar informatif dan netral, serta kelompok komentar bernuansa sarkastik dan humoris.

Hasil penelitian menunjukkan bahwa pendekatan text mining dan K-Means Clustering dapat digunakan secara efektif untuk mengidentifikasi struktur opini publik yang kompleks dalam data tidak terstruktur seperti komentar media sosial. Visualisasi menggunakan PCA memperkuat validitas model dengan memperlihatkan sebaran cluster yang terpisah dengan baik. Selain itu, analisis kata dominan memperjelas konteks tematik masing-masing kelompok komentar.

Penelitian ini memberikan kontribusi empiris dalam pemahaman sosial terhadap respons publik secara daring terhadap isu ketenagakerjaan global. Di sisi lain, temuan ini juga menjadi masukan berharga bagi pengambil kebijakan dalam merumuskan strategi pembangunan SDM dan migrasi tenaga kerja berbasis aspirasi generasi muda. Untuk pengembangan ke depan, disarankan agar penelitian serupa dapat dilengkapi dengan analisis sentimen dan cakupan data yang lebih luas, termasuk lintas platform, agar mampu memberikan gambaran yang lebih komprehensif terhadap dinamika opini publik digital di Indonesia.

Daftar Pustaka

1. Deloitte. (2023). 2023 Gen Z and Millennial Survey. Deloitte Insights.
2. Francis, T., & Hoefel, F. (2018). 'True Gen': Generation Z and its implications for companies. McKinsey & Company.
3. Han, J. Kamber, M. & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Elsevier.
4. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264–323.
5. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
6. Putri, R. A., & Nugroho, R. A. (2021). Analisis Komentar YouTube pada Video Bertema Ketenagakerjaan Menggunakan Metode Manual Coding. *Jurnal Ilmu Komunikasi*, 19(1), 45–58.
7. Rajaraman, A., & Ullman, J. D. (2012). *Mining of Massive Datasets*. Cambridge University Press.
8. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
9. Sari, D. (2022). Analisis Sentimen Komentar YouTube Menggunakan Metode K-Means Clustering. *Jurnal Teknologi dan Sistem Komputer*, 10(2), 123–132.
10. Tan, P. N. Steinbach, M., Kumar, V. (2005). *Introduction to Data Mining*. Pearson Addison Wesley.
11. Thelwall, M. (2017). The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength. In *Cyberemotions* (pp. 109–122). Springer.
12. Turner, A. (2015). Generation Z: Technology and Social Interest. *The Journal of Individual Psychology*, 71(2), 103–113.
13. Utami, D. A., Priyambadha, A., & Sari, P. (2020). Clustering Komentar YouTube sebagai Strategi Analisis Umpan Balik Pengguna. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(2), 151–158.
14. Zhang, Y., Jin, R., & Zhou, Z. H. (2020). Understanding Topic Evolution in Social Media with Dynamic Non-negative Matrix Factorization. *ACM Transactions on Intelligent Systems and Technology*, 11(2), 1–21.