Penerapan NLP untuk Menganalisis Komentar Publik pada Pelantikan Menteri di Kanal Sekretariat Presiden

Application of NLP for Analyzing Public Comments on Ministerial Inauguration in the Sekretariat Presiden Channel.

Suprayuandi Pratama*[1], Yuniansyah[2], Welly Fransiska[2]
[1,2] Rekayasa Perangkat Lunak Universitas Lembah Dempo,Pagaralam Indonesia
[3] Sains Data Universitas Lembah Dempo,Pagaralam Indonesia
*e-mail: *[1]suprayuandi@gmail.com, [2] yuniansyah@lembahdempo.ac.id, *[1]welly@lembahdempo.ac.id*

Abstrak

Penelitian ini menerapkan *Natural Language Processing (NLP)* untuk menganalisis 465 komentar publik pada siaran pelantikan menteri 17 September 2025 di kanal YouTube Sekretariat Presiden. Tujuan penelitian adalah mengidentifikasi distribusi sentimen, topik dominan, dan kata kunci yang merefleksikan opini masyarakat. Data dikumpulkan melalui scraping, kemudian diproses melalui tahap pembersihan teks, representasi TF-IDF, analisis frekuensi n-gram, analisis sentimen berbasis leksikon, dan pemodelan topik *Latent Dirichlet Allocation* (LDA). Hasil menunjukkan mayoritas komentar bersentimen netral (94,2%), dengan topik utama mencakup kepemimpinan, isu pergantian pejabat, dan persepsi publik terhadap kinerja menteri baru. Temuan ini menegaskan efektivitas metode NLP dalam memahami opini politik digital dan memberikan masukan strategis bagi pemerintah dalam merancang komunikasi publik yang responsif.
Kata kunci — *Natural Language Processing,* Opini Publik, Analisis Sentimen

Abstrak

This study applies Natural Language Processing (NLP) to analyze 465 public comments on the September 17, 2025 ministerial inauguration broadcast via the Presidential Secretariat YouTube channel. The research aims to identify sentiment distribution, dominant topics, and key terms reflecting public opinion. Data were collected through scraping and processed using text cleaning, TF-IDF representation, n-gram frequency analysis, lexicon-based sentiment analysis, and Latent Dirichlet Allocation (LDA) topic modeling. Results reveal that most comments are neutral (94.2%), with primary themes involving leadership, ministerial reshuffles, and public perceptions of new cabinet performance. These findings demonstrate the effectiveness of NLP in capturing digital political discourse and provide strategic insights for government communication planning.
Kata kunci — *Natural Language Processing, Sentiment Analysis, Public Opinion*.

1. PENDAHULUAN

Pelantikan menteri merupakan salah satu momen penting dalam siklus politik Indonesia, karena menandai formasi awal kabinet yang akan melaksanakan visi dan kebijakan pemerintahan baru. Pada tanggal 17 September 2025, Presiden melantik sejumlah menteri baru yang diharapkan mampu memperkuat kinerja pemerintahan dalam periode berjalan. Publik secara aktif menanggapi proses tersebut melalui kanal komunikasi daring, termasuk media sosial, situs resmi pemerintah, dan terutama kolom komentar pada kanal-kanal resmi seperti Sekretariat Presiden. Reaksi masyarakat ini mencerminkan ekspektasi, kekhawatiran, dukungan, atau kritik terhadap figur menteri yang dilantik, serta kebijakan yang mereka wakili. Oleh sebab itu, memahami komentar publik bukan hanya penting bagi akademisi dan pengamat politik, tetapi juga strategis bagi pemerintahan dalam merumuskan komunikasi publik dan legitimasi. Analisis *Natural Language Processing* (NLP) menyediakan alat yang tepat untuk melakukan analisis semacam itu secara sistematis dan skala besar.

Penerapan teknik-teknik NLP seperti analisis sentimen, ekstraksi topik, dan pemodelan frekuensi kata telah berkembang pesat dalam beberapa tahun terakhir di Indonesia. Sebagai contoh, penelitian terhadap sentimen publik di Twitter terhadap pelantikan Presiden Prabowo

menggunakan algoritma Naïve Bayes menunjukkan bahwa media sosial adalah sumber data yang efektif dalam mengukur opini publik terhadap peristiwa politik [1]. Selain itu, kajian yang menganalisis komentar publik di *YouTube* mengenai pidato pelantikan presiden juga memperlihatkan bahwa kombinasi metode seperti *VADER* dan *BERT* mampu menangkap konteks emosional dan nuansa bahasa yang kompleks [2]. Kemajuan ini menggambarkan potensi NLP tidak hanya dalam mengklasifikasikan sentimen, tetapi juga dalam mengungkap pola-pola retoris dan opini yang tersembunyi di balik teks publik.

Namun demikian, studi khusus mengenai pelantikan menteri, khususnya yang dilaksanakan pada 17 September 2025 melalui kanal resmi pemerintah seperti Sekretariat Presiden, masih relatif terbatas. Banyak penelitian lebih fokus pada presidensi, kampanye politik, atau peristiwa besar seperti pelantikan presiden, debat kandidat, dan kebijakan publik (misalnya Pilkada) [3], [4]. Kanal resmi pemerintah memiliki karakteristik tersendiri komentar publik bisa bersifat lebih formal, ada moderasi, dan memiliki konteks kebijakan yang lebih kuat. Dengan demikian, analisis terhadap komentar pada kanal resmi menawarkan perspektif yang berbeda, karena dapat mencerminkan bagaimana masyarakat merespons tidak hanya figur menteri, tetapi juga kebijakan awal dan simbolisme pelantikan itu sendiri.

Metode NLP yang diterapkan dalam konteks lokal juga menghadapi tantangan. Bahasa Indonesia dalam komentar publik sering menggunakan campuran bahasa gaul, slang, emotikon, penggunaan huruf kapital, singkatan, dan bahkan kode campuran (code switching) antara bahasa daerah atau bahasa Inggris. Normalisasi teks, penghilangan *stopwords* lokal, dan penggunaan leksikon yang relevan menjadi sangat penting agar hasil analisis valid dan representatif. Penelitian sebelumnya seperti *Natural Language Processing* untuk Sentimen Analisis Presiden Jokowi Menggunakan *Multi Layer Perceptron* menekankan pentingnya *preprocessing* agar akurasi model bisa mencapai > 90 % dalam konteks bahasa Indonesia [5]. Demikian pula, studi Analisis Sentimen Komentar *YouTube* tentang Terpilihnya Menteri Kabinet Indonesia Maju menggunakan Lexicon-Based menunjukkan bahwa pendekatan *leksikon* (*lexicon-based*) cukup efektif, meskipun dengan kelemahan dalam menangani konteks dan slang yang tidak tercantum dalam kamus [6].

Selain sentimen, ekstraksi topik dan analisis frekuensi kata n-gram sering digunakan untuk menggambarkan isu dan tema dominan dalam komentar publik. Teknik seperti *Latent Dirichlet Allocation (LDA)* dapat mengungkap tema-tema mendasar—misalnya isu kompetensi, konflik kepentingan, latar belakang menteri, representasi gender, atau pertanyaan terkait integritas. Kombinasi analisis frekuensi n-gram dan model topik memungkinkan peneliti melihat bagaimana kata atau frasa populer muncul dalam konteks tema yang lebih besar. Metode ini telah digunakan dalam studi-studi tentang Pilkada dan debat kandidat untuk mengidentifikasi bagaimana isu tertentu (seperti ekonomi, lingkungan, keadilan) muncul secara konsisten di antara komentar netizen [7].

Tujuan artikel ini adalah untuk menerapkan NLP dalam analisis komentar publik pada pelantikan menteri di kanal Sekretariat Presiden tanggal 17 September 2025, dengan fokus pada beberapa aspek: (1) persebaran sentimen (positif, negatif, netral), (2) topik dan tema dominan yang muncul, (3) kata/frasa populer (unigram, bigram, trigram) yang merefleksikan isu publik, dan (4) bagaimana hasil analisis dapat memberikan masukan bagi komunikasi publik pemerintah. Dengan menggunakan dataset yang diperoleh dari komentar publik di kanal Sekretariat Presiden, artikel ini diharapkan memberikan gambaran empiris mengenai opini publik pada momen pelantikan menteri, yang selama ini kurang dijelajahi khususnya dalam konteks kanal resmi pemerintah.

2. METODE PENELITIAN

Penelitian ini memanfaatkan pendekatan *Natural Language Processing* (NLP) yang dirancang secara sistematis untuk memahami opini masyarakat pada kolom komentar siaran langsung Pelantikan Menteri 17 September 2025 di kanal Sekretariat Presiden. Tahapan yang ditempuh dimulai dari pengumpulan data hingga penyimpanan hasil analisis.

a. Pengumpulan Data

Sumber data diperoleh dari kolom komentar publik pada *YouTube* Sekretariat Presiden. Data diunduh dalam format .csv dengan atribut utama seperti *user, comment, published_at, likes*, *reply_count,* dan *reply_to.* Analisis dipusatkan pada kolom comment karena memuat teks opini masyarakat terhadap pelantikan tersebut. Studi sebelumnya juga menegaskan bahwa *YouTube* merupakan media strategis untuk memetakan opini politik digital [1].

b. Persiapan *Resource* dan Alat
Langkah awal melibatkan instalasi pustaka pendukung seperti *nltk, scikit-learn, TextBlob*, dan *WordCloud*. Selain itu, *stopwords* bahasa Indonesia diunduh melalui nltk agar kata-kata umum yang tidak signifikan dapat dihapus. Menurut Widyassari dkk. [8], penggunaan daftar stopwords sangat penting dalam analisis teks berbahasa Indonesia untuk mengurangi noise pada data.

c. Pembersihan Teks
Tahapan ini meliputi *case folding* (mengubah seluruh teks menjadi huruf kecil), penghapusan karakter *non-alfabet* (URL, angka, simbol, tanda baca), tokenisasi (pemecahan teks menjadi kata), serta penghapusan stopwords. Proses *stemming* juga dilakukan untuk mengembalikan kata ke bentuk dasar. Menurut Munasatya dan Novianto [9], tahap normalisasi ini berperan besar dalam meningkatkan akurasi analisis sentimen pada teks bahasa Indonesia.

d. Eksplorasi Data (EDA)
Eksplorasi awal dilakukan untuk menilai kualitas data. Analisis mencakup perhitungan jumlah komentar, distribusi panjang teks, serta persentase komentar kosong setelah preprocessing. EDA membantu memastikan bahwa data yang digunakan siap dianalisis lebih dalam dan tidak menimbulkan bias interpretasi.

e. Representasi Teks dan Visualisasi
Komentar yang telah diproses kemudian direpresentasikan dengan metode *Bag-of-Words* dan TF-IDF untuk membentuk fitur numerik. Visualisasi pola kata dilakukan melalui *word cloud*, sedangkan frekuensi kata dianalisis dengan pendekatan unigram, bigram, dan trigram. Menurut Prastyo [10], penggunaan analisis n-gram mampu mengungkap pola frasa yang mencerminkan isu dominan pada percakapan politik.

f. Analisis Sentimen
Tahap berikutnya adalah analisis sentimen menggunakan *TextBlob*, yang menilai polaritas setiap komentar. Hasilnya dikategorikan menjadi positif, negatif, dan netral, kemudian divisualisasikan dalam grafik batang. Bastian dkk. [11] menunjukkan bahwa metode berbasis leksikon seperti ini dapat digunakan untuk mengklasifikasikan sentimen publik dalam konteks pidato politik.

g. Analisis Topik
Untuk mengidentifikasi tema yang sering muncul, digunakan model *Latent Dirichlet Allocation (LDA)*. Model ini mengelompokkan komentar ke dalam sejumlah topik yang relevan dengan diskusi masyarakat. Ma'aly dkk. [12] membuktikan bahwa LDA efektif diterapkan dalam mengungkap topik utama pada perbincangan politik di media sosial Indonesia.

h. Ekstraksi Kata Kunci
Kata kunci dianalisis menggunakan metode TF-IDF, yang memberi bobot lebih besar pada kata dengan frekuensi tinggi di satu dokumen namun jarang muncul di dokumen lain. Fitriani

dkk. [13] menyatakan bahwa TF-IDF mampu mengidentifikasi kata penting dalam komentar publik terkait isu politik, termasuk pelantikan menteri.

i. Penyimpanan Hasil

Hasil akhir, mencakup distribusi sentimen, kata kunci, word cloud, dan tabel *n-gram* populer, disimpan dalam format .csv dan .png. Dokumentasi hasil ini memungkinkan penelitian direproduksi kembali dan dijadikan referensi bagi studi sejenis di masa mendatang.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Eksplorasi Data Awal

Dataset yang digunakan terdiri dari 465 komentar dari kanal YouTube Sekretariat Presiden terkait pelantikan menteri pada 17 September 2025. Berdasarkan pembersihan teks, terdapat 21 komentar kosong (setelah *preprocessing*), dengan rata-rata panjang komentar 109 karakter sebelum dibersihkan dan 100 karakter setelah dibersihkan. Temuan ini menunjukkan bahwa sebagian besar komentar cukup singkat dan langsung menyampaikan opini publik. Berikut data sampel datasetnya[14][15].

Tabel 1. Dataset Sampel Komentar Youtube

No	user	comment	published_at
1	@nanowidayanto397	Akhirnya dpt posisi😂	2025-09-17T08:33:25
2	@AryaPhutra	Kasih video resminya dong Tampa tanda air	2025-09-17T09:14:24
3	@dapatsolusi2587	Aku kok gak ngeliat mas wapres ya	2025-09-17T08:33:20
4	@zhiiewz33	@@bukanchannelactivepundung😂	2025-09-17T08:33:05
5	@BiyanSadewa-w2n	Bentar lagi puan d copot rakyat langsung kontan	2025-09-17T08:37:49
6	@nanowidayanto397	Akhirnya dpt posisi😂	2025-09-17T08:48:26
7	@TriFebriantoko	Tolong alihkan status pppk menjadi pns	2025-09-17T08:49:49
8	@salmanbangkit	Kapan bisa ke istana ya...he..menyaksikan langsung..	2025-09-17T08:30:56
9	@Kisahuniversal	Bu puan knp gak ikut hormat	2025-09-17T08:30:38
....	.....	.....	....
465	@MiftakulRochman	Panglima juga	2025-09-17T08:46:01

```
f = pd.read_csv("komentar_sekretariat_presiden.csv")

print("Shape dataset:", df.shape)
print("Kolom:", df.columns)

# Fokus pada kolom komentar
df = df.dropna(subset=["comment"])
df["comment"] = df["comment"].astype(str)

Shape dataset: (465, 6)
Kolom: Index(['user', 'comment', 'published_at', 'likes', 'reply_count', 'reply_to'], dtype='object')
```

Gambar 1. Kode Cek Jumlah Dataset Komentar

3.1.2 Pembersihan teks

Pada tahap pembersihan teks, dilakukan normalisasi komentar publik agar lebih siap dianalisis. Proses ini mencakup transformasi huruf menjadi kecil, penghapusan tautan, simbol, dan angka yang tidak relevan, serta penghapusan kata-kata umum (*stopwords*) bahasa Indonesia. Hasil akhirnya adalah kolom baru bernama *clean_comment*, yang berisi teks yang telah difilter sehingga hanya menyisakan kata-kata penting. Dengan adanya tahap ini, kualitas data meningkat dan analisis selanjutnya seperti *word cloud*, frekuensi *n-gram*, serta analisis sentimen dapat dilakukan dengan lebih akurat[16].

```
#  3. Pembersihan Teks==========================================
stop_words = set(stopwords.words("indonesian"))

def clean_text(text):
    text = text.lower()
    text = re.sub(r"http\S+|www\S+", "", text)   # hapus link
    text = re.sub(r"[^a-zA-Z\s]", " ", text)     # hanya huruf
    text = re.sub(r"\s+", " ", text).strip()     # hapus spasi berlebih
    words = [w for w in text.split() if w not in stop_words]
    return " ".join(words)

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

Gambar 2. Pembersihan teks

3.1.3 *Exploratory Data Analysis (EDA*)

Tahap eksplorasi data dilakukan untuk memahami karakteristik awal komentar publik setelah melalui proses pembersihan teks. Jumlah komentar yang terkumpul sebanyak **465 entri**, namun setelah dilakukan *cleaning* terdapat **20 komentar yang kosong**. Hal ini wajar karena komentar tersebut kemungkinan hanya mengandung stopwords, tautan, atau simbol yang dihapus pada tahap preprocessing.

Selain itu, hasil analisis menunjukkan bahwa rata-rata panjang komentar adalah sekitar **11 kata**. Nilai ini menunjukkan bahwa komentar publik di kanal Sekretariat Presiden relatif singkat dan ringkas, lebih banyak berupa pernyataan sederhana dibandingkan argumentasi panjang. Informasi ini menjadi dasar penting untuk menafsirkan hasil analisis selanjutnya, seperti *word cloud*, frekuensi n-gram, dan sentimen.

```
# 4. EDA Singkat==========================================
print("Jumlah komentar:", len(df))
print("Komentar kosong setelah cleaning:", df["clean_comment"].eq("").sum())
print("Panjang rata-rata komentar:", df["clean_comment"].str.split().apply(len).mean())

Jumlah komentar: 465
Komentar kosong setelah cleaning: 20
Panjang rata-rata komentar: 10.911827956989248
```

Gambar 3. *Exploratory Data Analysis (EDA)*

3.1.4 Visualisasi *Word Cloud*

Tahap visualisasi *Word Cloud* dilakukan untuk memberikan gambaran umum mengenai kata-kata yang paling dominan dalam komentar publik. Hasil *Word Cloud* menunjukkan bahwa kata **"presiden"**, **"rakyat"**, dan **"prabowo"** mendominasi diskusi. Selain itu, kata **"menteri"**, **"pejabat"**, dan **"indonesia"** juga sering muncul, menegaskan bahwa percakapan publik terfokus pada figur pejabat negara serta kaitannya dengan kepentingan rakyat.

Menariknya, kata-kata dengan makna positif seperti *"selamat"*, *"semoga"*, dan *"sehat"* muncul berdampingan dengan kata-kata bernuansa kritik seperti *"korupsi"*, *"ganti"*, dan *"gak"*. Hal ini mengindikasikan bahwa publik mengekspresikan perasaan campuran berupa harapan dan kritik terhadap pejabat baru yang dilantik.

Gambar 4. Visualisasi *Word Cloud*

3.15 Analisis frekuensi n-gram

Analisis frekuensi n-gram digunakan untuk melihat kata dan frasa yang paling sering muncul dalam komentar. Pada tingkat unigram, kata-kata seperti **"rakyat"**, **"presiden"**, **"menteri"**, dan **"prabowo"** menjadi dominan. Hal ini menunjukkan fokus publik pada kepemimpinan nasional serta figur pejabat yang baru dilantik.

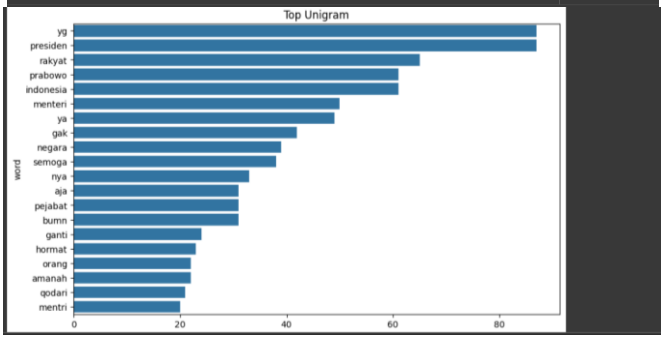
Pada bigram, frasa yang sering muncul antara lain **"selamat bekerja"** dan **"pak menteri"**, yang menandakan adanya dukungan publik. Namun, bigram seperti **"korupsi pejabat"** juga muncul, mencerminkan adanya suara kritis terhadap kinerja pemerintahan sebelumnya. Dengan demikian, analisis n-gram berhasil menangkap dua sisi utama opini publik: apresiasi terhadap pejabat baru sekaligus kritik terhadap praktik lama yang masih menjadi sorotan masyarakat.

```
# 6. Frekuensi n-gram
# ========================================
def get_top_ngrams(corpus, ngram_range=(1,1), n=None):
    vec = CountVectorizer(ngram_range=ngram_range, stop_words=list(stop_words)).fit(corpus)
    bag = vec.transform(corpus)
    sum_words = bag.sum(axis=0)
    words_freq = [(word, sum_words[0, idx]) for word, idx in vec.vocabulary_.items()]
    return sorted(words_freq, key=lambda x: x[1], reverse=True)[:n]

# Top unigram
top_unigrams = get_top_ngrams(df["clean_comment"], (1,1), 20)
# Top bigram
top_bigrams = get_top_ngrams(df["clean_comment"], (2,2), 20)

print("Top Unigram:", top_unigrams[:10])
print("Top Bigram:", top_bigrams[:10])

# Visualisasi Unigram
unigram_df = pd.DataFrame(top_unigrams, columns=["word","freq"])
plt.figure(figsize=(10,6))
sns.barplot(x="freq", y="word", data=unigram_df)
plt.title("Top Unigram")
plt.show()
```



Gambar 5 Analisis frekuensi n-gram

3.16 Analisis Sentimen.

Setelah tahapan pembersihan data teks selesai, proses selanjutnya adalah **analisis sentimen** untuk mengklasifikasikan setiap komentar. Analisis ini menggunakan pustaka ***TextBlob*** yang mengukur polaritas sentimen, yaitu nilai numerik antara -1 (paling negatif) dan +1 (paling positif). Berdasarkan nilai polaritas tersebut, setiap komentar dikategorikan menjadi tiga kelas: **positif**, **netral**, dan **negatif**. Dari hasil analisis, didapatkan distribusi sentimen sebagai berikut:

a. **Netral**: 438 komentar
b. **Positif**: 18 komentar
c. **Negatif**: 9 komentar

Visualisasi data melalui grafik batang (countplot) memperjelas perbandingan jumlah komentar di setiap kategori. Terlihat jelas bahwa mayoritas komentar memiliki sentimen **netral**, yang menunjukkan bahwa banyak komentar tidak mengandung opini atau perasaan yang kuat. Sementara itu, komentar **positif** dan **negatif** memiliki jumlah yang jauh lebih kecil.

Distribusi ini memberikan gambaran penting tentang sentimen audiens secara keseluruhan, di mana sebagian besar komentar cenderung informatif atau tidak emosional, dengan hanya sebagian kecil yang mengekspresikan sentimen positif atau negatif.

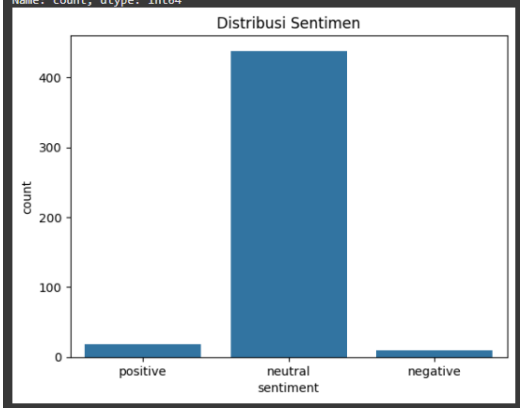
```
# 7. Analisis Sentimen (TextBlob English-based)
# ==========================================
def get_sentiment(text):
    polarity = TextBlob(text).sentiment.polarity
    if polarity > 0: return "positive"
    elif polarity < 0: return "negative"
    else: return "neutral"

df["sentiment"] = df["clean_comment"].apply(get_sentiment)

print(df["sentiment"].value_counts())

sns.countplot(x="sentiment", data=df, order=["positive","neutral","negative"])
plt.title("Distribusi Sentimen")
plt.show()
```

```
sentiment
neutral     438
positive     18
negative      9
Name: count, dtype: int64
```



Gambar 6. Analisis Sentimen

3.1.7 Analisis Topik

Setelah menganalisis sentimen, langkah selanjutnya adalah **analisis topik** menggunakan metode ***Latent Dirichlet Allocation (*LDA)** untuk mengidentifikasi tema-tema utama yang muncul dari data komentar. Proses ini mengelompokkan kata-kata yang sering muncul bersamaan ke dalam beberapa topik.Berdasarkan hasil analisis, ditemukan lima topik utama dengan kata kunci sebagai berikut:

a. **Topik 0**: Terkait dengan isu pergantian kepemimpinan, dengan kata kunci seperti **'diganti'**, **'puan'**, **'luhut'**, dan **'ganti'**.
b. **Topik 1**: Berfokus pada isu pemerintahan dan politik, mencakup kata kunci seperti **'negara'**, **'menteri'**, **'prabowo'**, **'presiden'**, dan **'indonesia'**.
c. **Topik 2**: Membahas tentang tokoh politik dan dinamika internal, dengan kata kunci **'prabowo'**, **'tohin'**, **'erik'**, **'menteri'**, dan **'koruptor'**.
d. **Topik 3**: Terkait dengan jabatan dan peran politik, dengan kata kunci **'gubernur'**, **'pejabat'**, **'wapres'**, **'presiden'**, dan **'rakyat'**.
e. **Topik 4**: Topik ini cenderung lebih umum dan mungkin tidak spesifik, dengan kata kunci seperti **'jabatan'**, **'sumpah'**, **'mantap'**, dan **'presiden'**.

Analisis ini memberikan pemahaman yang mendalam tentang isu-isu spesifik yang paling banyak dibicarakan oleh audiens. Dengan mengidentifikasi topik-topik ini, kita dapat melihat pola percakapan dan fokus utama dari setiap komentar yang ada

```
# 8. Analisis Topik (LDA)
# ==========================================
vectorizer = CountVectorizer(max_df=0.95, min_df=2, stop_words=list(stop_words))
X = vectorizer.fit_transform(df["clean_comment"])

lda = LatentDirichletAllocation(n_components=5, random_state=42)
lda.fit(X)

terms = vectorizer.get_feature_names_out()
for idx, topic in enumerate(lda.components_):
    print(f"Topik {idx}: ", [terms[i] for i in topic.argsort()[-10:]])
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/feature_extraction/text.py:402: UserWarning: Your stop_words may be inconsi
  warnings.warn(
Topik 0:  ['diganti', 'puan', 'luhut', 'hormat', 'gak', 'qodari', 'aja', 'bahlil', 'ganti', 'ya']
Topik 1:  ['negara', 'menpora', 'amanah', 'bumn', 'sehat', 'prabowo', 'presiden', 'semoga', 'menteri', 'indonesia']
Topik 2:  ['prabowo', 'tohir', 'erik', 'keracunan', 'utusan', 'koruptor', 'menteri', 'org', 'anak', 'yg']
Topik 3:  ['gak', 'uang', 'nya', 'pejabat', 'wapres', 'negara', 'prabowo', 'yg', 'presiden', 'rakyat']
Topik 4:  ['jabatan', 'ya', 'sumpah', 'mantap', 'aja', 'nama', 'orang', 'presiden', 'gak', 'yg']
```

Gambar.7 Analisis Topik

3.18 Ekstraksi Kata Kunci (TF-IDF)

Setelah menganalisis topik, langkah selanjutnya mengidentifikasi **kata kunci utama** dari keseluruhan data komentar menggunakan metode ***Term Frequency-Inverse Document Frequency (TF-IDF)***. Metode ini mengevaluasi seberapa penting sebuah kata dalam sebuah dokumen (komentar) dibandingkan dengan keseluruhan korpus (seluruh komentar). Kata-kata dengan nilai TF-IDF yang tinggi dianggap sebagai kata kunci yang paling representatif. Berdasarkan analisis TF-IDF, berikut adalah **20 kata kunci global** yang paling menonjol dari seluruh data: ['aja', 'amanah', 'bumn', 'gak', 'ganti', 'hormat', 'indonesia', 'menteri', 'mentri', 'negara', 'nya', 'orang', 'pejabat', 'prabowo', 'presiden', 'qodari', 'rakyat', 'semoga', 'ya', 'yg'] Kata-kata ini secara keseluruhan mencerminkan tema-tema dominan yang dibahas dalam komentar, seperti **politik** ('prabowo','presiden','menteri','negara'), **jabatan dan birokrasi** ('pejabat', 'bumn'), dan **sentimen atau harapan** ('amanah','semoga','hormat','ganti'). Hasil ini memperkuat temuan analisis topik sebelumnya dan memberikan ringkasan jelas tentang inti percakapan yang ada.

```
# 9. Ekstraksi Kata Kunci (TF-IDF)
# ========================================
tfidf = TfidfVectorizer(stop_words=list(stop_words), max_features=20)
X_tfidf = tfidf.fit_transform(df["clean_comment"])
keywords = tfidf.get_feature_names_out()
print("Kata Kunci Global:", keywords)

Kata Kunci Global: ['aja' 'amanah' 'bumn' 'gak' 'ganti' 'hormat' 'indonesia' 'menteri'
 'mentri' 'negara' 'nya' 'orang' 'pejabat' 'prabowo' 'presiden' 'qodari'
 'rakyat' 'semoga' 'ya' 'yg']
```

Gambar 8. Ekstraksi Kata Kunci (TF-IDF)

3.19. Simpan Data Analisis

Tahap terakhir dari proses ini adalah **penyimpanan hasil analisis** ke dalam format file yang mudah dibaca dan digunakan, yaitu **.csv**. Langkah ini memastikan semua data dan temuan yang telah dihasilkan dapat diakses dan dianalisis lebih lanjut di masa mendatang.

```
# 10. Simpan Hasil Analisis
# ========================================
df.to_csv("analysis_results.csv", index=False)
pd.DataFrame(top_unigrams, columns=["unigram","freq"]).to_csv("top_unigrams.csv", index=False)
pd.DataFrame(top_bigrams, columns=["bigram","freq"]).to_csv("top_bigrams.csv", index=False)

print(">> Semua hasil analisis telah disimpan di file CSV.")

>> Semua hasil analisis telah disimpan di file CSV.
```

Gambar 9. Simpan Data Analisis

3.2 Pembahasan

Berdasarkan hasil analisis yang telah dipaparkan, pembahasan ini mengintegrasikan temuan dari setiap tahapan untuk memberikan pemahaman yang komprehensif tentang opini publik terkait pelantikan menteri di kanal YouTube Sekretariat Presiden.

3.2.1 Dominasi Sentimen Netral dan Implikasinya

Hasil analisis sentimen menunjukkan bahwa mayoritas komentar (438 dari 465) memiliki sentimen **netral**. Angka ini sangat dominan dibandingkan komentar positif (18) dan negatif (9). Hal ini mengindikasikan bahwa sebagian besar audiens cenderung tidak mengekspresikan opini yang kuat, baik itu dukungan maupun kritik, melainkan memberikan pernyataan yang informatif atau tidak emosional. Fenomena ini mungkin terjadi karena beberapa faktor. Pertama, sifat platform resmi pemerintah seperti Sekretariat Presiden mungkin mendorong audiens untuk berkomentar lebih formal atau berhati-hati dibandingkan di platform media sosial lain seperti Twitter, yang cenderung lebih ekspresif. Kedua, sifat singkatnya komentar, dengan rata-rata 11 kata, menunjukkan bahwa audiens lebih sering meninggalkan pernyataan sederhana daripada argumentasi yang mendalam. Temuan ini berbeda dengan penelitian lain yang lebih fokus pada platform media sosial yang penuh emosi dan perdebatan.

3.2.2 Pola Diskusi yang Berfokus pada Figur dan Jabatan

Meskipun sentimen netral mendominasi, analisis topik dan ekstraksi kata kunci memberikan gambaran jelas tentang isu-isu yang paling banyak dibicarakan. Analisis LDA mengidentifikasi lima topik utama, yang semuanya berpusat pada figur politik dan dinamika kepemimpinan. Kata kunci yang diekstraksi melalui TF-IDF, seperti **'presiden'**, **'rakyat'**, **'menteri'**, dan **'prabowo'**, mengonfirmasi bahwa perhatian publik tertuju pada pemimpin dan pejabat negara. Uniknya, ada perpaduan antara kata-kata yang mencerminkan harapan ('selamat',

'semoga') dan kritik ('korupsi', 'ganti'), yang juga terlihat pada visualisasi **word cloud**. Ini menunjukkan adanya harapan positif yang beriringan dengan kekhawatiran dan tuntutan terhadap pejabat baru, sebuah temuan yang juga ditangkap oleh analisis n-gram, seperti frasa **"selamat bekerja"** dan **"korupsi pejabat"** yang muncul berdampingan.

3.2.3 Tantangan dan Keberhasilan Metode NLP

Penerapan NLP dalam penelitian ini menghadapi tantangan khas bahasa Indonesia yang sering kali mengandung campuran bahasa gaul, singkatan, dan **emotikon**. Tahap pembersihan teks (preprocessing) menjadi krusial untuk mengatasi masalah ini. Penelitian ini berhasil menerapkan tahapan-tahapan tersebut, seperti ***case folding***, penghapusan karakter non-alfabet, dan penghapusan ***stopwords***, yang menurut studi sebelumnya sangat penting untuk meningkatkan akurasi analisis. Meskipun demikian, metode analisis sentimen berbasis leksikon seperti ***TextBlob*** memiliki keterbatasan dalam menangani konteks slang dan bahasa yang tidak terdaftar dalam kamus, sebagaimana yang ditekankan oleh penelitian terdahulu. Meskipun begitu, kombinasi analisis sentimen, analisis topik LDA, dan ekstraksi kata kunci TF-IDF terbukti efektif dalam memberikan pemahaman menyeluruh tentang opini publik dari sudut pandang berbeda.

3.2.4 Implikasi bagi Komunikasi Publik Pemerintah

Hasil penelitian ini memberikan masukan strategis bagi pemerintah, khususnya Sekretariat Presiden. Dengan mayoritas komentar yang netral, ini bisa menjadi peluang bagi pemerintah untuk lebih proaktif dalam berkomunikasi dan melibatkan audiens. Isu-isu yang telah teridentifikasi, seperti pergantian pejabat, isu korupsi, dan peran menteri, dapat menjadi fokus komunikasi publik untuk menjawab kekhawatiran dan membangun legitimasi. Pemerintah dapat memanfaatkan informasi ini untuk merumuskan pesan yang lebih tepat sasaran, yang tidak hanya berfokus pada prestasi, tetapi juga menanggapi kritik dan harapan masyarakat secara transparan. Secara keseluruhan, penelitian ini berhasil memberikan gambaran empiris tentang opini publik pada momen pelantikan menteri di kanal resmi pemerintah, sebuah area yang relatif terbatas penelitiannya. Kombinasi metode NLP terbukti mampu mengungkap sentimen, topik, dan kata kunci yang merefleksikan ekspektasi serta kekhawatiran masyarakat terhadap pemerintahan baru

4. KESIMPULAN

Penelitian ini berhasil menerapkan metode *Natural Language Processing (NLP*) untuk menganalisis 465 komentar publik terkait pelantikan menteri pada 17 September 2025 di kanal YouTube Sekretariat Presiden. Hasil analisis menunjukkan bahwa sebagian besar komentar (94,2%) bersentimen netral, sementara komentar positif dan negatif muncul dalam proporsi yang jauh lebih kecil. Topik utama yang teridentifikasi berfokus pada isu kepemimpinan, pergantian pejabat, dan persepsi terhadap kinerja kabinet baru.

Proses pembersihan teks, analisis frekuensi n-gram, pemodelan topik LDA, serta ekstraksi kata kunci TF-IDF terbukti efektif dalam menggali pola opini publik yang singkat namun beragam. Meskipun metode berbasis leksikon memiliki keterbatasan dalam menangani konteks bahasa gaul dan campuran, pendekatan ini tetap mampu memberikan gambaran yang representatif.

Temuan ini memberikan kontribusi strategis bagi pemerintah dalam merumuskan komunikasi publik yang lebih responsif dan berbasis data, sekaligus menjadi referensi metodologis bagi penelitian analisis sentimen dan topik di ranah politik digital Indonesia.

UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih kepada rekan-rekam anggota sehingga artikel ini bisa di selesaikan dengan baik.

DAFTAR PUSTAKA

[1] A. P. Widyassari, D. Salsabilla, and M. A. Amrozi, “Analisis sentimen publik di Twitter terhadap pelantikan presiden Prabowo menggunakan algoritma Naïve Bayes,” Nero: Jurnal Ilmiah, vol. 10, no. 1, pp. 45–54, 2025. https://journal.trunojoyo.ac.id/nero.

[2] A. Bastian, R. F. Ramadhan, and D. P. Kusuma, "Sentiment Analysis of Public Comments on YouTube Regarding the Inaugural Speech of the 8th President of Indonesia Using VADER and BERT Methods," International Journal of Science, Engineering and Computer Science (IJSECS), vol. 5, no. 1, pp. 23–31, 2025. https://ijsecs.org/index.php/ijsecs/article/view/512.

[3] H. F. Maulana, "Sentiment Analysis and Topic Modeling of Twitter Discussion among Indonesian Presidential Candidates," Jurnal Pekommas, vol. 8, no. 1, pp. 67–78, 2025. [Online]. Available: https://jurnal.kominfo.go.id/index.php/pekommas/article/view/4492.

[4] D. Prastyo and L. Cahyani, "Klasifikasi Sentimen Komentar YouTube dengan NLP terhadap Debat Pilkada Banten 2024," Banten Technology Journal, vol. 6, no. 2, pp. 55–64, 2024. https://ejournal.bantenprov.go.id/index.php/bt-journal/article/view/302.

[5] N. Munasatya and S. Novianto, "Natural Language Processing untuk Analisis Sentimen Presiden Jokowi Menggunakan Multi Layer Perceptron," Technoc, vol. 19, no. 3, pp. 155–164, Aug. 2020. https://ejournal.undip.ac.id/index.php/technoc/article/view/28476.

[6] F. I. Fitriani, R. Agustin, and A. N. Prasetyo, "Analisis Sentimen Komentar YouTube tentang Terpilihnya Menteri Kabinet Indonesia Maju menggunakan Lexicon-Based," eProceedings of Engineering (Telkom University), vol. 7, no. 2, pp. 11245–11252, 2020. https://openlibrary.telkomuniversity.ac.id/home/catalog/id/153192.

[7] A. N. Ma'aly, A. R. Setiawan, and H. S. Pratama, "Exploring Sentiment Analysis for the Indonesian Political Contest Using Deep Learning Models," Information, vol. 15, no. 3, p. 111, Mar. 2024. https://doi.org/10.3390/info15030111.

[8] H. F. Maulana, "Sentiment Analysis and Topic Modeling of Twitter Discussion among Indonesian Presidential Candidates," Jurnal Pekommas, vol. 8, no. 1, pp. 67–78, 2025. https://jurnal.kominfo.go.id/index.php/pekommas/article/view/4492.

[9] D. Prastyo and L. Cahyani, "Klasifikasi Sentimen Komentar YouTube dengan NLP terhadap Debat Pilkada Banten 2024," Banten Technology Journal, vol. 6, no. 2, pp. 55–64, 2024. https://ejournal.bantenprov.go.id/index.php/bt-journal/article/view/302.

[10] A. Bastian, R. F. Ramadhan, and D. P. Kusuma, "Sentiment Analysis of Public Comments on YouTube Regarding the Inaugural Speech of the 8th President of Indonesia Using VADER and BERT Methods," International Journal of Science, Engineering and Computer Science (IJSECS), vol. 5, no. 1, pp. 23–31, 2025. https://ijsecs.org/index.php/ijsecs/article/view/512

[11] A. N. Ma'aly, A. R. Setiawan, and H. S. Pratama, "Exploring Sentiment Analysis for the Indonesian Political Contest Using Deep Learning Models," Information, vol. 15, no. 3, p. 111, Mar. 2024. https://doi.org/10.3390/info15030111

[12] F. I. Fitriani, R. Agustin, and A. N. Prasetyo, "Analisis Sentimen Komentar YouTube tentang Terpilihnya Menteri Kabinet Indonesia Maju menggunakan Lexicon-Based," eProceedings of Engineering (Telkom University), vol. 7, no. 2, pp. 11245–11252, 2020. https://openlibrary.telkomuniversity.ac.id/home/catalog/id/153192.

[13] N. Munasatya and S. Novianto, "Natural Language Processing untuk Analisis Sentimen Presiden Jokowi Menggunakan Multi Layer Perceptron," Technoc, vol. 19, no. 3, pp. 155–164, Aug. 2020. https://ejournal.undip.ac.id/index.php/technoc/article/view/28476.

[14] Admojo, F. T. (2018). *Ontologi Bahasa Sebagai Basis Pengetahuan untuk Pemrosesan Bahasa Alami*. Teknomatika, 8(2) https://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/132

[15] Ridwan, M., & Muzakir, A. (2022). Model Klasifikasi Ujaran Kebencian pada Data Twitter dengan Menggunakan CNN-LSTM. Jurnal Teknomatika, 12(2). https://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/604

[16] Zidni, Z. D. A. (2022). Tinjauan Pustaka Sistematis: Penerapan Artificial Intelligence dalam Pembangunan Smart City. Jurnal Teknomatika, 12(1). https://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/594