
Pemodelan Topik dan Analisis Sentimen pada 'Voices of History: 50 Iconic Speeches' Menggunakan Pendekatan Natural Language Processing

Topic Modeling and Sentiment Analysis in 'Voices of History: 50 Iconic Speeches' Using a Natural Language Processing Approach

Juniana Husna^{1*}, Margareth Dyah Anggraini Widirahayu²

^{1,2} Sekolah Tinggi Multi Media "MMTC" Yogyakarta

^{1,2} Alamat: Jl. Magelang Km. 6 Yogyakarta 55284

*Email korespondensi: juniana.husna@mmtc.ac.id

Diterima: 25 Juni 2024 || Revisi: 27 Juni 2024 || Disetujui: 30 Juni 2024

Abstract

This study examines various iconic speeches throughout history using Natural Language Processing (NLP) approaches. The selected speeches address a wide range of historical and social issues, ranging from the struggle for civil rights upto the calls for peace.. This analysis is carried out to reveal the sentiments, themes and emotions expressed in each speech, as well as to understand how orators use language to influence and inspire their audience. This study identifies common patterns and themes in these speeches by utilizing text vectorization methods, topic extraction using Latent Dirichlet Allocation (LDA), as well as sentiment and emotion analysis. The findings of this analysis show that the most influential speeches mostly include strong themes such as social justice, freedom, and unity. Furthermore, good thoughts and powerful emotions such as optimism and courage are frequently featured in the most motivating speeches. This research provides valuable insight for historians, researchers, and oratory enthusiasts into how iconic speeches have shaped public opinion and influenced social change.

Keywords: *Natural Language Processing, Sentiment Analysis, Topic Extraction, Text Clustering.*

Abstrak

Penelitian ini membahas analisis mendalam terhadap beberapa pidato ikonik dalam sejarah menggunakan teknik Natural Language Processing (NLP). Pidato-pidato yang dipilih mencakup berbagai konteks historis dan sosial, mulai dari perjuangan hak-hak sipil hingga seruan untuk perdamaian. Analisis ini dilakukan untuk mengungkapkan sentimen, tema, dan emosi yang terkandung dalam setiap pidato, serta untuk memahami bagaimana orator menggunakan bahasa untuk mempengaruhi dan menginspirasi audiens mereka. Dengan memanfaatkan metode vektorisasi teks, ekstraksi topik menggunakan Latent Dirichlet Allocation (LDA), serta analisis sentimen dan emosi, penelitian ini mengidentifikasi pola dan tema yang umum dalam pidato-pidato tersebut. Hasil dari analisis ini menunjukkan bahwa pidato yang paling berpengaruh sering kali mengandung tema-tema yang kuat seperti keadilan sosial, kemerdekaan, dan persatuan. Selain itu, sentimen positif dan emosi kuat seperti harapan dan keberanian sering kali menonjol dalam pidato-pidato tersebut. Penelitian ini memberikan wawasan yang berharga bagi sejarawan, peneliti, dan penggemar oratori tentang bagaimana pidato ikonik telah membentuk opini publik dan mempengaruhi perubahan sosial.

Kata kunci: Pemrosesan Bahasa Alamliah, Analisis Sentimen, Ekstraksi Topik, Pengelompokan Teks

PENDAHULUAN

Pidato memiliki peran penting dalam sejarah umat manusia, digunakan oleh para pemimpin, aktivis, dan pemikir untuk menginspirasi, memotivasi, dan mengarahkan masyarakat. Pidato ikonik seperti "*I Have a Dream*" oleh Martin Luther King Jr. dan "*We Shall Fight on the Beaches*" oleh Winston Churchill telah meninggalkan jejak yang mendalam dalam ingatan kolektif masyarakat, berkontribusi pada perubahan sosial dan politik yang signifikan. Meskipun kekuatan pidato-pidato ini sering kali diakui secara luas, memahami elemen-elemen bahasa yang membuat mereka begitu berdampak membutuhkan analisis yang lebih mendalam.

Selama dekade terakhir, penelitian mengenai analisis pidato menggunakan metode *Natural Language Processing* (NLP) telah mengalami perkembangan pesat. Studi oleh Pennebaker dkk., (2015) mengeksplorasi hubungan antara pilihan kata dalam pidato politik dan persepsi publik, menunjukkan bahwa penggunaan kata-kata emosional dan moral memiliki dampak signifikan pada audiens. Sementara itu, penelitian oleh Biyani dkk., (2014) menggunakan analisis sentimen untuk menilai pidato politik di media sosial, menemukan bahwa pidato dengan sentimen positif cenderung lebih viral. Lebih lanjut, penelitian oleh Iyyer, dkk., (2014) mengembangkan model klasifikasi untuk mendeteksi niat dalam pidato, sementara Mac & Oki (2020) fokus pada analisis retorika dalam pidato kampanye presiden AS, menyoroti pentingnya pola linguistik dalam mempengaruhi pemilih.

Penelitian Jianqiang, dkk., (2018) tentang analisis sentimen dalam pidato politik mengungkapkan bagaimana perasaan dan sikap pembicara dapat diekstraksi dan dianalisis secara otomatis menggunakan teknik NLP. Munezero, dkk., (2014) memfokuskan pada analisis emosi dalam pidato motivasi, menemukan bahwa elemen emosional tertentu dalam pidato dapat meningkatkan keterlibatan dan motivasi pendengar.

Meski banyak penelitian telah dilakukan, terdapat kesenjangan dalam literatur mengenai analisis mendalam terhadap pidato-pidato ikonik dari berbagai era dan konteks menggunakan teknik NLP yang komprehensif. Sebagian besar studi fokus pada pidato politik kontemporer atau pidato dalam konteks tertentu tanpa menggabungkan berbagai dimensi

analisis seperti sentimen, topik, dan emosi dalam satu kerangka studi. Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengaplikasikan teknik NLP pada pidato-pidato ikonik dalam sejarah, dengan tujuan memahami bagaimana elemen-elemen linguistik berkontribusi pada kekuatan dan dampak pidato tersebut. Dengan memanfaatkan alat-alat seperti vektorisasi teks, *Latent Dirichlet Allocation* (LDA) untuk ekstraksi topik, serta analisis sentimen dan emosi, penelitian ini berusaha mengidentifikasi tema-tema umum dan sentimen yang mendasari pidato-pidato yang paling berpengaruh dalam sejarah.

Dengan pendekatan ini, diharapkan dapat mengungkap pola-pola linguistik yang mungkin tidak terlihat melalui analisis tradisional, serta memberikan wawasan yang lebih mendalam tentang kekuatan retorika dalam pidato ikonik dan bagaimana elemen-elemen bahasa tertentu dapat mempengaruhi audiens secara signifikan. Dalam konteks ini, NLP dapat menjadi alat yang kuat untuk menganalisis teks secara komprehensif.

KAJIAN PUSTAKA

Penelitian ini berlandaskan pada beberapa teori dan metode dalam bidang *Natural Language Processing* (NLP), seperti analisis sentimen, dan studi retorika. NLP adalah cabang dari kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia. *Natural Language Processing* melibatkan teknik untuk memahami, menafsirkan, dan menghasilkan bahasa alami dengan cara yang bermanfaat (Jurafsky & Martin, 2023). Beberapa teknik NLP yang sering dimanfaatkan adalah analisis sentiment, analisis emosi dan *Latent Dirichlet Allocation* (LDA).

Analisis sentimen adalah proses mengidentifikasi dan mengkategorikan opini yang diekspresikan dalam suatu teks, terutama untuk menentukan apakah sikap penulis terhadap subjek tertentu bersifat positif, negatif, atau netral. Teknik ini memungkinkan peneliti untuk mengukur emosi dan perasaan yang terkandung dalam pidato, memberikan wawasan tentang bagaimana orator mempengaruhi audiens mereka secara emosional (Zhang, Lei, dkk., 2018).

Analisis emosi adalah yang digunakan untuk mengidentifikasi dan mengukur emosi spesifik yang diekspresikan dalam teks. Pendekatan ini membantu dalam memahami bagaimana pidato memanfaatkan emosi untuk menciptakan hubungan dengan audiens dan mempengaruhi mereka. Ekstraksi emosi dalam teks dapat dilakukan menggunakan pendekatan berbasis leksikon atau model pembelajaran mesin (Buechel & Hahn, 2016).

Latent Dirichlet Allocation (LDA) adalah model generatif yang digunakan untuk pemodelan topik dalam kumpulan dokumen. Model ini mengasumsikan bahwa setiap dokumen terdiri dari campuran beberapa topik, dan setiap topik terdiri dari sekumpulan kata dengan probabilitas tertentu. Proses ini melibatkan dua tahap utama: pertama, menentukan distribusi topik untuk setiap dokumen; kedua, menentukan distribusi kata untuk setiap topik. Model ini membantu mengungkapkan struktur tersembunyi dari dokumen, memungkinkan identifikasi tema atau topik utama yang dibahas dalam teks (Jelodar, dkk., 2019). Dalam penelitian ini, LDA digunakan untuk mengidentifikasi tema-tema utama yang terdapat dalam sejumlah pidato ikonik, sehingga memudahkan pemahaman tentang isu-isu yang ditekankan oleh para orator.

Selain konsep di atas, salah satu teori lain yang dimanfaatkan dalam penelitian ini adalah teori retorika yang mempelajari seni persuasi melalui penggunaan bahasa, berfokus pada cara-cara untuk meyakinkan atau mempengaruhi audiens. Aristoteles mengidentifikasi tiga elemen utama dalam retorika: *ethos* (karakter dan kredibilitas pembicara), *pathos* (kemampuan untuk membangkitkan emosi audiens), dan *logos* (penggunaan logika dan alasan). Pendekatan ini membantu dalam menganalisis bagaimana struktur dan gaya bahasa dalam pidato berkontribusi terhadap efektivitas dan dampaknya (Tindale, 2019).

Para peneliti, telah melakukan sejumlah penelitian terkait pemanfaatan NLP, diantaranya adalah (L Bnoit., 2015) mengeksplorasi penggunaan gaya bahasa dalam pidato, menunjukkan bagaimana perubahan gaya bahasa dapat mencerminkan perubahan dalam konteks sosial dan politik. Penggunaan gaya bahasa untuk menyampaikan pesan-pesan yang relevan dengan isu-isu masyarakat seringkali

memberi pengaruh dalam kampanye politik (Enjang, 2022). Penyampaian dengan gaya politainment memiliki daya tarik yang kuat dan melekat diingatan masyarakat karena disajikan dengan format yang menghibur dan mengesankan (Indah, 2021).

Christopher dan kawan-kawan pada 2021 melakukan penelitian mengenai analisis otomatis emosi dalam pidato politik berdasarkan transkrip. Analisis sentimen otomatis digunakan secara luas dalam ilmu politik. Digitalisasi transkrip legislatif telah meningkatkan potensi penerapan alat yang ada untuk analisis otomatis emosi dalam teks. Hal ini menunjukkan bahwa teknik NLP dapat secara otomatis mengevaluasi perasaan dan sikap.

Eksplorasi penggunaan *Latent Dirichlet Allocation (LDA)* untuk ekstraksi topik dari koleksi dokumen menunjukkan bagaimana LDA dapat digunakan untuk mengidentifikasi tema utama dalam berbagai jenis teks, termasuk artikel ilmiah, berita, dan pidato publik (Blei & Jordan, 2003). Bhuvaneshwari dan kawan-kawan pada 2023 mengidentifikasi topik dalam pidato politik yang disampaikan oleh politisi dari partai yang sama dan membandingkannya dengan politisi dari partai lain di negara bagian Telangana. Eksperimen ini menggunakan alat NLP untuk analisis wacana politik, memberikan wawasan tentang bagaimana topik dan gagasan utama yang disampaikan oleh para pemimpin, yang memengaruhi pendengar.

Penelitian-penelitian tersebut menunjukkan bagaimana berbagai teknik NLP, analisis sentimen, dan analisis topik telah digunakan dalam berbagai konteks untuk memahami dan mengeksplorasi teks-teks penting.

METODE

Metode yang digunakan dalam penelitian ini terdiri dari beberapa langkah penting yang mencakup persiapan data, pra-pemrosesan teks, analisis sentimen, analisis emosi, pembuatan model topik, dan klustering. Berikut adalah uraian dari setiap tahapannya:

1. Persiapan Data

Data yang digunakan adalah kumpulan teks dari pidato-pidato ikonik yang disimpan dalam format CSV. Data diperoleh dari platform open data: Kaggle. Proses awal meliputi pembacaan

data dan penghapusan baris yang memiliki nilai kosong.

```
df = pd.read_csv ('50_Famous_Speechs.csv',
encoding='latin1')
df.dropna(inplace=True)
```

2. Pra-pemrosesan Teks

Pra-pemrosesan teks melibatkan tokenisasi, penghapusan stopwords, dan penghapusan tanda baca. Ini dilakukan untuk setiap teks dalam dataset.

```
all_stopwords =
set(stopwords.words('english')).union(addition
al_stopwords)
def preprocess_text(text):
tokens = word_tokenize(text.lower())
tokens = [token for token in tokens if token
not in all_stopwords]
tokens = [token for token in tokens if token
not in string.punctuation]
return tokens
texts = [preprocess_text(text) for text in
df['Speech']]
```

3. Analisis Sentimen

Analisis sentimen dilakukan menggunakan TextBlob untuk mendapatkan polaritas dan subjektivitas dari setiap teks pidato.

```
def analyze_sentiment(text):
analysis = TextBlob(text)
return analysis.sentiment.polarity,
analysis.sentiment.subjectivity
df['Sentiment_Polarity'],
df['Sentiment_Subjectivity'] =
zip(*df['Speech'].apply(lambda x:
analyze_sentiment('
.join(preprocess_text(x))))))
```

4. Analisis Emosi

Analisis emosi dilakukan dengan menghitung frekuensi kata-kata yang terkait dengan emosi tertentu dalam teks pidato. Empat kategori emosi yang dianalisis adalah 'joy', 'sadness', 'anger', dan 'fear'.

```
def analyze_emotion(text):
```

```
emotion_scores = {'joy': 0, 'sadness': 0,
'anger': 0, 'fear': 0}
total_words = 0
tokens = word_tokenize(text.lower())
total_words = len(tokens)
for word in tokens:
if word in emotions:
emotion_scores[emotions[word]] += 1
if total_words == 0:
return {'joy': 0, 'sadness': 0, 'anger': 0,
'fear': 0}
return {key: value/total_words for key,
value in emotion_scores.items()}
```

5. Pembuatan Model Topik

Model topik dibuat menggunakan Latent Dirichlet Allocation (LDA) untuk menemukan topik-topik yang mendasari kumpulan teks.

```
dictionary = corpora.Dictionary(texts)
corpus = [dictionary.doc2bow(text) for text in
texts]
lda_model = models.LdaModel(corpus,
num_topics=5, id2word=dictionary,
passes=15)
for idx, topic in lda_model.print_topics(-1):
print(f"Topic: {idx} \nWords: {topic}")
word_counts = df['Speech'].apply(lambda x:
len(word_tokenize(x)))
```

6. Clustering

Clustering dilakukan menggunakan K-Means berdasarkan distribusi topik yang diperoleh dari model LDA.

```
# Vektorisasi teks menggunakan
CountVectorizer
vectorizer = CountVectorizer
(max_df=0.95,min_df=2,stop_words='english')
dtm = vectorizer.fit_transform (df['Speech'])

# LDA untuk ekstraksi topik
num_topics = 5
lda =LatentDirichletAllocation
(n_components=num_topics,random_state=42)
lda.fit(dtm)

# Mengambil dan normalisasi distribusi topic
topic_distribution = lda.transform(dtm)
topic_distribution =
normalize(topic_distribution, norm='l1')
```

```
# Clustering dokumen menggunakan K-means
num_clusters = 5
km = KMeans(n_clusters= num_clusters,
random_state=42, n_init=10)
clusters = km.fit_predict (topic_distribution)
df['Cluster'] = clusters

# Reduksi dimensi menggunakan PCA.
pca = PCA(n_components=2)
reduced_topic_distribution =
pca.fit_transform(topic_distribution)
```

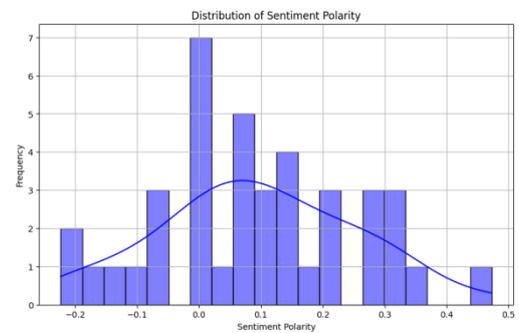
HASIL DAN PEMBAHASAN

Penelitian ini memberikan gambaran tentang isi dan tema pidato-pidato terkenal dalam “*Voices of History: 50 Iconic Speeches*”. Dari distribusi sentimen dan emosi, kita bisa melihat bagaimana pidato-pidato tersebut mencerminkan perasaan dan opini dari pembicara. Model topik LDA dan *clustering K-Means* membantu mengidentifikasi tema utama dan kelompok dalam pidato-pidato ini, memberikan gambaran umum tentang isu-isu yang paling sering dibahas. Visualisasi seperti awan kata dan histogram frekuensi kata membantu memperkuat temuan ini dengan cara yang mudah dipahami. Secara keseluruhan, analisis ini menunjukkan bahwa pidato-pidato ini sangat berfokus pada isu-isu sosial, politik, dan kemanusiaan, dengan berbagai emosi dan topik yang terlibat. Gambaran detail tentang hasil analisa dapat dijelaskan sebagai berikut:

1. Analisis Sentimen

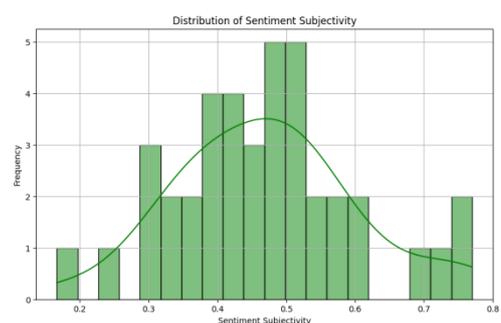
- **Distribusi Sentimen Polaritas:** Kami mengukur polaritas (positif atau negatif) dari setiap pidato menggunakan TextBlob. Nilai polaritas sentimen menunjukkan nuansa emosional dari masing-masing pidato. Polaritas positif (di atas 0) menunjukkan sentimen positif, sedangkan polaritas negatif (di bawah 0) menandakan sentimen negatif. Nilai polaritas mendekati atau sama dengan 0 mengindikasikan sentimen netral. Misalnya: untuk pidato bertema “*I Have a Dream from MLK*”, polaritas sentimennya adalah 0.09776, yang menunjukkan sentimen yang sedikit positif. Hasil menunjukkan bahwa sebagian besar pidato memiliki polaritas yang cukup netral, dengan beberapa pidato menunjukkan polaritas yang sangat positif

atau sangat negatif. Hal ini mengindikasikan bahwa pidato-pidato tersebut cenderung memiliki nada yang seimbang. Pidato dengan polaritas sangat positif seringkali menekankan harapan dan optimisme, sementara yang sangat negatif menyoroti tantangan dan kritik keras.



Gambar 1. Distribusi Sentimen Polaris

- **Distribusi Sentimen Subjektivitas:** untuk mengukur sejauh mana teks bersifat subjektif atau objektif. Sebagian besar pidato memiliki tingkat subjektivitas yang cukup tinggi, menunjukkan bahwa pidato ini cenderung mengandung opini dan perasaan pribadi daripada hanya fakta. Nilai subjektivitas sentimen berkisar dari 0 hingga 1, di mana 0 sangat objektif (faktual) dan 1 sangat subjektif (berpendapat atau emosional). Misalnya: untuk pidato ‘*Tilbury Speech*’ oleh Ratu Elizabeth I, subjektivitas sentimennya adalah 0.677366, menunjukkan konten yang lebih subjektif atau emosional. Pidato dengan subjektivitas tinggi sering digunakan untuk menginspirasi atau mempengaruhi audiens dengan menyam-paikan perasaan dan keyakinan pem-bicara.



Gambar 2. Distribusi Sentimen Subjektivitas

6. Model Topik dengan LDA

Kami menggunakan *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi lima topik utama dalam pidato-pidato ini. Topik-topik ini mencakup kata-kata seperti "freedom", "nation", "people", dan "world", mengindikasikan bahwa isu-isu ini sangat dominan dalam pidato-pidato tersebut. Topik-topik seperti kebebasan, perjuangan, harapan, dan persatuan muncul, mencerminkan keprihatinan dan aspirasi yang umum yang diangkat oleh para pembicara.

Tabel 1. Model Topik dengan LDA

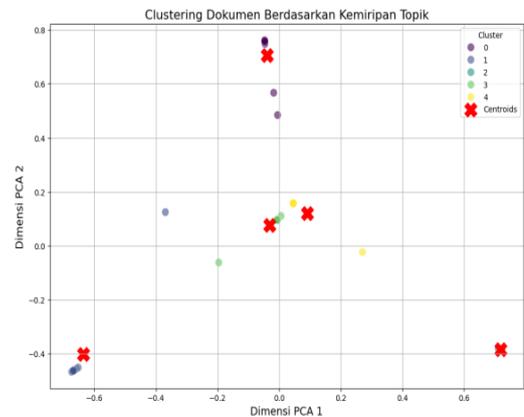
No.	Model Topik
0	Words: 0.012*"women" + 0.008*"world" + 0.006*"right" + 0.004*"peace" + 0.004*"country" + 0.004*"america" + 0.004*"men" + 0.004*"free"
1	Words: 0.014*"party" + 0.013*"war" + 0.011*"hope" + 0.006*"russia" + 0.005*"risk" + 0.004*"struggle" + 0.004*"great" + 0.004*"army" + 0.004*"fighting"
2	Words: 0.007*"freedom" + 0.006*"great" + 0.006*"woman" + 0.006*"power" + 0.005*"women" + 0.004*"world" + 0.004*"congress"
3	Words: 0.007*"world" + 0.006*"god" + 0.005*"women" + 0.005*"love" + 0.005*"oligarchy" + 0.005*"fear" + 0.004*"future"
4	Words: 0.009*"black" + 0.007*"freedom" + 0.007*"rise" + 0.006*"country" + 0.006*"right" + 0.006*"men" + 0.006*"struggle" + 0.005*"white" + 0.005*"hope" + 0.005*"world"

Hasil pemodelan topik menunjukkan 5 topik teratas yang diidentifikasi dari pidato-pidato tersebut. Setiap topik diwakili oleh daftar kata kunci (istilah) yang paling khas dari topik tersebut, beserta bobotnya (probabilitas).

7. Klustering Teks Berdasarkan Topik

Dalam penelitian ini, digunakan K-Means klustering untuk mengelompokkan pidato berdasarkan distribusi topik yang dihasilkan oleh LDA. Hasil clustering menunjukkan adanya lima kelompok utama dalam pidato-pidato tersebut. Setiap kluster memiliki

karakteristik topik yang berbeda, yang dapat dilihat dari *centroid* atau pusat massa dari setiap cluster. Misalnya, satu kluster didominasi oleh pidato yang berfokus pada kebebasan dan hak asasi manusia, sementara kluster lain lebih banyak tentang isu-isu nasional dan patriotisme.



Gambar 7. Dokumen Klustering Berdasarkan Kemiripan Topik

Gambar 7 memperlihatkan pengelompokan pidato-pidato menggunakan metode K-Means, berdasarkan hasil tersebut terdapat beberapa pola yang dapat diamati:

Cluster 0:

Dokumen-dokumen dalam kluster ini cenderung berkaitan dengan kekuatan perempuan, hak asasi manusia, dan isu-isu sosial. Pidato-pidato seperti "Tilbury Speech" oleh Queen Elizabeth I, "The Audacity of Hope" oleh Barack Obama, dan "Nobel Speech" oleh Malala Yousafzai menyoroti tema-tema seperti kesetaraan, keadilan sosial, dan pengaruh perempuan dalam masyarakat. Sentimen yang dominan di sini adalah inspiratif dan memotivasi untuk perubahan sosial.

Cluster 1:

Kluster ini didominasi oleh pidato-pidato yang memiliki fokus pada perjuangan politik dan hak asasi manusia. Contohnya termasuk "I have a dream" oleh Martin Luther King Jr., "The Gettysburg Address" oleh Abraham Lincoln, dan "I Am Prepared to Die" oleh Nelson Mandela. Pidato-pidato ini sering kali mengekspresikan aspirasi untuk perubahan sosial dan politik yang lebih besar. Sentimen yang dominan di sini adalah penuh semangat,

aspiratif, dan menginspirasi untuk perubahan sosial dan politik.

Cluster 2:

Dokumen-dokumen dalam klaster ini mencakup pidato-pidato yang mengangkat isu-isu feminisme, perjuangan hak-hak perempuan, dan kebebasan. Termasuk di antaranya adalah "*Ain't I A Woman?*" oleh Sojourner Truth, "*Freedom or Death*" oleh Emmeline Pankhurst, dan "*Nobel Lecture*" oleh Mother Teresa. Klaster ini menyoroti perjuangan sejarah untuk kesetaraan gender dan kebebasan individual. Sentimen yang ada di sini mencakup pendorong perubahan sosial, reflektif, dan filosofis.

Cluster 3:

Klaster ini terdiri dari pidato-pidato yang fokus pada kebijakan politik dan strategi, seperti pidato "*Their Finest Hour*" oleh Winston Churchill, "*Atoms for Peace Speech*" oleh Dwight Eisenhower, dan "*Speech on Vietnam*" oleh Lyndon Johnson, serta pernyataan politik yang kuat seperti "*The Ballot or The Bullet*" oleh Malcolm X. Pidato-pidato ini menunjukkan pemikiran strategis dan politik dari para pemimpin dalam situasi-situasi penting. Sentimen yang dominan di sini adalah strategis, taktis dalam implikasinya terhadap kebijakan dan politik.

Cluster 4:

Dokumen-dokumen dalam klaster ini cenderung mengangkat tema-tema edukasi, ilmu pengetahuan, dan inspirasi pribadi. Terdapat juga topik lainnya seperti perjuangan hak-hak sipil dan hak asasi manusia. Contohnya termasuk "*Questioning the Universe*" oleh Stephen Hawking, "*Final Commencement Speech*" oleh Michelle Obama dan "*Still I Rise*" oleh Maya Angelou. Klaster ini menyoroti pengaruh dan inspirasi dalam konteks pendidikan dan pengetahuan. Sentimen yang mungkin dominan di sini adalah determinasi, perjuangan, dan keadilan sosial.

Analisis Umum:

Klaster-klaster ini menggambarkan keberagaman topik dan tema yang dihadirkan oleh pidato-pidato ikonik tersebut. Meskipun ada beberapa overlap dalam tema seperti hak asasi manusia dan keadilan sosial, setiap klaster

menunjukkan fokus yang unik terhadap isu-isu tertentu. Hal ini mencerminkan kompleksitas dalam pengaruh dan dampak yang dimiliki oleh pidato-pidato ini terhadap sejarah dan masyarakat. Melalui analisis ini, kita dapat memahami bagaimana berbagai topik dan tema dalam pidato ikonik dikelompokkan bersama berdasarkan kemiripan topik mereka, yang dapat memberikan wawasan yang lebih dalam tentang pesan-pesan yang disampaikan oleh para orator kepada audiens mereka.

SIMPULAN

Penelitian ini menghasilkan pemahaman yang mendalam tentang isi dan tema dari kumpulan pidato terkenal. Berikut adalah beberapa kesimpulan utama dari analisis yang telah dilakukan:

1. **Distribusi Sentimen dan Emosi:** Pidato-pidato ini menunjukkan variasi dalam polaritas sentimen dan emosi yang digunakan. Sebagian besar pidato memiliki polaritas netral atau sedikit positif, dengan emosi *joy* dan *anger* yang dominan. Ini mengindikasikan bahwa pidato-pidato ini mencoba untuk menyampaikan pesan dengan menggabungkan elemen-elemen positif dan kritik.
2. **Topik Utama Pidato:** Berdasarkan model LDA, kita dapat mengidentifikasi lima topik utama dalam pidato-pidato ini, seperti kebebasan, identitas nasional, hak asasi manusia, dan masalah global. Hal ini menunjukkan bahwa pidato-pidato ini sering mengangkat isu-isu yang penting secara sosial dan politik.
3. **Awan Kata dan Frekuensi Kata:** Kata-kata seperti "*freedom*", "*nation*", "*people*", dan "*world*" adalah kata-kata yang paling sering muncul dalam pidato-pidato ini. Ini mencerminkan fokus utama pidato-pidato tersebut pada isu-isu yang mencakup kebebasan, identitas nasional, dan kepedulian terhadap masyarakat global.
4. **Klustering Dokumen:** Melalui teknik K-Means klustering, pidato-pidato ini dapat dikelompokkan berdasarkan kesamaan topiknya. Ini membantu dalam memahami bagaimana tema-tema yang berbeda saling terkait dan bagaimana pidato-pidato ini bisa dikelompokkan berdasarkan konteks dan fokusnya.

5. Penelitian ini diharapkan dapat memberikan kontribusi signifikan bagi bidang studi retorika, sejarah, dan komunikasi, serta memberikan wawasan praktis bagi para pemimpin dan orator masa kini tentang cara menggunakan bahasa untuk mempengaruhi dan menginspirasi audiens mereka. Selain itu, penelitian ini juga menunjukkan potensi penerapan teknik NLP dalam analisis teks historis dan kontemporer.

SARAN

Berikut adalah beberapa saran untuk penelitian selanjutnya berdasarkan temuan dari penelitian ini. Pertama, pengembangan model sentimen yang lebih kompleks dapat menangkap nuansa dan perubahan dalam sentimen sepanjang pidato, memberikan pemahaman lebih dalam tentang perkembangan perasaan pembicara dari awal hingga akhir. Kedua, analisis komparatif antara pidato dari berbagai pembicara atau waktu berbeda dapat membantu melihat perubahan dan konsistensi dalam tema dan pesan yang disampaikan. Ketiga, analisis sentimen yang terfokus pada topik-topik spesifik, seperti kebebasan, hak asasi manusia, atau isu-isu global, dapat memberikan wawasan lebih mendalam tentang perkembangan sentimen dalam konteks tertentu. Keempat, penggunaan teknik NLP lanjutan, seperti pengenalan entitas, pemodelan bahasa canggih, atau analisis sintaksis, dapat menggali lebih dalam struktur dan hubungan dalam teks pidato. Terakhir, pengujian terhadap kumpulan data yang lebih luas, mencakup pidato dari berbagai budaya atau konteks politik yang berbeda, dapat mengungkap bagaimana perbedaan tersebut mempengaruhi tema dan sentimen dalam pidato.

DAFTAR PUSTAKA

- Bhuvaneshwari Kondeti, S.A. Jyothirani, Haragopal Venkata Vajjha. (2023). Topic Modelling from Repository of Political Speeches. *International Journal For Innovative Research In Multi-disciplinary Field*, ISSN: 2455-0620, Volume 9, Issue 9.
- Biyani, P., Bhatia, S., Caragea, C., & Mitra, P. (2014). Using Non-lexical Features for Identifying Factual and Opinionative Threads in Online Forums. *Knowledge-Based Systems*, 91, 49-56.
- Blei, D. M., & Jordan, M. I. (2003). Latent Dirichlet Allocation for Topic Modeling. *Journal of Machine Learning Research*, 3, 993-1022.
- Buechel, S., & Hahn, U. (2016). Emotion Analysis as a Regression Problem: Dimensional Models and Their Implications on Emotion Representation and Metrical Evaluation. *Proceedings of the 22nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 111-121.
- Cochrane, C., Rheault, L., Godbout, J. F., Whyte, T., Wong, M.W. C., & Borwein, S. (2021). The Automatic Analysis of Emotion in Political Speech Based on Transcripts. *Political Communication*, 39(1), 98-121. <https://doi.org/10.1080/10584609.2021.1952497>.
- Enjang Pera Irawan (2022), Political Campaign Strategy of Young PKS Politicians In Seeking Constituent Support, *Jurnal Ilmiah Manajemen Informasi dan Komunikasi*, Vol 6, No. 1, ISSN : 2614-1272.
- Indah Suryawati (2021), Political Entertainment Of Central Parlemen Celebrity In Online Media Construction, *Jurnal Ilmiah Manajemen Informasi dan Komunikasi*, Vol 5, No. 1, ISSN : 2614-1272.
- Iyyer, M., Enns, P., Boyd-Graber, J., & Resnik, P. (2014). Political Ideology Detection Using Recursive Neural Networks. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 1113-1122.
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey. *Multimedia Tools and Applications*, 78(11), 15169-15211.
- Jianqiang, Z., Xiaolin, G., & Xuejun Z. (2018). Deep Convolution Neural Networks for Twitter Sentiment Analysis. *Applied Sciences*, 6(3), 52.
- Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd ed.). Pearson.

- Mac Aditiawarman, Oki Agustriawan (2020), An Analysis of Speech The Presidential Candidates Debate of United States in 2020. *Jurnal Ilmiah Langue and Parole*, Volume 6, Nomor 1, E-ISSN : 2581-1819.
- Munezero, M., Montero, C. S., Sutinen, E., & Pajunen, J. (2014). Are They Different? Affect, Feeling, Emotion, Sentiment, and Opinion Detection in Text. *IEEE Transactions on Affective Computing*, 5(2), 101-111.
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC 2015*. Austin, TX: University of Texas at Austin.
- Tindale, C. W. (2019). *The Art of Persuasion in Aristotle's Rhetoric*. Routledge.
- William L Benoit (2015). *Political Election Debates: Informing Voters about Policy and Character*. Lexington Books, University of Alabama Birmingham, ISBN-13978-1498515610.
- Zhang, L., Wang, S., & Liu, B. (2018). Deep Learning for Sentiment Analysis: A Survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253.