



Papua dalam Perspektif Komentar Youtube: Studi Pemodelan Topik dan Analisis Sentimen dengan Pendekatan Text Mining

Yan Nazala Bisoumi¹, Jimy Munandar², Setiawan Amrullah³, Muhammad Tegar Pandiriyani⁴, Kaia Raissa Akmalia⁵, Fatkhurrohman Fauzi⁶

^{1, 2, 3, 4, 5, 6} Program Studi Statistika Universitas Muhammadiyah Semarang

¹ ynazalabisoumi@gmail.com

² jimymunandar11@gmail.com

³ setiawanamrullah17@gmail.com

⁴ moh.tegarr@gmail.com

⁵ kaiaraissaakmalia@gmail.com

⁶ fatkhurokhanff@unimus.ac.id

Corresponding author email: ynazalabisoumi@gmail.com

Abstract: This research analyzes comments about Papua on the YouTube platform using a text mining approach to identify the main topics in the comment section discussions. The best topics are determined by comparing the accuracy of the Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) methods. The results show that the NMF model is superior with a coherence score of 0.8546 for 5 topics, compared to LDA with a coherence score of 0.7710 for 10 topics. However, the highest coherence score for the best LDA topic reached 0.9702, surpassing NMF at 0.9407. The topics from the LDA model include words such as "no, in, and, they, which," while NMF includes "ripe, mission, boiler, white, red." This difference indicates variations in the sentiment analysis results. Sentiment analysis using the VADER method was applied to the best NMF model and showed that the five topics tend to have a neutral sentiment with an average of 85.92%. This research provides in-depth insights into public perception of Papua and demonstrates the effectiveness of the NMF method in topic modeling.

Keywords: Papua issues, Topic modeling, Sentimen Analysis, Text Mining, Youtube

Abstrak: Penelitian ini menganalisis komentar tentang Papua di platform YouTube menggunakan pendekatan *text mining* untuk mengidentifikasi topik utama dalam diskusi kolom komentar. Topik terbaik ditentukan dengan membandingkan akurasi metode *Latent Dirichlet Allocation* (LDA) dan *Non-Negative Matrix Factorization* (NMF). Hasil menunjukkan bahwa model NMF lebih unggul dengan nilai *coherence* sebesar 0.8546 untuk 5 topik, dibandingkan LDA dengan nilai *coherence* 0.7710 untuk 10 topik. Namun, nilai *coherence* tertinggi pada topik terbaik LDA mencapai 0.9702, mengungguli NMF yang sebesar 0.9407. Topik dari model LDA mencakup kata-kata seperti "tidak, di, dan, mereka, yang", sedangkan NMF mencakup "matang, misi, kualiti, putih, merah". Perbedaan ini menunjukkan adanya variasi dalam hasil analisis sentimen. Analisis sentimen menggunakan metode VADER diterapkan pada model NMF terbaik dan menunjukkan bahwa kelima topik cenderung memiliki sentimen netral dengan rata-rata 85,92%. Hasil penelitian ini memberikan wawasan mendalam mengenai persepsi publik terhadap Papua dan menunjukkan efektivitas metode NMF dalam pemodelan topik.

Kata kunci: Papua, Pemodelan Topik, Analisis Sentimen, *Text Mining*, Youtube

I. PENDAHULUAN

Pulau Papua merupakan pulau terbesar nomor dua di dunia setelah Greenland yang terletak di utara Australia. Pulau Papua mempunyai luas total 785.753 km² yang terbagi menjadi dua bagian. Bagian pertama yaitu wilayah yang masuk Negara Indonesia kurang lebih seluas 418.707,7 km² dan bagian kedua kurang lebih seluas 367.045,3 km² merupakan wilayah Negara Papua Nugini. Meskipun Negara Papua Nugini berada di Pulau Papua, namun istilah "Papua" dewasa ini digunakan guna merujuk wilayah Indonesia yang ada dalam pulau tersebut. Wilayah yang luas menjadikan Papua mengalami beberapa kali pemekaran provinsi dan beragamnya suku bangsa [1]. Wilayah hasil pemekaran tersebut hingga saat ini sudah terbentuk enam provinsi yang sudah sah secara administratif yaitu: Provinsi Papua, Provinsi Papua Barat, Provinsi Papua Barat Daya, Provinsi Papua Pegunungan, Provinsi Papua



Selatan dan Provinsi Papua Tengah [2]. Suku bangsa yang tercatat tinggal di Papua kurang lebih sebanyak 255 suku dengan penduduk asli Papua yang disebut Orang Asli Papua (OAP) [1].

Pulau Papua merupakan pulau yang indah dengan kepadatan penduduk sesuai data BPS tahun 2021 hanya sekitar 11-14 jiwa/km². Jumlah penduduk yang sedikit dapat menimbulkan efek positif dan negatif bagi Papua. Dampak positif dari hal tersebut adalah terjaganya kekayaan alam yang terdapat di Papua, terbukti dengan konsep kehidupan penduduknya yang menjadikan tanah, air dan hutan sebagai ibu bagi mereka [3]. Prinsip tersebut menjadikan penduduknya senantiasa menjaga alam mereka karena percaya dengan konsep timbal balik [3]. Banyak wilayah di Papua yang tidak terkena dampak negatif aktivitas manusia sebab proporsi jumlah penduduk dan luas wilayah yang tidak sepadan. Selain memiliki alam yang indah, banyak pula karya seni yang dimiliki oleh Papua. Karya seni tersebut berupa kebudayaan, musik daerah maupun *modern*, adat istiadat, dan lain-lain. Selain dampak positif dan kelebihan yang tampak, terdapat pula dampak negatif dari rendahnya angka kepadatan penduduk di wilayah Papua. Isu ketimpangan-ketimpangan yang terjadi di Indonesia merupakan salah satu dampak negatifnya [4]. Isu-isu terkait ketidakratahan pembangunan, ketergantungan pangan, tingginya harga bahan pokok di Papua merupakan sebagian kecil permasalahan yang sering menjadi sorotan [4]. Adanya konflik militer yang terjadi di Papua dengan munculnya Kelompok Kriminal Separatis Bersenjata (KKB), Kelompok Separatis Teroris Papua (KSTP) dan Organisasi Papua Merdeka (OPM) juga menjadi sorotan beberapa dekade terakhir karena kekerasan terhadap manusia dan pelanggaran HAM yang tak kunjung mereda [4][5]. Isu terbaru yang sedang diperbincangkan adalah mengenai penggusuran lahan.

Informasi dari sisi positif dan negatif [6] Papua selalu beredar di media sosial tanpa jeda waktu [7], [8]. Salah satu platform media sosial yang menyediakan kedua informasi tersebut adalah *YouTube* [9]. *YouTube* merupakan situs yang menyediakan berbagai video-video menarik milik perusahaan *Google Inc* [10]. Konten yang tersedia dalam platform *YouTube* lebih beragam dibandingkan dengan media sosial lain seperti *Instagram*, *facebook*, *X* dll [10]. Konten yang dikhususkan untuk *meng-upload* atau menonton video dan audio dengan durasi panjang, sudah pasti menjadikan *YouTube* media informasi visual yang lebih unggul dari media sosial lain. Keberagaman video yang disediakan di *Youtube* memudahkan untuk mengakses informasi baik yang berkaitan dengan musik, kebudayaan, sarana belajar maupun berita mengenai isu-isu terbaru yang sedang diperbincangkan [10]. Setiap konten video yang diupload ke *YouTube* akan tersedia komentar dari topik yang di bahas. Komentar tersebut ada yang positif, negatif atau bahkan netral [11], [12], [13].

Perbincangan topik yang beragam mengenai Papua di *platform YouTube* ini, mendorong penulis untuk melakukan penelitian mengenai pemodelan topik dan analisis sentimen terhadap masing-masing topik yang dibahas. Pemodelan topik yang digunakan adalah metode *Latent Dirichlet Allocation (LDA)* dan *Non-Negative Matrix Factorization (NMF)* [14]. Metode yang digunakan tersebut diharapkan dapat mengetahui kompleksitas topik Papua, sebelum dilakukan analisis sentimen tiap topik yang dibahas. Analisis sentimen menggunakan metode *VADER* [15], [16], [17], [18] bertujuan untuk secara otomatis mengklasifikasikan topik yang diperoleh menjadi sentimen positif, negatif, atau netral. Penelitian ini bertujuan mengidentifikasi topik-topik yang dibahas pada platform *YouTube* secara keseluruhan mengenai Papua. Topik-topik terpilih akan diklasifikasikan untuk mengetahui emosi atau sentimen yang terkandung dalam opini yang disampaikan melalui kolom komentar *Youtube*, sehingga diketahui emosi positif, negatif atau netral terhadap topik yang dibahas.



II. METODE PENELITIAN

2.1. Jenis Data dan Sumber Data Penelitian

Penelitian mengenai prespektif masyarakat terhadap Papua ini menggunakan data primer berupa text. Data yang diperoleh merupakan komentar pengguna Youtube di berbagai chanel. Periode waktu yang digunakan adalah dari 1 Januari 2024 hingga 31 Mei 2024.

2.2. Tahapan Analisis Data

Tahapan analisis data yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Mengumpulkan data

Data yang digunakan dalam penelitian ini berasal dari hasil *scraping YouTube* dengan *auth-token*. Pengambilan data hanya difokuskan pada satu *keyword* yaitu “Papua” tanpa memberi filter pada *chanel* tertentu. Data yang diperoleh sebanyak 142.804 komentar *YouTube* yang kemudian disimpan dalam bentuk *Comma Separated Value (CSV)*.

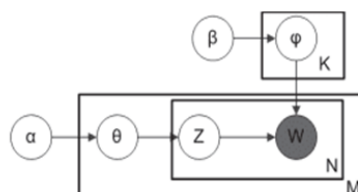
Tabel 1. Preview 5 Komentar Teratas pada Hasil *Scraping*

<i>Author</i>	<i>Published At</i>	<i>Comment</i>	<i>Video Title</i>	<i>Channel Title</i>
@hashpling	2024-05-30T17:06:32Z	<ahref="https://www.youtube.com/watch?v=i5Hwdb4ufEg&mp	<i>The Gangs of Papua New Guinea</i>	<i>Best Document ary</i>
@Identificador4225	2024-05-30T12:25:56Z	<i>Luego quejaros cuando vengan a europa a saquear</i>	<i>The Gangs of Papua New Guinea</i>	<i>Best Document ary</i>
@rasyidirahman940	2024-05-30T10:19:52Z	<ahref="https://www.youtube.com/watch?v=i5Hwdb4ufEg&mp	<i>The Gangs of Papua New Guinea</i>	<i>Best Document ary</i>
@giancarlomaster2535	2024-05-29T22:00:29Z	<i>Ahora entiendo porque las montaAsas â°i, □ sepultaron 2000personas inocente</i>	<i>The Gangs of Papua New Guinea</i>	<i>Best Document ary</i>
@user-gl7we7nb9u	2024-05-29T19:59:41Z	<i>We got crips and bloods. They got papas and noogies</i>	<i>The Gangs of Papua New Guinea</i>	<i>Best Document ary</i>

2. Mengestimasi model terbaik dari teknik pemodelan text dengan metode LDA dan NMF

- Metode LDA

Latent Dirichlet Allocation (LDA) adalah salah satu metode yang populer dalam *topic modeling* [19]. LDA adalah suatu probabilitas *generative model* yang memungkinkan dokumen untuk dirpresentasikan sebagai kombinasi topik-topik [20]. LDA yang mengekspresikan masing-masing topik sebagai distribusi probabilitas terhadap kata-kata dalam kamus, namun mengalami penurunan kinerja yang signifikan pada teks pendek karena kekurangan informasi mengenai kemunculan kata-kata secara bersamaan. LDA digunakan untuk menyimpulkan korelasi statistik antara saluran visual dan kata kunci [21].



Gambar 1. Model Probabilistik dari Latent Dirichlet Allocation [20]



Berdasarkan Gambar 1, a dan b adalah parameter pada tingkat korpus. q adalah variabel pada tingkat dokumen. Sementara z dan w adalah variabel pada tingkat kata. Variabel adalah simbol objek yang dapat diisi dengan konten alternatif tetapi memiliki distribusi tertentu. Dalam hal ini, z dan w dapat diubah atau ditukar. Karena perubahan nilai (mengikuti distribusi), sebuah variabel memiliki parameter tertentu. Distribusi *joint* dari DAG pada Gambar 1 dapat dituliskan dalam (1).

$$p(z, \theta, \varphi | \alpha, \beta) = \int_0^{\infty} p(\theta | \alpha) \left(\prod_j \sum_{z_k} p(z_k | \theta) p(\varphi_j | \beta) p(w | z_k, \varphi) \right) \quad (1)$$

- $p(\theta | \alpha)$ = distribusi yang mendefinisikan gabungan topik
- $p(z_k | \theta)$ = topik probabilitas z_k untuk gabungan ini
- $p(w | z_k, \varphi)$ = probabilitas item yang diberikan topik
- α = distribusi *Dirichlet*
- β = menentukan probabilitas $p(w | z_k, \varphi)$ yang ingin dicari

- Metode *Non-Negative Matrix Factorization* (NMF)

Non-Negative Matrix Factorization (NMF) merupakan model statistik yang dikembangkan untuk faktorisasi matriks pada data yang mempunyai struktur kompleks seperti data text [22]. *Non-Negative Matrix Factorization* (NMF) digunakan untuk mereduksi dimensi dan pemodelan topik dengan cara membuat faktor sebuah matriks menjadi dua matriks *non-negatif* [23], [24]. Teknik pembelajaran Mechine Learning pada metode *Non-Negative Matrix Factorization* (NMF) difokuskan untuk menemukan topik-topik tersirat atau topik yang tidak langsung terlihat di dalam data text yang digubakan (*hidden topic*) [23]. *Non-Negative Matrix Factorization* (NMF) yang berhasil menemukan topik-topik laten [25], dapat menjelaskan hubungan kompleks antar kata dan memberikan akurasi yang lebih tinggi dan terperinci.

$$\begin{array}{ccccccc}
 & \text{matriks } N \times P & & \text{matriks } N \times K & & & \text{matriks } K \times P \\
 x_{1,1} & \cdots & \cdots & \cdots & x_{1,K} & \theta_{1,1} & \cdots & \theta_{1,K} & & \phi_{1,1} & \cdots & \cdots & \cdots & \phi_{1,P} \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \times & \vdots & \vdots & \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \phi_{K,1} & \cdots & \cdots & \cdots & \phi_{K,P} \\
 x_{N,1} & \cdots & \cdots & \cdots & x_{N,P} & \theta_{N,1} & \cdots & \theta_{N,K} & & & & & &
 \end{array} \approx \quad (2)$$

3. Menghitung *topic coherence* terbaik dari hasil pemodelan.

Topik yang dihasilkan dari pemodelan tidak selalu bisa di interpretasikan Oleh karena itu, perhitungan *coherence* topik digunakan untuk membedakan antara topik yang baik dan buruk [20], [24]. *Coherence* topik mengukur tingkat kesamaan makna antara kata-kata dengan skor tertinggi dalam suatu topik [26], [27].

$$coherence(V) = \sum_{(v_i, v_j) \in V} score(v_i, v_j, \epsilon) \quad (3)$$

$$score(v_i, v_j, \epsilon) = \log \frac{D(v_i, v_j) + \epsilon}{D(v_j)} \quad (4)$$



V = Nilai *coherence*

$D(v_i, v_j)$ = jumlah kata dari v_i dan v_j

$D(v_j)$ = jumlah kata yang muncul dari v_j

4. Menganalisis sentimen topik

Model Topik terbaik yang telah diidentifikasi pada tahapan sebelumnya kemudian diklasifikasi sentimennya, apakah dihasilkan kecenderungan sentimen positif, negatif maupun netral. Analisis sentimen topik dalam penelitian ini menggunakan metode VADER [28], [29] guna mengetahui emosi tersirat dari opini yang tertulis di kolom komentar YouTube.

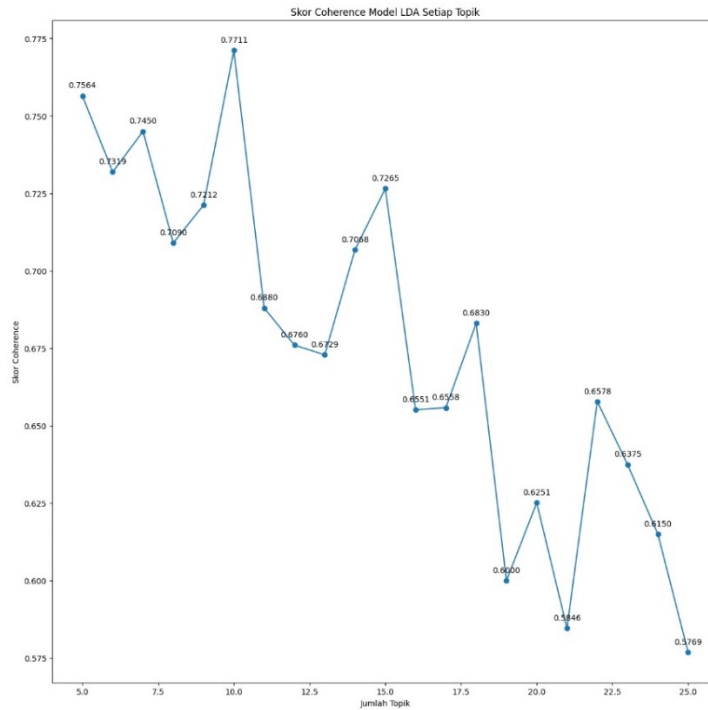
III. HASIL DAN PEMBAHASAN

Penelitian ini menganalisis topik “Papua” dengan *scraping YouTube* pada rentang waktu 1 Januari 2024 hingga 31 Mei 2024. Data yang diperoleh sebanyak 142.804 komentar dan dilakukan preprocessing serta beberapa filter sehingga dihasilkan data bersih sebanyak 50.287 data. Setelah dilakukan analisis deskriptif, kemudian dilakukan analisis topik dengan metode LDA dan NMF dari rentang 5 topik hingga 25 topik. Penentuan topik terbaik akan dilihat dari nilai *coherence* tertinggi yang dihasilkan oleh kedua metode.

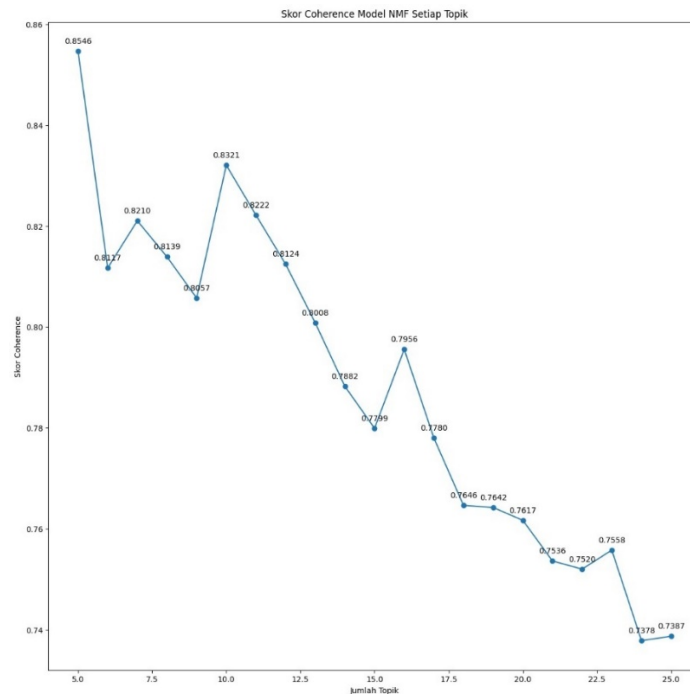
Tabel 2. Perbandingan Model Terbaik berdasarkan Nilai *Coherence*

Jumlah Topik Pemodelan Metode LDA	Nilai Coherence	Jumlah Topik Pemodelan Metode NMF	Nilai Coherence
5	0.7564	5	0.8546
6	0.7319	6	0.8117
7	0.7450	7	0.8210
8	0.7090	8	0.8139
9	0.7212	9	0.8057
10	0.7711	10	0.8321
11	0.6880	11	0.8222
12	0.6760	12	0.8124
13	0.6729	13	0.8008
14	0.7068	14	0.7882
15	0.7265	15	0.7799
16	0.6551	16	0.7956
17	0.6558	17	0.7780
18	0.6830	18	0.7646
19	0.6000	19	0.7642
20	0.6251	20	0.7617
21	0.5846	21	0.7536
22	0.6578	22	0.7520
23	0.6375	23	0.7558
24	0.6150	24	0.7378
25	0.5769	25	0.7387

Dari Tabel 2. dapat dilihat perbandingan model dari kedua metode yang digunakan. Model terbaik yang diperoleh dari metode LDA adalah 10 topik dengan nilai *coherence* 0.7711. Model terbaik dari metode NMF adalah 5 topik dengan nilai *coherence* sebesar 0.8546. Hasil perbandingan tersebut menunjukkan bahwa pemodelan topik dengan metode NMF lebih baik daripada model yang dihasilkan oleh metode LDA.



Gambar 2. Grafik nilai Coherence Metode LDA



Gambar 3. Grafik nilai Coherence Metode LDA

Jumlah topik terbaik masing-masing metode dilihat kata kuncinya sebanyak 5 kata dan dilihat kembali nilai *coherence*-nya untuk mengetahui model terbaik dalam setiap topiknya.



Tabel 3. Perbandingan Model Terbaik masing-masing topik setiap Metode berdasarkan Nilai *Coherence*

Topik	Model LDA	Nilai Coherence	Model NMF	Nilai Coherence
1	tidak, di, dan, mereka, yang	0.9702	semangat, moga, sehat, bang, bobon	0.8166
2	Papua, bang, moga, sehat, bobon	0.6987	kkn, tni, Papua, ui, bem	0.8565
3	ui, ham, bem, kkb, tni	0.8739	bon, banget, bang, keren, lu	0.7864
4	konten, lu, keren, bobon, bang	0.7579	matang, misi, kualiti, putih, merah	0.9407
5	bang, yesus, kati, bobon, tuhan	0.8085	bobon, berkat, yesus, kati, tuhan	0.8730
6	nya, lagu, indonesia, rohingnya, Papua	0.4613	-	-
7	Papua, program, prabowo, gratis, makan	0.7684	-	-
8	Papua, berani, doang, ui, bem	0.8476	-	-
9	dia, di, ini, itu, yang	0.9364	-	-
10	bang, kkn, bem, ui, Papua	0.5886		

Berdasarkan analisis pemodelan topik didapatkan bahwa model terbaik metode LDA adalah 10 topik dengan topik terbaik berada pada topik pertama. Nilai *coherence* dari topik terbaik menggunakan metode LDA adalah sebesar 0.9702. Sedangkan, dari lima topik pemodelan dengan metode NMF, model terbaik didapatkan pada topik keempat dengan nilai *coherence* sebesar 0.9407. Analisis model secara bersama-sama didapatkan bahwa pemodelan topik menggunakan NMF lebih baik daripada LDA, sehingga hasil pemodelan NMF dilakukan klasifikasi berdasarkan topik yang didapatkan.

Tabel 4. Klasifikasi Topik Berdasarkan Model Terbaik

No	Komentar	Klasifikasi Topik dengan NMF
1.	bro cowok yang pake baju nike kuning mirip banget sama lebron james	3
2.	ini tidak ada di saluran sejarah kehidupan orang kulit hitam itu penting	2
3.	realitas manusia setelah kematian bersikap baik dan rendah hati terhadap semua orang rasisme membenci penindasan tidak pernah bertahan lama	1
4.	apa yang terjadi pada orang-orang ini dan mengapa mereka masih duduk seperti itu sungguh menyedihkan	5
5.	mereka semua menderita kuru karena memakan budaya mereka yang menjijikkan dan jahat	5
...
...
50285	kali nonton bang bobon ngakak emosi lo gue bahagia sedih haru lo bang moga tuhan memberkatimu moga sehat limpah rezeki amen	1
50286	Papua perhati pemerintah kkb ngk	2
50287	maaf kk sy ijin titip cinta kasih u adik sodara Papua terima kasih layanan kk bobon santoso dgn video moga gerak hati bantu sodara Papua rasa merdeka sungguh adil sosial rakyat indonesia tuhan kati terima kasih	5

Topik yang sudah diklasifikasi kemudian dianalisis sentimennya, sehingga dapat diketahui emosi dalam opini yang tertulis dalam kolom komentar tersebut. Analisis sentimen yang dihasilkan dapat mengetahui topik mana saja yang mengundang prespektif negatif, positif maupun netral.



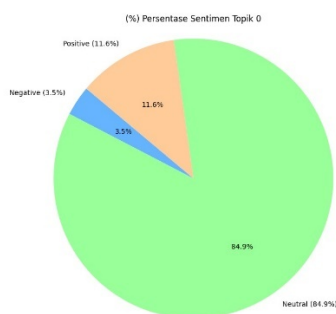
Tabel 6. Klasifikasi Sentimen Topik Berdasarkan Model Terbaik

No	Topik	Klasifikasi Topik dengan NMF	Sentiment
1.	bro cowok yang pake baju nike kuning mirip banget sama lebron james	3	Neutral
2.	ini tidak ada di saluran sejarah kehidupan orang kulit hitam itu penting	2	Neutral
3.	realitas manusia setelah kematian bersikap baik dan rendah hati terhadap semua orang rasisme membenci penindasan tidak pernah bertahan lama	1	Neutral
4.	apa yang terjadi pada orang-orang ini dan mengapa mereka masih duduk seperti itu sungguh menyedihkan	5	Neutral
5.	mereka semua menderita kuru karena memakan budaya mereka yang menjijikkan dan jahat	5	Neutral
...
...
50285	kali nonton bang bobon ngakak emosi lo gue bahagia sedih haru lo bang moga tuhan memberkatimu moga sehat limpah rezeki amen	1	Neutral
50286	Papua perhati perintah kkb ngk	2	Neutral
50287	maaf kk sy ijin titip cinta kasih u adik sodara Papua terima kasih layan kk bobon santoso dgn video moga gerak hati bantu sodara Papua rasa merdeka sungguh adil sosial rakyat indonesia tuhan kati terima kasih	5	Positive

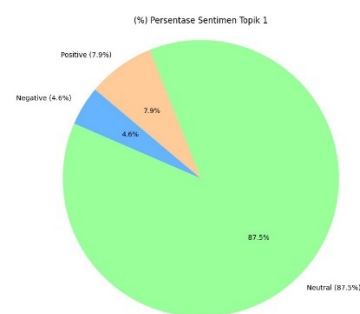
Tabel 5. Analisis Jumlah Sentimen Topik Berdasarkan Model Terbaik

Topic	Negative	Neutral	Positive
1	496	12126	1663
2	1023	19353	1736
3	206	6773	988
4	100	2139	294
5	92	2974	323

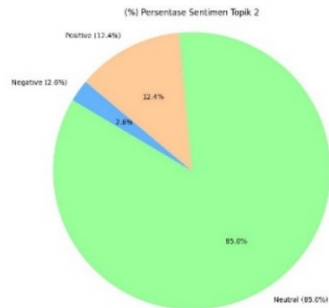
Topik yang digunakan dalam analisis sentimen ini adalah topik yang diperoleh dari metode NMF. Kecenderungan sentimen yang Terdapat dalam kelima topik terpilih adalah netral dengan presentasi lebih dari 80% tiap topiknya.



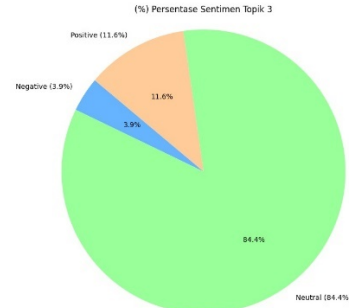
Gambar 4a.



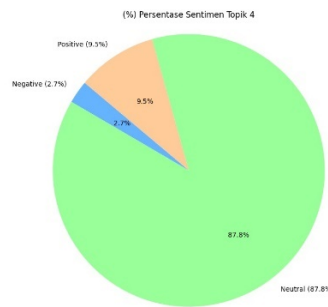
Gambar 4b.



Gambar 4c.



Gambar 4d.



Gambar 4e.

Gambar 4. Visualisasi sentimen tiap topik : analisis sentimen topik ke-1 (4a), analisis sentimen topik ke-2 (4b), analisis sentimen topik ke-3 (4c), analisis sentimen topik ke-4 (4d), analisis sentimen topik ke-5 (4e).

IV. KESIMPULAN

Hasil penelitian ini dapat disimpulkan bahwa bahwa model yang diperoleh dari Metode *Non-Negative Matrix Factorization* (NMF) dan model yang dihasilkan metode *Latent Dirichlet Allocation* (LDA) secara bersama dapat menjelaskan topik yang berada di 142.804 komentar YouTube, dalam rentang waktu 1 Januari 2024 hingga 31 Mei 2024. Model yang diperoleh dari Metode *Non-Negative Matrix Factorization* (NMF) lebih baik daripada model yang dihasilkan metode *Latent Dirichlet Allocation* (LDA). Kesimpulan model terbaik ini diperoleh dari hasil analisis nilai *coherence* NMF sebesar 0.8546 dengan 5 topik, sedangkan nilai *coherence* LDA sebesar 0.7710 dengan 10 topik. Namun, nilai *coherence* dalam topik terbaik model LDA lebih tinggi daripada nilai *coherence* model topik terbaik yang dihasilkan oleh metode NMF. Model topik terbaik metode LDA memiliki nilai *coherence* sebesar 0.9702 sedangkan model topik terbaik metode NMF memiliki nilai *coherence* sebesar 0.9407.

Isi topik pada model terbaik menggunakan metode LDA berbeda dengan isi topik pada model terbaik yang dianalisis menggunakan metode NMF. Model yang dianalisis dengan metode LDA berisi “tidak, di, dan, mereka, yang” dan model topik yang dianalisis dengan metode NMF berisi “matang, misi, kualiti, putih, merah”. Perbedaan bentuk pemodelan ini memungkinkan bahwa hasil analisis sentimen yang dihasilkan juga mengalami perbedaan. Analisis sentimen dengan metode VADER dalam penelitian ini hanya dilakukan terhadap model terbaik yaitu metode NMF. Hasil analisis metode VADER menunjukkan bahwa kelima topik pada metode NMF cenderung memiliki sentimen netral dengan presentasi rata-rata 85,92%.



Saran yang diberikan untuk penelitian serupa adalah perlunya mencari metode yang lebih detail untuk analisis sentimen apabila menggunakan perbandingan dua metode. Metode analisis sentimen yang tepat memungkinkan adanya pembandingan terhadap suatu metode dengan metode analisis topik yang lain.

UCAPAN TERIMA KASIH

Terima kasih kepada dosen pembimbing yang sudah memberikan dukungan serta kontribusi dalam penelitian ini. Ucapan terima kasih penulis berikan atas dukungan dan kerjasama dari berbagai pihak, tim dan organisasi yang terlibat dalam terselesaikannya penelitian ini. Semoga hasil penelitian ini dapat bermanfaat dan menjadi wawasan bagi perkembangan ilmu pengetahuan dan pihak yang terkait dengan bidang statistika, matematika, data science dan sejenisnya.

REFERENSI

1. C. Stella Kakisina, “ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KEMISKINAN DI KABUPATEN JAYAWIJAYA,” *JURNAL EKONOMI & BISNIS*, vol. 7, no. 2, pp. 43–52, Jan. 2016, [Online]. Available: www.bappenas.go.id
2. Nikita Rosa Damayanti Waluyo, “3 Provinsi Baru Papua Disahkan , Ini Sejarah Pemekaran Daerah di Indonesia Sejarah Pemekaran Daerah di Indonesia,” *detikedu, detikpedia*. Accessed: Jun. 17, 2024. [Online]. Available: <https://www.detik.com/edu/detikpedia/d-6156710/3-provinsi-baru-papua-disahkan-ini-sejarah-pemekaran-daerah-di-indonesia>
3. J. Karay, P. Pariwisata, Y. Merauje, and H. H. Pigai, “Relasi Manusia dan Alam di Tanah Papua,” *econusa*. Accessed: Jun. 17, 2024. [Online]. Available: <https://econusa.id/id/ecoblog/relasi-manusia-dan-alam-di-tanah-papua/>
4. U. Pangarti, S. D. Purnomo, D. Retnowati, and B. Adhitya, “Studi komparatif ketimpangan antar pulau di Indonesia Ufit,” *FORUM EKONOMI: Jurnal Ekonomi, Manajemen dan Akuntansi* ISSN: 1411-1713 (Print) 2528-150X (Online), vol. 2, no. 2, pp. 288–298, 2022, doi: 10.29264/jfor.v24i2.10988.
5. B. Anugerah, “Papua : Mengurai Konflik dan Merumuskan Solusi,” no. 23, pp. 51–65, 2024.
6. M. Farid, “Dampak Negatif Dan Positif Budaya Osuleng Bagi Masyarakat Banggai Dalam Upacara Adat Pernikahan Khususnya Di Desa Oluno,” 2014. [Online]. Available: <https://api.semanticscholar.org/CorpusID:194225401>
7. H. Sama, A. Angeline, G. Phua, R. Rahel, and S. Stephanie, “Dampak Positif dan Negatif Etika Informasi dalam Media Sosial: Sebuah Penelitian Kualitatif dengan Metode Studi Literatur,” *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, 2023, [Online]. Available: <https://api.semanticscholar.org/CorpusID:259878911>
8. Ruslan, “Penelitian Ini Membahas Tentang Dampak Media Sosial Dalam Penyebaran Informasi Positif Dan Negatif Di Kecamatan Teunom, Kabupaten Aceh Jaya,” 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:214139433>
9. K.-I. Kwon and S. Shin, “Analysis of YouTube comments related to ‘self-harm’ using topic modeling,” *Korean Association For Learner-Centered Curriculum And Instruction*, 2023, [Online]. Available: <https://api.semanticscholar.org/CorpusID:263210016>
10. D. Wiryany and T. V. Pratami, “KEKUATAN MEDIA BARU YOUTUBE DALAM MEMBENTUK BUDAYA POPULER,” vol. 02, no. 02, pp. 25–30, 2019.
11. M. H. Joviansyah, L. N. Alyssa, I. S. Rahadatul’aisyi, G. N. Bakry, and N. Aristi, “ANALISIS SENTIMEN DAN JARINGAN KOMENTAR VIDEO YOUTUBE NAJWA SHIHAB ‘PIALA DUNIA U-20 GAGAL DIGELAR DI INDONESIA. MARI LIHAT DARI DUA PERSPEKTIF,’” *Jurnal Komunikasi dan Media*, 2023, [Online]. Available: <https://api.semanticscholar.org/CorpusID:266077302>



12. M. S. T. Allo, R. R. Onsu, and L. E. Tulung, “Pengaruh Konten Youtube Cretivox Friends with Benefit Terhadap Perilaku Negatif Universitas Sam Ratulangi,” *JURNAL ADMINISTRASI BISNIS*, 2023, [Online]. Available: <https://api.semanticscholar.org/CorpusID:260847151>
13. D. Mualfah et al., “Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM,” *JURNAL FASILKOM*, 2023, [Online]. Available: <https://api.semanticscholar.org/CorpusID:260038340>
14. J. Munandar and M. Al Haris, “Analisis Topic Modeling pada Tweet Pengguna Twitter tentang Calon Presiden Indonesia 2024 Topic Modeling Analysis on Twitter User Tweet Regarding the 2024 Indonesia Presidential Candidates,” *JURNAL SAINS DAN INFORMATIKA Research of Science and Informatic V4.I1*, no. April 2023, 2024.
15. W. A. Anggraeni, F. F. Roji, and M. Alkautsar, “Analisis Sentimen Publik terhadap Kebijakan Insentif Perpajakan Dengan Pendekatan VADER (Valence Aware Dictionary And Sentiment Reasoner),” *Jurnal Proaksi*, 2023, [Online]. Available: <https://api.semanticscholar.org/CorpusID:265093431>
16. K. E. Rofran and F. J. Kaunang, “Analisis Sentimen Pengguna Instagram terhadap Kebijakan Nadiem Makarim yang Memperbolehkan Mahasiswa Lulus Tanpa Skripsi menggunakan Metode Analisis VADER dan Metode Klasifikasi Naïve Bayes,” *INTECOMS: Journal of Information Technology and Computer Science*, 2024, [Online]. Available: <https://api.semanticscholar.org/CorpusID:270308753>
17. V. Nurcahyawati and Z. Mustaffa, “Vader Lexicon and Support Vector Machine Algorithm to Detect Customer Sentiment Orientation,” *Journal of Information Systems Engineering and Business Intelligence*, vol. 9, no. 1, pp. 108–118, Apr. 2023, doi: 10.20473/jisebi.9.1.108-118.
18. S. Biswas, K. Young, and J. Griffith, “A Comparison of Automatic Labelling Approaches for Sentiment Analysis,” Nov. 2022, doi: 10.5220/0011265900003269.
19. A. Meddeb and L. Ben Romdhane, “Using Topic Modeling and Word Embedding for Topic Extraction in Twitter,” *Procedia Comput Sci*, vol. 207, pp. 790–799, 2022, doi: <https://doi.org/10.1016/j.procs.2022.09.134>.
20. S. Qomariyah, N. Iriawan, and K. Fithriasari, “Topic modeling Twitter data using Latent Dirichlet Allocation and Latent Semantic Analysis,” *AIP Conf Proc*, vol. 2194, no. 1, p. 020093, Dec. 2019, doi: 10.1063/1.5139825.
21. Y. Chen, J.-B. Bordes, and D. Filliat, “An experimental comparison between NMF and LDA for active cross-situational object-word learning,” Jun. 2016. doi: 10.1109/DEVLRN.2016.7846822.
22. M. Z. Ghaly and A. D. Laksito, “Topic Modeling of Natural Disaster in Indonesia Using NMF,” 2023 Eighth International Conference on Informatics and Computing (ICIC), pp. 1–6, 2023, [Online]. Available: <https://api.semanticscholar.org/CorpusID:266905580>
23. J. Informasi, R. Perucha, and F. Afidh, “Pemodelan Topik Menggunakan n-Gram dan Non-negative Matrix Factorization,” *Jurnal Informasi dan Teknologi*, vol. 5, no. 1, pp. 2–9, 2023, doi: 10.60083/jidt.v5i1.385.
24. R. Egger and J. Yu, “A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts,” *Frontiers in Sociology*, vol. 7, pp. 1–16, May 2022, doi: 10.3389/fsoc.2022.886498.
25. I. Vayansky and S. Kumar, “A review of topic modeling methods,” *Inf Syst*, vol. 94, p. 101582, Jun. 2020, doi: 10.1016/j.is.2020.101582.
26. N. Suaysom and W. Gu, “Expert Opinion and Coherence Based Topic Modeling,” *International Journal on Natural Language Computing*, vol. 7, no. 2, pp. 01–14, Apr. 2018, doi: 10.5121/ijnlc.2018.7201.
27. J. Blad and K. Svensson, “Exploring NMF and LDA Topic Models of Swedish News Articles,” 2020. [Online]. Available: <http://www.teknat.uu.se/student>
28. M. H. A. H. A. K. J. S. M. A. A. A. Badriya Murdhi Alenzi Muhammad Badruddin Khan, “Automatic Annotation Performance of TextBlob and VADER on Covid Vaccination Dataset,” *Intelligent Automation & Soft Computing*, vol. 34, no. 2, pp. 1311–1331, 2022, doi: 10.32604/iasc.2022.025861.



SENADA
Seminar Nasional Sains Data

Seminar Nasional Sains Data 2024 (SENADA 2024)

UPN “Veteran” Jawa Timur

E-ISSN 2808-5841

P-ISSN 2808-7283

29. M. K. B. Laxmi, “TWITTER SENTIMENT ANALYSIS USING VADER ON PYTHON,” JETIR, 2020.
[Online]. Available: www.jetir.org