

# Cek Plagiasi Ulfah

*by* Alfian Indra

---

**Submission date:** 11-Mar-2023 10:53AM (UTC+0900)

**Submission ID:** 2014633042

**File name:** jurnal\_ulfah\_cekplagiasi.pdf (880.85K)

**Word count:** 7059

**Character count:** 43666

**TOPIC MODELING IN COVID-19 VACCINATION REFUSAL CASES USING LATENT DIRICHLET ALLOCATION AND LATENT SEMANTIC ANALYSIS**Ulfah Malihatn Sholihah<sup>1</sup>, Yulian Findawati<sup>2</sup>, Uce Indahyanti<sup>3</sup><sup>1,2,3</sup>Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sidoarjo  
Email: <sup>1</sup>191080200172@umsida.ac.id, <sup>2</sup>yulianfindawati@umsida.ac.id, <sup>3</sup>uceindahyanti@umsida.ac.id

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

**Abstract**

COVID -19 vaccination is a program provided by the Indonesian government to minimize the spread of the virus. The Covid-19 vaccination program in Indonesia runs alongside circulating issues, causing pros and cons. Some people reject the COVID -19 vaccination and write their opinions on social media such as Twitter. Therefore, this study uses the Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA) methods to summarize frequently discussed topics and hidden topics in 1797 Twitter scraping data. Both models require a set of words that have been converted into a matrix, so before conducting LDA topic modeling, the dataset will undergo a bag of word (BOW) calculation. Meanwhile, in LSA topic modeling, the existing dataset will undergo word weighting of frequently occurring words using Term Frequency - Inverse Document Frequency (TF-IDF). This study was conducted to find and summarize hidden information in the form of frequently discussed topics, thus understanding public opinions related to the COVID -19 vaccination refusal case. LDA and LSA methods will display topics based on the probability and mathematical calculations of word occurrences in each topic in the document. The topics that appear will be further analyzed through coherence score by applying a limit of 20 topics to display the best value. Further modeling experiments are carried out to display topics through LDA and LSA models again. Based on the evaluated results of the models, this study takes 4 topics with the highest coherence value to be interpreted into both topic modeling methods.

**Keywords:** Covid-19 vaccination Refusal, Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), Topic Modelling, Twitter.

**PEMODELAN TOPIK PADA KASUS TOLAK VAKSINASI COVID-19 MENGGUNAKAN LATENT DIRICHLET ALLOCATION DAN LATENT SEMANTIC ANALYSIS****Abstrak**

Vaksinasi COVID -19 adalah program yang diberikan pemerintah Indonesia dalam upaya meminimalisir terjadinya penularan virus tersebut. Program vaksinasi COVID -19 di Indonesia berjalan beriringan dengan isu – isu yang beredar sehingga menimbulkan pro dan kontra. Sejumlah masyarakat menolak adanya vaksinasi COVID -19, beberapa menulis opini tersebut di media sosial seperti Twitter. Oleh karena itu, untuk meringkas topik yang sering dibahas dan topik – topik yang tersembunyi pada 1797 data hasil *scrapping* Twitter, pada penelitian ini menggunakan metode *Latent Dirichlet Allocation (LDA)* dan *Latent Semantic Analysis (LSA)*. Kedua model tersebut membutuhkan sekumpulan kata – kata yang telah dirubah ke dalam suatu matriks, sehingga sebelum melakukan pemodelan topik metode LDA, dataset tersebut akan dilakukan perhitungan *bag of word (BOW)*. Sedangkan pada pemodelan topik LSA, dataset yang ada akan dilakukan pembobotan kata – kata yang sering muncul menggunakan *Term Frequency – Inverse Document Frequency (TF-IDF)*. Penelitian ini dilakukan untuk menemukan dan meringkas informasi tersembunyi berupa topik – topik yang sering dibahas sehingga dapat mengetahui opini – opini masyarakat terkait kasus penolakan vaksinasi COVID – 19. Metode LDA dan LSA akan menampilkan topik – topik berdasarkan hasil dari perhitungan probabilitas dan matematis kemunculan kata pada setiap topik dalam dokumen. Topik yang muncul akan dianalisa lagi melalui *coherence score* dengan menerapkan batas topik yang akan ditampilkan sebanyak 20 topik nilai terbaik. Percobaan pemodelan selanjutnya dilakukan untuk menampilkan topik melalui model LDA dan LSA lagi. Berdasarkan hasil evaluasi model yang telah divisualisasikan, pada penelitian ini diambil 4 jumlah topik dengan nilai koherensi tertinggi yang akan diinterpretasikan ke dalam kedua metode pemodelan topik tersebut.

2  
**Kata kunci:** *Latent Dirichlet Allocation (LDA)*, *Latent Semantic Analysis (LSA)*, *Pemodelan Topik*, *Tolak Vaksin Covid-19*, *Twitter*.

## 1. PENDAHULUAN

COVID-19 adalah virus yang ditemukan pada tahun 2019 di China yang menyerang saluran pernapasan. Virus ini menimbulkan pandemi di seluruh dunia karena proses penularan sangat cepat dan belum ada obat untuk mencegah. Akibatnya semua kegiatan harus diberlakukan protokol kesehatan, beberapa sektor pekerjaan ikut terdampak, pemutusan hubungan kerja di sejumlah tempat dilakukan secara besar – besaran, pengalihan pembelajaran ke media dalam jaringan (daring), hingga *lockdown*.

Banyak penelitian yang dilakukan oleh badan kesehatan hamper di seluruh dunia dalam upaya menciptakan vaksin COVID-19, sehingga pada tahun 2021 vaksin tersebut pertama kali ditetapkan pemerintah sebagai upaya untuk menekan kenaikan angka dan mencegah mengurangi penyebaran COVID-19. Vaksinasi dilakukan secara masal di berbagai daerah di Indonesia secara bertahap karena pelayanan kesehatan seperti ini merupakan bentuk dari pemenuhan hak dan kewajiban masyarakat Indonesia. Terdapat peraturan vaksin COVID -19 dalam Perpres No. 99 Tahun 2020 tentang Pengadaan Vaksin dan Pelaksanaan Vaksinasi Dalam Rangka Penanggulangan Pandemi COVID-19, Keputusan Menteri Kesehatan No. HK.01.02./MENKES/12758/2021 tentang Penetapan Jenis Vaksin untuk Pelaksanaan Vaksinasi COVID-19, Permenkes No 84 Tahun 2020 tentang Pelaksanaan Vaksinasi Dalam Rangka Penanggulangan Pandemi COVID19 [1].

Namun, vaksinasi yang diselenggarakan pemerintah menimbulkan pro dan kontra. Terdapat isu – isu tentang vaksin COVID-19 yang ramai dibicarakan melalui media sosial seperti Twitter sehingga muncul kasus penolakan vaksin COVID-19. Banyak masyarakat menuliskan tanggapan di Twitter dan menjadi *trending topic* di Indonesia. Hal ini ditandai dengan munculnya sebuah *hashtag* #TolakDiVaksinSinovac yang menjadi *trending* Twitter pada tanggal 12 Januari 2021 [2]. Twitter menjadi media sosial paling mudah untuk memaparkan pendapat atau opini – opini penggunaannya, sehingga hal seperti ini dapat membantu dalam mengambil suatu informasi, data penelitian, jajak pendapat, analisa sentimen, dan kepentingan lainnya.

Pada penelitian ini, pendapat masyarakat Indonesia yang menolak vaksinasi COVID-19 di Twitter akan dilakukan klasifikasi untuk melihat topik yang sering dibicarakan. Pengambilan data untuk diringkan tersebut menggunakan metode *Latent Dirichlet Allocation (LDA)* dan *Latent Semantic Analysis (LSA)*. Pendekatan LDA menghasilkan kata-kata yang sering muncul

bersamaan dan juga mengaturnya ke dalam topik yang berbeda. Untuk menghitung jumlah topik yang paling tepat berdasarkan model *coherence-gensim*[3] pada 21 batasan perhitungan model topik dengan nilai tertinggi yang dibentuk [4] setelah melakukan percobaan pemodelan pertama.

Sejumlah topik dari dataset setelah melalui *preprocessing* dan *cleaning data* selanjutnya data akan dilakukan pembentukan matriks dari sekumpulan korpus yaitu *bag of word (BOW)* pemilihan fitur ini membuat pemodelan lebih efektif dan relevan. Hasil dari evaluasi *topic coherence* akan diproses menggunakan model LDA dan divisualisasikan topik terbaik melalui modul *pyLDAvis*.

Untuk mengetahui topik tersembunyi atau pandangan lainnya dan polaritas masyarakat Indonesia terhadap isu penolakan vaksinasi COVID-19, pada penelitian ini dapat dilakukan metode pemodelan topik yang berbeda. Hal ini penting agar opini publik terhadap kasus penolakan vaksin dapat dengan cepat dianalisa sehingga dapat *dicounter* oleh pelaku masa depan. Model LSA memberikan cara berbeda dalam pemrosesan matriks setiap kata dalam korpus. Pada LSA sebelum menguji model tersebut terdapat proses perhitungan *Document Term Matrix (DTM)* yang bisa diberikan pembobotan kata melalui *Term Frequency - Inverse Document Frequency (TF - IDF)*. Bobot kata yang dihasilkan akan dijadikan sebagai percobaan pemodelan topik pertama pada metode LSA sebelum evaluasi model menggunakan *coherence score*.

## 2. METODE PENELITIAN



Gambar 1. Flowchart Penelitian

Terdapat beberapa tahapan yang digunakan dalam penelitian ini. Adapun tahapan – tahapan tersebut adalah sebagai berikut:

### 2.1. Pengambilan Dataset

Data yang akan diolah merupakan data hasil *scraping* pada media sosial Twitter menggunakan *library* Python yaitu *Tweepy*. Dataset yang didapat disimpan ke dalam bentuk berkas yang nerekstensi *.csv*.

### 2.2. Pra-pemrosesan Teks (Text Peprocessing)

Dataset terlebih dahulu dilakukan pembersihan dari beberapa prosedur. Tahap ini dapat dilakukan untuk membuat semua teks menjadi teks yang berhuruf kecil (*lowercase folding*). Pada dasarnya proses *tokenizing* ialah memisahkan setiap kata yang menyusun suatu dokumen[5]. Menghilangkan karakter – karakter yang ada di dalamnya seperti emoji, tanda baca, link, tagar, maupun URL melalui tahapan *tokenizing*. Kemudian, teks yang telah bersih akan dilakukan penghapusan kata – kata yang tidak memiliki pengaruh atau makna, tahap ini dinamakan *stopwords*. *Stopwords removal* dapat meningkatkan rasio *signal-to-noise* dalam teks yang tidak terstruktur dan dengan demikian meningkatkan signifikansi statistik dari istilah yang bisa jadi penting[6]. Pada dataset penelitian ini, terdapat kata – kata berulang yang harus dihapuskan untuk mendapatkan hasil terbaik saat melakukan pemodelan (*normalization*) dan menghilangkan kata dengan imbuhan berupa prefiks atau sufiks melalui tahapan *stemming*.

### 2.3. Pemrosesan Teks (Text Processing)

Penulis akan melakukan perhitungan kata – kata dari dataset yang telah bersih menjadi sebuah matriks BOW sebelum dijadikan ke pemodelan topik. Metode ini mengubah data tabular (korpus) menjadi document term yang merupakan data numerik. BOW hanya menghitung kata unik, jika terdapat kata duplikat atau kata yang sama antar topik akan ditulis satu kali saja.

Pada pemodelan topik LSA, terdapat korpus berupa sekumpulan kata yang harus dirubah juga menjadi sebuah vektor. TF – IDF merupakan pembobotan kata yang digunakan sebelum melakukan pemodelan topik LSA. Pada persamaan (1) menjelaskan bahwa TF menghitung jumlah seberapa sering kemunculan kata dalam sebuah dokumen dibanding dengan jumlah kata pada dokumen.

$$tf_{td} = \frac{n_{td}}{\text{Total number of term document}} \quad (1)$$

Persamaan (2) IDF menilai kata yang *relative* sering muncul dianggap sebagai kata penting atau bukan berdasarkan kemunculan kata pada setiap korpus. IDF menghitung level korpus secara keseluruhan bukan perdokumen individual.

$$IDF_d = \log\left(\frac{\text{Number of document}}{\text{Number of doc with term } t'}\right) \quad (2)$$

2

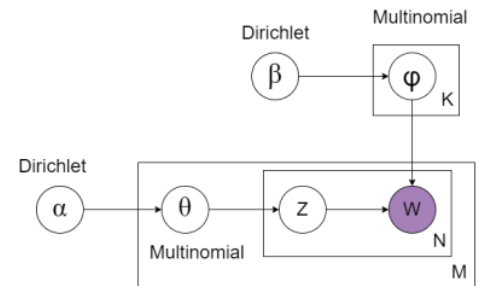
#### 2.3.1. Latent Dirichlet Allocation (LDA)

*Latent Dirichlet Allocation* (LDA) merupakan model probabilitas generatif yang mengasumsikan bahwa setiap topik adalah campuran dari

sekumpulan kata potensial atau biasa disebut token, dan setiap dokumen (korpus) adalah campuran dari sekumpulan topik probabilitas yang disebut latent. Fungsi dari LDA bisa digunakan untuk melakukan klasterisasi, memroses data yang besar karena daftar topik yang dihasilkan LDA memiliki pembobotan.

Algoritma LDA bekerja dengan melakukan inisialisasi parameter berupa:

- Jumlah dokumen (M)
- Jumlah topik (K) yang akan ditampilkan
- Jumlah iterasi (i)
- Jumlah kata dalam dokumen (N)
- Koefisien LDA ( $\alpha, \beta$ )



Gambar 2. Alur Kerja LDA

LDA melabeli masing – masing kata dengan topik yang telah ditentukan dalam suatu dokumen secara acak (*random topic assignment*) melalui *random distribution*. Pada proses iterasi *resample each word* akan menghasilkan parameter yang dapat menentukan distribusi dari jumlah topik dalam suatu dokumen. *Resample* dilakukan berdasarkan kelaziman suatu kata dalam suatu topik, dan mengevaluasi kelaziman topik berada dalam sebuah dokumen.

Distribusi *dirichlet*  $\alpha$  merupakan parameter yang mengontrol distribusi topik sebuah dokumen, semakin tinggi nilai alfa maka dinyatakan bahwa dokumen tersebut memiliki beragam topik. Hal ini dapat direpresentasikan melalui distribusi *multinomial* ( $\theta$ ), di mana setiap distribusi acak dapat memuat topik ( $Z$ ) dan kata yang spesifik ( $W$ ) dalam dokumen sehingga akan diperoleh banyaknya dokumen ( $M$ ) dan memuat beberapa jumlah kata dalam setiap dokumen ( $N$ ).

Sebaliknya, jika nilai  $\alpha$  terlalu rendah, maka tidak ada banyak sebaran topik yang tercampur dalam satu dokumen sehingga tidak terdistribusi secara merata. Distribusi *dirichlet*  $\beta$  merupakan parameter yang mengontrol persebaran kata setiap topik ( $\phi$ ). Nilai  $\beta$  yang tinggi dari suatu topik memuat persebaran kata yang ada di topik lain. Sedangkan nilai  $\beta$  cenderung rendah menunjukkan kata – kata dalam satu topik hanya terdistribusi pada topik tertentu ( $K$ ) sehingga lebih spesifik. Pada model LDA variabel yang diobservasi adalah  $W$ .

### 2.3.2. Latent Semantic Analysis (LSA)

LSA adalah metode natural language yang menganalisis hubungan antara sekumpulan dokumen, kata – kata, dan istilah yang dikandungnya menggunakan distribusi *univariat* (dekomposisi nilai tunggal), teknik matematika yang mencari data tidak terstruktur untuk menemukan hubungan tersembunyi antara istilah dan konsep. Istilah atau konsep kata yang terkandung dalam tulisan akan menjadi acuan pembandingan tanpa melihat ciri kebahasaan suatu tulisan[7]. Refleksi LSA dari pengetahuan manusia telah ditunjukkan dalam berbagai cara diantaranya skor tumpang tindih dengan kosakata. LSA meniru urutan kata manusia dan penilaian kategori, menyimulasikan data priming leksikal.

$$A_{txd} = U_{txn} \times S_{txn} \times V_{dxn}^T \quad (3)$$

Masukan untuk metode ini adalah sekumpulan teks atau korpus yang direpresentasikan ke dalam matriks yang dianggap mampu mengambil inti sari dari suatu dokumen setelah dibandingkan per-kata uniknya (*term*). LSA dapat digunakan untuk memberikan nilai padateksi dengan mengkonversi teks menjadi matriks - matriks yang diberi nilai pada masing - masing *term* untuk dicari kesamaan dengan *term referensi* [8]. Pada Persamaan 3 LSA menggunakan *Singular Vector Decomposition* (SVD) untuk menghitung matriks tersebut, matriks dibuat menjadi 3 komponen yaitu S, U, dan V untuk dilakukan dekomposisi sehingga mampu menghasilkan matriks yang baru. Dokumen yang direpresentasikan dalam matriks jika mengandung kata – kata yang mirip, maka vektor akan menampilkan kata serupa. LSA didasarkan pada *term frequency*, keunggulannya adalah menghasilkan hasil intuitif [9]. Hal ini diterapkan ke dalam tahapan LSA pada saat menghimpun kata – kata menjadi sebuah vektor yang direpresentasikan ke dalam TF IDF. Hal ini melibatkan pembobotan kata dan pencarian kata penting berdasarkan nilai relatif kemunculan kata dari setiap dokumen.

### 2.3.3. Coherence Score

Coherence score atau topik koherensi adalah ukuran untuk mengevaluasi pemodelan topik. Model yang baik menghasilkan topik dengan nilai koherensi yang tinggi [10], maka langkah pengujian dilakukan dengan iterasi sebanyak topik n dan dimulai dari topik i. . Topik koherensi menggunakan statistik dan probabilitas yang diambil dari korpus referensi (kumpulan teks), berfokus pada konteks kata, untuk memberikan skor koherensi pada suatu topik.

Topik koherensi telah diusulkan sebagai metode evaluasi intrinsik untuk pemodelan topik, dinyatakan sebagai rata – rata atau median dari word similarity yang berpasangan dan dibentuk oleh kata

– kata paling atas dengan persamaan topik. Semakin tinggi nilai koherensi maka semakin bagus hasil interpretasi pemodelan topik [11].

## 3. HASIL DAN PEMBAHASAN

### 3.1. Pengambilan Dataset

Pengambilan data diambil pada rentang tanggal 7 Januari 2021 sampai 13 Januari 2021 dan memperoleh 1797 data dari Twitter. Pengambilan data ini dianalisa menggunakan *library* Pandas sebelum disimpan menjadi bentuk csv. Dataset awal seperti pada Tabel 1 masih berupa data kotor yang belum dilakukan pembersihan isi – isi yang dianggap tidak berpengaruh pada proses pemodelan nantinya.

Tabel 1. Dataset Tolak Vaksinasi Covid-19

Date	Tweet
2021-01-13	@Laudate_Dom @Pipelawer165 @dayatia Beliau dan dr. Agni serta semua yang ada di HEAL tidak menolak vaksin Mbak. Kami tidak menolak apapun anjuran pemerintah. Kami juga kampanye 3M kok. Yang kami keitisi adalah hal substansial soal pcr dan metode test lainnya. Plus edukasi how to...
2021-01-13	Orang bule, vaccine = Antichrist Juga dianggap sebagai sarana kontrol populasi penduduk di dunia.. Jadi wajar kalau ada yang menolak program vaksinasi dan anti vaksin🙄 <a href="https://t.co/GwNvveV9TU">https://t.co/GwNvveV9TU</a>
2021-01-07	Pemerintah Provinsi DKI Jakarta akan memberikan sanksi berupa denda bagi warga yang menolak vaksinasi Covid-19. #vaksin <a href="https://t.co/ZqPsi4cvxx">https://t.co/ZqPsi4cvxx</a>
2021-01-08	vaksin belum lulus uji BP POM dan ternyata ga semua orang bisa divaksin..anehnya udah tuh vaksin distribusikan ke daerah, dan menolak divaksin dikenakan denda..Gagal Paham 🙄🙄🙄🙄🙄 <a href="https://t.co/JZL3oYy29f">https://t.co/JZL3oYy29f</a>
2021-01-10	Terus liat video nakes menolak vaksin. Itu pasti video buat seru-seruan kan? Iya kan dok? Sus?

### 3.2. Text Preprocessing

Dataset tidak langsung digunakan oleh sistem. Oleh karena itu, beberapa preprocessing harus dilakukan untuk sedikit memodifikasi data guna meningkatkan kualitas data yang digunakan.

#### 3.2.1. Lowercase Folding

Data awal masih berupa teks yang belum dilakukan pemrosesan apapun, sehingga bentuk dan strukturnya masih original. Untuk mengubah semua huruf pada dokumen atau data tersebut menjadi huruf kecil maka dapat memanfaatkan fungsi Python yaitu `str.lower()`. Proses ini tidak hanya mengubah menjadi huruf kecil semua akan tetapi juga menghilangkan karakter selain a sampai z atau

delimiter. Untuk hasil *lowercase folding* dapat dilihat pada Tabel 2.

Tabel 2. Perbandingan Sebelum dan Sesudah Lowercase Folding

Sebelum lowercase folding	Sesudah lowercase folding
@Laudate Dom @Pipelpower165 @dayatia Beliau dan dr. Agni serta semua yang ada di HEAL tidak menolak vaksin Mbak. Kami tidak menolak apapun anjuran pemerintah. Kami juga kampanye 3M kok. Yang kami keitisi adalah hal substansial soal per dan metode test lainnya. Plus edukasi how to...	@laudate_dom @pipelpower165 @dayatia beliau dan dr. agni serta semua yang ada di heal tidak menolak vaksin mbak. kami tidak menolak apapun anjuran pemerintah. kami juga kampanye 3m kok. yang kami keitisi adalah hal substansial soal per dan metode test lainnya. plus edukasi how to...
Orang bule, vaccine = Antichrist Juga dianggap sebagai sarana kontrol populasi penduduk di dunia.. Jadi wajar kalau ada yang menolak program vaksinasi dan anti vaksin🙏 <a href="https://t.co/GwNvveV9TU">https://t.co/GwNvveV9TU</a>	orang bule, vaccine = antichrist juga dianggap sebagai sarana kontrol populasi penduduk di dunia.. jadi wajar kalau ada yang menolak program vaksinasi dan anti vaksin.🙏 <a href="https://t.co/gwnvvev9tu">https://t.co/gwnvvev9tu</a>
1 Pemerintah Provinsi DKI Jakarta akan memberikan sanksi berupa denda bagi warga yang menolak vaksinasi Covid-19. #vaksin <a href="https://t.co/ZqPsi4cvxx">https://t.co/ZqPsi4cvxx</a>	1 pemerintah provinsi dki jakarta akan memberikan sanksi berupa denda bagi warga yang menolak vaksinasi covid-19. #vaksin <a href="https://t.co/zqpsi4cvxx">https://t.co/zqpsi4cvxx</a>
vaksin belum lulus uji BP POM dan ternyata ga semua orang bisa divaksin..anehnya udah tuh vaksin distribusikan ke daerah, dan menolak divaksin dikenakan denda..Gagal Paham 🙏🙏🙏 <a href="https://t.co/JZL3oYy29f">https://t.co/JZL3oYy29f</a>	vaksin belum lulus uji bp pom dan ternyata ga semua orang bisa divaksin..anehnya udah tuh vaksin distribusikan ke daerah, dan menolak divaksin dikenakan denda..gagal paham 🙏🙏🙏 <a href="https://t.co/jzl3oyy29f">https://t.co/jzl3oyy29f</a>
Terus liat video nakes menolak vaksin. Itu pasti video buat seru-seruan kan? Iya kan dok? Sus?	@ridwanhr ntar katanya bisa dipidanakan kalau menolak vaksin

### 3.2.2. Tokenizing

Data hasil case folding selanjutnya akan dipisah menjadi potongan – potongan karakter seperti kata, tanda baca, angka, dan simbol sebagai token. Untuk memisah kalimat dalam sebuah dokumen menjadi kata per kata dapat menggunakan kelas `word_tokenize()` dengan mengimpor modul `nlk.tokenize` dan `nlk.probability` terlebih dahulu. Pada Tabel 3 kalimat *tweet* yang ada terdapat emoji, link, tagar, URL tidak sempurna, garis spasi, dan tab akan dihilangkan juga. Proses menghilangkan tanda baca melalui *remove\_punctuation sehingga* data yang ada hanya berupa potongan – potongan kalimat berbentuk kata – kata saja.

Tabel 3. Perbandingan Sebelum dan Sesudah Tokenizing

Sebelum Tokenizing	Sesudah Tokenizing
--------------------	--------------------

@laudate_dom @pipelpower165 @dayatia beliau dan dr. agni serta semua yang ada di heal tidak menolak vaksin mbak. kami tidak menolak apapun anjuran pemerintah. kami juga kampanye 3m kok. yang kami keitisi adalah hal substansial soal per dan metode test lainnya. plus edukasi how to...	['dom', 'beliau', 'dan', 'dr', 'agni', 'serta', 'semua', 'yang', 'ada', 'di', 'heal', 'tidak', 'menolak', 'vaksin', 'mbak', 'kami', 'tidak', 'menolak', 'apapun', 'anjuran', 'pemerintah', 'kami', 'juga', 'kampanye', 'kok', 'yang', 'kami', 'keitisi', 'adalah', 'hal', 'substansial', 'soal', 'per', 'dan', 'metode', 'test', 'lainnya', 'plus', 'edukasi', 'how', 'to']
orang bule, vaccine = antichrist juga dianggap sebagai sarana kontrol populasi penduduk di dunia.. jadi wajar kalau ada yang menolak program vaksinasi dan anti vaksin. 🙏 <a href="https://t.co/gwnvvev9tu">https://t.co/gwnvvev9tu</a>	['orang', 'bule', 'vaccine', 'antichrist', 'juga', 'dianggap', 'sebagai', 'sarana', 'kontrol', 'populasi', 'penduduk', 'di', 'dunia', 'jadi', 'wajar', 'kalau', 'ada', 'yang', 'menolak', 'program', 'vaksinasi', 'dan', 'anti', 'vaksin']
1 pemerintah provinsi dki jakarta akan memberikan sanksi berupa denda bagi warga yang menolak vaksinasi covid-19. #vaksin <a href="https://t.co/zqpsi4cvxx">https://t.co/zqpsi4cvxx</a>	1 ['pemerintah', 'provinsi', 'dki', 'jakarta', 'akan', 'memberikan', 'sanksi', 'berupa', 'denda', 'bagi', 'warga', 'yang', 'menolak', 'vaksinasi', 'covid', '19', '#vaksin']
vaksin belum lulus uji bp pom dan ternyata ga semua orang bisa divaksin..anehnya udah tuh vaksin distribusikan ke daerah, dan menolak divaksin dikenakan denda.gagal paham 🙏🙏🙏 <a href="https://t.co/jzl3oyy29f">https://t.co/jzl3oyy29f</a>	['vaksin', 'belum', 'lulus', 'uji', 'bp', 'pom', 'dan', 'ternyata', 'ga', 'semua', 'orang', 'bisa', 'divaksin', 'anehnya', 'udah', 'tuh', 'vaksin', 'distribusikan', 'ke', 'daerah', 'dan', 'menolak', 'divaksin', 'dikenakan', 'denda', 'gagal', 'paham']
@ridwanhr ntar katanya bisa dipidanakan kalau menolak vaksin	['terus', 'liat', 'video', 'nakes', 'menolak', 'vaksin', 'itu', 'pasti', 'video', 'buat', 'seruseruan', 'kan', 'iya', 'kan', 'dok', 'sus']

### 3.2.3 Stopwords

Pada *text mining*, susuan kata pada suatu kalimat dianggap kurang bagus untuk pemodelan topik dan analisa sentimen karena dapat memengaruhi keakuratan hasil analisis. Oleh karena itu, upaya penghapusan kata – kata yang tidak memiliki makna namun muncul dengan frekuensi cukup banyak dilakukan menggunakan `nlk.corpus`. Contoh *stopwords* dalam bahasa Indonesia menggunakan modul Sastrawi adalah “dan”, “pada”, “pd”, “begitu”, “gini”, “gitu”, “yang”, “adalah”, “jika”, “maka”, dan lain sebagainya. Perbedaan data sebelumnya dengan data yang telah diproses ke *stopwords* dapat dilihat pada Tabel 4.

Tabel 4. Perbandingan Sebelum dan Sesudah Stopwords

Sebelum Stopwords	Sesudah Stopwords
['dom', 'beliau', 'dan', 'dr', 'agni', 'serta', 'semua', 'yang', 'ada', 'di', 'heal', 'tidak', 'menolak', 'vaksin', 'mbak', 'kami', 'tidak', 'menolak', 'apapun', 'anjuran', 'pemerintah', 'kami', 'juga', 'kampanye', 'kok', 'yang', 'kami', 'keitisi', 'adalah', 'hal', 'substansial', 'soal', 'per', 'dan', 'metode', 'test', 'lainnya', 'plus', 'edukasi', 'how', 'to']	['dom', 'beliau', 'dr', 'agni', 'heal', 'menolak', 'vaksin', 'mbak', 'menolak', 'apapun', 'anjuran', 'pemerintah', 'kami', 'juga', 'kampanye', 'keitisi', 'adalah', 'hal', 'substansial', 'soal', 'per', 'dan', 'metode', 'test', 'lainnya', 'plus', 'edukasi', 'how', 'to']

'apun', 'anjuran', 'substansial', 'pcr', 'metode', 'pemerintah', 'kami', 'juga', 'kampanye', 'kok', 'yang', 'kami', 'keitis', 'adalah', 'hal', 'substansial', 'soal', 'pcr', 'dan', 'metode', 'test', 'lainnya', 'plus', 'edukasi', 'how', 'to']	
['orang', 'bule', 'vaccine', 'antichrist', 'juga', 'dianggap', 'sebagai', 'sarana', 'kontrol', 'populasi', 'penduduk', 'di', 'dunia', 'jadi', 'wajar', 'kalau', 'ada', 'yang', 'menolak', 'program', 'vaksinasi', 'dan', 'anti', 'vaksin']	['orang', 'bule', 'vaccine', 'antichrist', 'dianggap', 'sarana', 'kontrol', 'populasi', 'penduduk', 'dunia', 'wajar', 'menolak', 'program', 'vaksinasi', 'anti', 'vaksin']
1 ['pemerintah', 'provinsi', 'dki', 'jakarta', 'akan', 'memberikan', 'sanksi', 'berupa', 'denda', 'bagi', 'warga', 'yang', 'menolak', 'vaksinasi', 'covid']	1 ['pemerintah', 'provinsi', 'dki', 'jakarta', 'sanksi', 'denda', 'warga', 'menolak', 'vaksinasi', 'covid']
['vaksin', 'belum', 'lulus', 'uji', 'bp', 'pom', 'dan', 'ternyata', 'ga', 'semua', 'orang', 'bisa', 'divaksinanehnya', 'udah', 'tuh', 'vaksin', 'distribusikan', 'ke', 'daerah', 'dan', 'menolak', 'divaksin', 'dikenakan', 'dendagagal', 'paham']	['vaksin', 'lulus', 'uji', 'bp', 'pom', 'orang', 'divaksinanehnya', 'tuh', 'vaksin', 'distribusikan', 'daerah', 'menolak', 'divaksin', 'dikenakan', 'dendagagal', 'paham']
['terus', 'liat', 'video', 'nakes', 'menolak', 'vaksin', 'itu', 'pasti', 'video', 'buat', 'seruseruan', 'kan', 'iya', 'kan', 'dok', 'sus']	['ntar', 'dipidanakan', 'menolak', 'vaksin']

Hasil dari proses ini dapat divisualisasikan melalui wordcloud. Visualisasi dapat memperlihatkan kata – kata yang dicetak dalam ukuran besar dan kecil yang merepresentasikan frekuensi kemunculan. Semakin besar ukuran kata maka semakin sering kata tersebut muncul dalam satu dokumen atau data.



Gambar 3. Wordcloud

### 3.2.4. Text Normalization

Sebelum melakukan normalisasi teks, data terlebih dahulu didefinisikan dari tweet\_token yang tersimpan selama proses tokenizing sama dengan tweet\_tokens WSW agar mudah memanggil fungsi tersebut ketika melakukan normalisasi dan diunduh. Selanjutnya, hasil unduhan dibaca kembali oleh

mesin dan di inialisasi. Normalisasi teks bertujuan untuk mengurangi karakter yang berulang sehingga tidak ada duplikat kata seperti pada Tabel 5.

Tabel 5. Perbandingan Sebelum dan Sesudah Normalization

Sebelum Normalization	Sesudah Normalization
['dom', 'beliau', 'dr', 'agni', 'heal', 'menolak', 'vaksin', 'mbak', 'menolak', 'apun', 'anjuran', 'pemerintah', 'kampanye', 'keitis', 'substansial', 'pcr', 'metode', 'test', 'plus', 'edukasi', 'how', 'to']	['dom', 'beliau', 'dr', 'agni', 'heal', 'vaksin', 'mbak', 'menolak', 'apun', 'anjuran', 'pemerintah', 'kampanye', 'keitis', 'substansial', 'pcr', 'metode', 'test', 'plus', 'edukasi', 'how', 'to']
['orang', 'bule', 'vaccine', 'antichrist', 'dianggap', 'sarana', 'kontrol', 'populasi', 'penduduk', 'dunia', 'wajar', 'menolak', 'program', 'vaksinasi', 'anti', 'vaksin']	['orang', 'bule', 'vaccine', 'antichrist', 'dianggap', 'sarana', 'kontrol', 'populasi', 'penduduk', 'dunia', 'wajar', 'menolak', 'program', 'vaksinasi', 'anti', 'vaksin']
1 ['pemerintah', 'provinsi', 'dki', 'jakarta', 'sanksi', 'denda', 'warga', 'menolak', 'vaksinasi', 'covid']	1 ['pemerintah', 'provinsi', 'dki', 'jakarta', 'sanksi', 'denda', 'warga', 'menolak', 'vaksinasi', 'covid']
['vaksin', 'lulus', 'uji', 'bp', 'pom', 'orang', 'divaksinanehnya', 'tuh', 'vaksin', 'distribusikan', 'daerah', 'menolak', 'divaksin', 'dikenakan', 'dendagagal', 'paham']	['vaksin', 'lulus', 'uji', 'bp', 'pom', 'orang', 'divaksinanehnya', 'tuh', 'vaksin', 'distribusikan', 'daerah', 'menolak', 'divaksin', 'dikenakan', 'dendagagal', 'paham']
['ntar', 'dipidanakan', 'menolak', 'vaksin']	['ntar', 'dipidanakan', 'menolak', 'vaksin']

### 3.2.4. Stemming

Pada proses ini, data hasil normalisasi berupa pemanggilan fungsi tweet\_normalized diolah kembali. Modul Python Sastrawi dapat membantu menghilangkan kata imbuhan dan infleksi kata ke bentuk dasar, menghilangkan prefiks serta sufiks berdasarkan bahasa Indonesia. Tahap akhir dari stemming adalah mendefinisikan tweet\_normalized sama dengan fungsi tweet\_token\_stemmed. Proses ini membutuhkan kamus bahasa Indonesia yaitu Sastrawi. Untuk melihat perbedaann dari data yang belum di stemming dengan yang sudah terdapat pada Tabel 6.

Tabel 6. Perbandingan Sebelum dan Sesudah Stemming

Sebelum Stemming	Sesudah Stemming
['dom', 'beliau', 'dr', 'agni', 'heal', 'vaksin', 'mbak', 'menolak', 'apun', 'anjuran', 'pemerintah', 'kampanye', 'keitis', 'substansial', 'pcr', 'metode', 'test', 'plus', 'edukasi', 'how', 'to']	['dom', 'beliau', 'dr', 'agni', 'heal', 'tolak', 'vaksin', 'mbak', 'tolak', 'apa', 'anjur', 'perintah', 'kampanye', 'kampanye', 'keitis', 'substansial', 'pcr', 'substansial', 'pcr', 'metode', 'test', 'plus', 'edukasi', 'how', 'to']
['orang', 'bule', 'vaccine', 'antichrist', 'dianggap', 'sarana', 'kontrol', 'populasi', 'penduduk', 'dunia', 'wajar', 'menolak', 'program', 'vaksinasi', 'anti', 'vaksin']	['orang', 'bule', 'vaccine', 'antichrist', 'anggap', 'sarana', 'kontrol', 'populasi', 'duduk', 'dunia', 'wajar', 'tolak', 'program', 'vaksinasi', 'anti', 'vaksin']

'vaksinasi', 'anti', 'vaksin']	'vaksin']
1 ['pemerintah', 'provinsi', 'dki', 'jakarta', 'sanksi', 'denda', 'warga', 'menolak', 'vaksinasi', 'covid']	['perintah', 'provinsi', 'dki', 'jakarta', 'sanksi', 'denda', 'warga', 'tolak', 'vaksinasi', 'covid']
['vaksin', 'lulus', 'uji', 'bp', 'pom', 'orang', 'divaksinanehnya', 'tuh', 'distribusikan', 'daerah', 'menolak', 'divaksin', 'dikenakan', 'dendagagal', 'paham']	['vaksin', 'lulus', 'uji', 'bp', 'pom', 'orang', 'divaksinanehnya', 'tuh', 'vaksin', 'distribusi', 'daerah', 'tolak', 'vaksin', 'kena', 'dendagagal', 'paham']
['ntar', 'dipidanakan', 'menolak', 'vaksin']	['ntar', 'pidana', 'tolak', 'vaksin']

### 3.3. Pembuatan Bag of Word (BOW)

Langkah selanjutnya setelah *text preprocessing* selesai adalah membuat dokumen dari dataset menjadi sebuah vektor yang dikelompokkan menjadi satu. Model Bag of Words mengklasifikasikan teks dan mengekstrak beberapa fitur dari teks [5]. Metode ini akan menghitung kemunculan setiap kata pada satu dokumen, sehingga pada penelitian ini didapatkan BOW untuk indeks ke 1794 dapat dilihat pada Tabel 7.

Tabel 7. Frekuensi Kemunculan Kata

Kata	Frequency kemunculan
Perintah	1
Tolak	1
Vaksin	1
Denda	1
Warga	1
Juta	1
Badan	1
Milik	2
Jakarta	1
hahahahaha	1

### 3.3. Term Frequency – Inverse document Frequency (TF-IDF)

Pembobotan *term* atau kata perdokumen dimulai dari perhitungan *term frequency*. Kemunculan setiap kata di dalam satu dokumen di dataset ini dihitung dan dinyatakan dalam nilai desimal. Setelah itu, pembobotan dari IDF menghasilkan nilai yang sama seperti TF yaitu nilai desimal. TF-IDF sering digunakan sebagai skema pembobotan untuk menentukan pentingnya sebuah kata dalam sebuah dokumen atau korpus. Semakin tinggi nilai TF-IDF, semakin penting kata tersebut dianggap.

Pada Tabel 8 kata "milik" memiliki nilai IDF tertinggi, yang menunjukkan bahwa kata tersebut relatif jarang digunakan dalam korpus. Di sisi lain, "Badan" memiliki nilai IDF tertinggi kedua, yang menunjukkan bahwa kata tersebut juga relatif tidak umum.

Dalam hal nilai TF, "milik" memiliki nilai tertinggi, yang berarti kata tersebut muncul paling

sering dalam dokumen. "hahahahah" memiliki nilai IDF terendah, menunjukkan bahwa kata tersebut yang sangat umum dalam korpus.

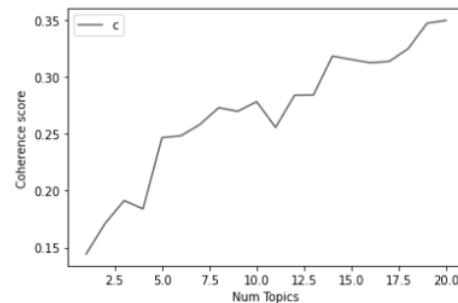
Tabel 8. Tem Frequency dan Inverse Document Frequency

Kata	TF	IDF
Perintah	0.09090909090909091	0.20717619356132486
Tolak	0.09090909090909091	0.000813054863081583
Vaksin	0.09090909090909091	0.002565325149954772
Denda	0.09090909090909091	0.24016760207854018
Warga	0.09090909090909091	0.2763205991390956
Juta	0.09090909090909091	0.35057161155065214
Badan	0.09090909090909091	0.5183740379595912
Milik	0.18181818181818182	0.9265415661791252
Jakarta	0.09090909090909091	0.5043603397934769
hahahahah	0.09090909090909091	0.6182478823839649

### 3.4. Latent Dirichlet Allocation (LDA)

Model LDA dapat menentukan topik yang diringkas secara subjektif, akan tetapi pada penelitian ini untuk menginterpretasikan topik – topik yang diperoleh secara maksimal, maka dilakukan perhitungan koherensi sebelum melakukan pemodelan topik. Perhitungan ini membutuhkan sekumpulan korpus yang telah dibuat, dan *dictionary*.

*Coherence score* dapat menampilkan topik dengan nilai atau *score* terbaiknya dengan menetapkan batasan atau jumlah maksimal topik yang akan ditampilkan. Pada penelitian ini, koherensi dihitung dengan memberi batasan sebanyak 21 dan dimulai dari 1 topik terbaik yang diinterpretasi, dan diiterasi 1000 kali.

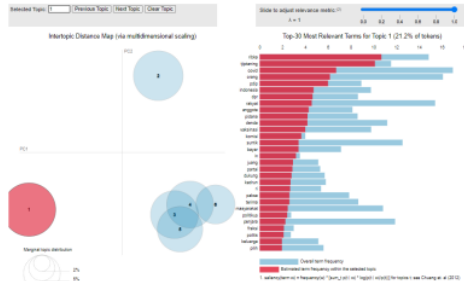


Gambar 4. Hubungan Topik dengan Corence Score

Untuk memilih jumlah topik yang optimal untuk model topik, dipilih jumlah topik yang memberikan skor koherensi tertinggi [12]. Pada grafik Gambar 4 menunjukkan *Num Topics* yang memiliki nilai tertinggi terdapat 6. Analisa ini akan dijadikan untuk menampilkan pemodelan LDA sebanyak jumlah topik yang paling tinggi.

Percobaan selanjutnya adalah pemodelan topik yang menentukan jumlah topik berdasarkan nilai koherensi terbaik, yaitu 6 jumlah topik dan menentukan jumlah kata per topik untuk

ditampilkan. Pada penelitian ini adalah 10 kata dan iterasi sebanyak 1000 kali. Kemudian hasil yang didapat dari pemodelan LDA akan divisualisasikan ke dalam modul pyLDAvis sehingga mudah untuk dianalisis topik apa saja yang muncul.



Gambar 5. Visualisasi Hasi LDA

Setiap kata yang memiliki frekuensi kemunculan berdasarkan topik yang dipilih akan ditampilkan pada pyLDAvis seperti pada contoh Gambar. Topik ke 1 memiliki nilai koherensi tinggi ketika dipilih akan menampilkan kata – kata dengan relevan kemunculan paling sering dengan topik diurutkan pada bagan paling atas.

Pemodelan topik LDA memungkinkan untuk melihat persebaran kata pada setiap topik yang telah ditentukan sebelum pemodelan, yaitu melalui *wordcloud* pada Gambar 6.



Gambar 6. Wordcloud Setiap Topik dari Pemodelan LDA

Pada Gambar 6 hasil analisa untuk Topik 0 dapat dilihat memiliki kata-kata seperti "PKI", "DPR", "FPI", dan "rakyat". Potensi topik yang mungkin terkait dengan kata - kata tersebut adalah topik seputar kebijakan pemerintah terkait vaksinasi dan pembahasan organisasi politik di Indonesia.

Topik 1 dalam hasil LDA pada Gambar 6 mempunyai kata “ribka”, “tjipaning”, “covid” sehingga menggambarkan pembicaraan tentang kontroversi beberapa waktu lalu yang dilakukan oleh Ribka Tjipaning. Pada topik 1 juga terdapat pembicaraan hukuman pidana yang berlaku.

Distribusi kata pada topik 2 dapat disimpulkan topik yang dibicarakan seputar hak asasi manusia yang ada terkait beredarnya anggapan masyarakat menganggap vaksin COVID-19 adalah pemaksaan. Kata “cina” menyinggung isu yang sempat *trend* dari dugaan masyarakat bahwa vaksin bagian dari strategi Cina.

Topik 3 mempunyai kata - kata terkait seperti “covid”, “masyarakat”, “warga”, “wajib”, “vaksin”, “hak”, “denda”. Kata-kata menunjukkan perbincangan terhadap kepatuhan aturan kesehatan yang harus diikuti masyarakat dalam vaksinasi dan hukuman bagi ketidakpatuhan berupa denda.

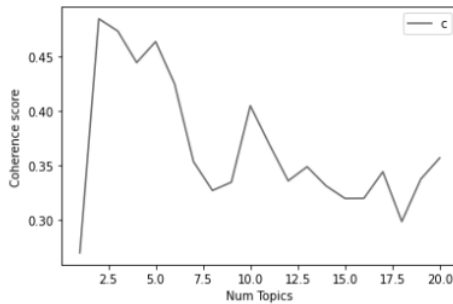
Sedangkan pada topik 4 dari kata – kata yang ada akan membentuk beberapa potensi seperti hal-hal seperti keamanan, kehalalan, manfaat, dan distribusi suatu produk. Kata-kata yang mendominasi dalam topik ini adalah "aman", "halal", "distribusi", "jamin", dan "manjur". Hal ini mungkin terkait dengan topik seperti sertifikasi halal, penjaminan mutu produk, dan regulasi distribusi produk di suatu negara atau daerah. Namun, untuk memahami konteks sebenarnya, perlu dilihat dokumen atau teks yang diolah untuk membentuk model LDA ini.

Distribusi pada topik 5 membahas vaksinasi COVID-19 dan isu terkait di Indonesia. Beberapa kata kunci dalam topik ini adalah "penjara", "sinovac", "suntik", "rakyat", "covid", "negara", "sanksi", "indonesia", dan "masyarakat". Kemungkinan topik ini membahas masalah seperti distribusi vaksin, keengganan untuk divaksinasi, penegakan mandat vaksinasi, dan sanksi potensial bagi mereka yang tidak mematuhi kebijakan vaksinasi.

### 3.4. Latent Semantic Analysis (LSA)

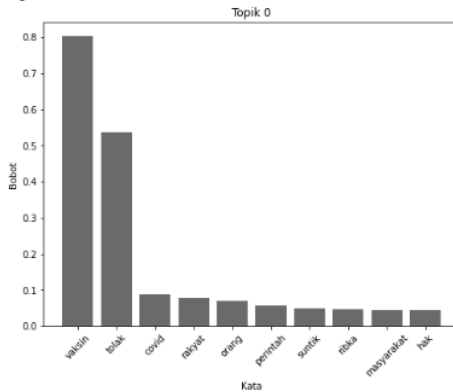
LSA mampu menemukan topik yang tersembunyi pada suatu *dataset*, sehingga hasil topik yang dihasilkan ketika melakukan pemodelan akan berbeda dengan pemodelan topik metode LDA. Untuk melakukan pemodelan topik setelah pembobotan korpus, LSA akan menghitung nilai koherensi sama seperti LDA.

Semua term hasil *preprocessing* dari semua pasal dibentuk menjadi sebuah matriks (Matrik M). Baris merepresentasikan semua term unik pada semua pasal dan Kolom merepresentasikan pasal. Bobot hasil pembobotan TF-IDF dimasukkan ke matriks berdasarkan *term* dan topiknya [13]. Matriks tersebut akan didekomposisi melalui SVD sehingga menghasilkan matriks U, S, dan V untuk mengurangi dimensinya sebelum dicari nilai koherensi setiap topik.



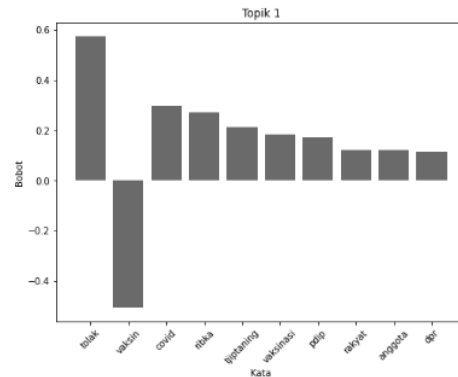
Gambar 7 Hubungan Topik dengan Coherence Score

Perhitungan *coherence score* pada Gambar 7 melibatkan maksimal perhitungan atau batas perhitungan, pada penelitian ini batas perhitungan topik terbaik yang akan diinterpretasi adalah 21 kali perhitungan pemodelan topik dan dimulai dari 1. Hasil koherensi menunjukkan nilai paling tertinggi 0.484607 pada jumlah topik ke 2. Untuk melihat persebaran topik yang dihasilkan pada model LSA berdasarkan nilai koherensi, dilakukan visualisasi menggunakan *bar chart* untuk masing – masing topik.



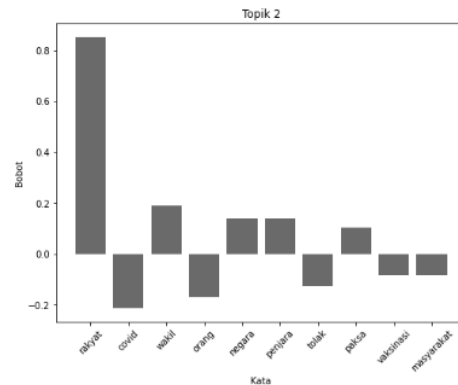
Gambar 8. Visualisasi Topic 0 dari Pemodelan LSA

Gambar 8 merupakan grafis dari topik 0 tersebut kemungkinan adalah penolakan terhadap vaksinasi COVID-19. Hal ini diduga berdasarkan kata-kata kunci dalam topik, seperti "tolak", "vaksin", "covid", "rakyat", "orang", dan "perintah". Kemungkinan topik ini membahas tentang alasan-alasan yang membuat sebagian orang menolak untuk divaksinasi, serta upaya-upaya pemerintah dalam meyakinkan masyarakat akan keamanan dan manfaat dari vaksinasi COVID-19. Ada juga kemungkinan bahwa topik ini membahas tentang hak individu dalam memilih untuk divaksinasi atau tidak. Kata "Ribka" mungkin merujuk pada nama seseorang yang terlibat dalam diskusi atau konteks tertentu, atau mungkin merupakan kata yang tidak relevan dengan topik tersebut.



Gambar 9. Visualisasi Topik 1 dari Pemodelan LSA

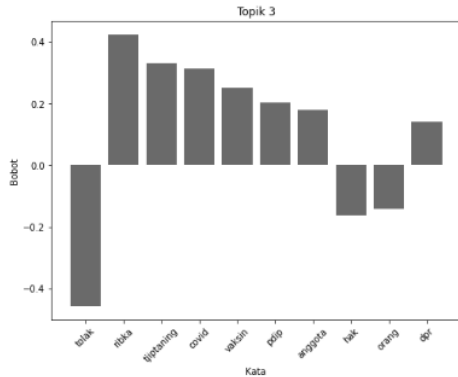
Hasil LSA Gambar 9 menunjukkan Topik ke-1, di mana kata - kata yang paling berkontribusi terhadap topik tersebut adalah "ribka" dan "covid". Topik tersebut berkaitan dengan topik yang dapat disimpulkan menjadi isu-isu yang terkait dengan Ribka Tjiptaning, yang terkenal karena kontroversi beberapa waktu lalu seputar vaksinasi COVID-19 di Indonesia. Topik ini juga terkait dengan penolakan terhadap vaksin, dengan kata kunci utama "tolak" dan "vaksin". Namun, terdapat beberapa kata kunci tambahan seperti "vaksinasi", "pdip", "rakyat", "anggota", dan "dpr". Kata kunci tersebut menunjukkan bahwa topik ini membahas tentang alasan penolakan vaksin, persepsi masyarakat dan anggota DPR terhadap vaksinasi, serta pandangan partai politik seperti PDIP tentang vaksin.



Gambar 10. Visualisasi Topik 2 dari Pemodelan LSA

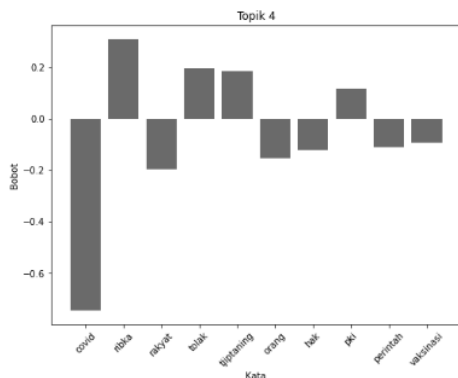
Visualisasi hasil LSA pada Gambar 10 terdiri dari beberapa kata kunci, yaitu "rakyat", "covid", "wakil", "orang", "negara", "penjara", "tolak", "paksa", "vaksinasi", dan "masyarakat". Hasil ini menunjukkan kemungkinan pembahasan yang berkaitan dengan pandemi Covid-19 di Indonesia dan mungkin juga membahas tentang penolakan atau paksaan terhadap vaksinasi dari perspektif rakyat,

negara, dan masyarakat. Kata kunci seperti "penjara" dan "paksa" dapat mengindikasikan adanya tindakan tegas dari pihak berwenang dalam menangani situasi pandemi yang mungkin tidak disetujui oleh sebagian masyarakat.



Gambar 11. Visualisasi Topik 3 dari Pemodelan LSA

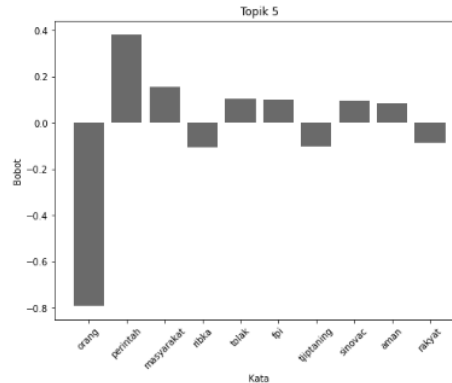
Berdasarkan hasil LSA pada Gambar 11, topik 3 didominasi oleh kata-kata seperti "tolak", "ribka", "tjiptaning", "covid", "vaksin", "pdip", "anggota", "hak", "orang", dan "dpr". Kemungkinan topik ini membahas tentang penolakan terhadap vaksin COVID-19 oleh sekelompok orang atau kelompok tertentu yang diwakili oleh tokoh-tokoh seperti Ribka Tjiptaning dan anggota DPR dari PDIP. Terdapat pula kemungkinan pembahasan tentang hak individu untuk menolak vaksinasi, meskipun hal ini bertentangan dengan upaya pemerintah untuk menangani pandemi COVID-19.



Gambar 12. Visualisasi Topik 4 dari Pemodelan LSA

Distribusi kata – kata berdasarkan Gambar 12 untuk topik 4 kata kunci yang muncul dalam topik ini adalah "covid", "ribka", "tolak", "tjiptaning", "orang", "hak", "pki", dan "vaksinasi". Terdapat kemungkinan pembahasan dalam topik ini adalah penolakan terhadap tindakan pemerintah dalam menangani pandemi COVID-19, seperti *lockdown* atau vaksinasi. Selain itu, mungkin juga membahas

mengenai pendapat orang tentang efektivitas vaksin, hak asasi manusia dalam situasi pandemi, dan potensi campur tangan partai politik seperti PKI dalam penanganan pandemi.



Gambar 13. Visualisasi Topik 5 dari Pemodelan LSA

Tema topik ini terkait dengan orang-orang, dengan kata kunci utamanya seperti yang divisualisasikan pada Gambar 13 adalah "orang", "perintah", "masyarakat", "ribka", dan "tolak". Kemungkinan topik ini membahas masalah sosial atau politik terkait dengan orang-orang, seperti pengaruh kebijakan pemerintah terhadap masyarakat, aksi tolak atau protes dari kelompok tertentu, atau peran individu tertentu dalam isu-isu sosial dan politik. Ada kemungkinan juga topik ini membahas isu-isu terkait hak asasi manusia dan keamanan masyarakat.

#### 4. DISKUSI

Pemodelan topik menggunakan metode LDA dan LSA masing – masing menghasilkan 6 topik utama yang telah divisualisasikan. Setiap topik mengandung kata – kata dengan nilai tertinggi, pada penelitian ini ditampilkan hanya 10 kata. Topik yang terbentuk berkaitan dengan tolak vaksin COVID-19, kontroversi Ribka Tjiptaning, denda dan hukum pidana terhadap kasus yang ada, pandangan partai politik PDIP terhadap vaksin, kontroversi vaksin Sinovac, dan perdebatan antara hak atau paksaan vaksinasi COVID-19. Untuk melihat bobot setiap kata dalam topik – topik yang dihasilkan oleh kedua model tersebut diuraikan dalam Tabel 9 dan Tabel 10.

Tabel 9. Hasil Pembentukan Model LDA

Topik 0	Topik 1	Topik 2	Topik 3	Topik 4	Topik 5
bu (0.003)	ribka (0.012)	ham (0.005)	hak (0.006)	aman (0.006)	penjara (0.006)
fpi (0.003)	tjiptaning (0.011)	langgar (0.005)	wajib (0.005)	halal (0.006)	sinovac (0.005)
orang (0.003)	covid (0.007)	paksa (0.003)	masyar akat	alas (0.006)	suntik (0.005)

rakyat (0.003)	orang (0.007)	suntik (0.003)	warga (0.004)	jamin (0.005)	rakyat (0.004)
anak (0.003)	pdip (0.007)	kasih (0.002)	perintah (0.004)	manjur (0.005)	covid (0.004)
pki (0.003)	indonesia (0.005)	infeksi (0.002)	terima (0.004)	distribusi (0.005)	negara (0.004)
dpr (0.003)	dpr (0.005)	kalo (0.002)	denda (0.004)	perintah (0.005)	orang (0.004)
perintah (0.002)	rakyat (0.005)	covid (0.002)	vaksin (0.004)	rakyat (0.004)	sanksi (0.003)
penjara (0.002)	anggota (0.005)	orang (0.002)	covid (0.004)	orang (0.003)	indonesia (0.003)
beliau (0.002)	pidana (0.005)	cina (0.002)	sehat (0.004)	hak (0.003)	masyarakat (0.003)

Tabel 10. Hasil Pembentukan Model LSA

Topik 0	Topik 1	Topik 2	Topik 3	Topik 4	Topik 5
vaksin (0.801)	tolak (0.576)	rakyat (0.854)	tolak (- (0.457)	covid (- (0.746)	orang (- (0.792)
tolak (0.536)	vaksin (- (0.507)	covid (- (0.214)	ribka (0.424)	ribka (0.311)	perintah (0.383)
covid (0.089)	covid (0.295)	wakil (0.191)	tjiptani (0.330)	rakyat (- (0.198)	masyarakat (0.157)
rakyat (0.078)	ribka (0.269)	orang (- (0.170)	covid (0.314)	tolak (0.196)	ribka (- (0.106)
orang (0.071)	tjiptani (0.212)	negara (0.140)	vaksin (0.252)	tjiptani (- (0.184)	tolak (0.105)
perintah (0.057)	vaksinasi (0.183)	penjara (0.139)	pdip (0.204)	orang (- (0.156)	fpi (0.100)
suntik (0.049)	pdip (0.174)	tolak (- (0.127)	anggota (0.180)	hak (- (0.122)	tjiptani (- (0.100)
ribka (0.046)	rakyat (0.123)	paksa (0.103)	hak (- (0.162)	pki (0.118)	sinovac (0.098)
masyarakat (0.045)	anggota (0.120)	vaksinasi (- (0.084)	orang (- (0.142)	perintah (- (0.112)	aman (0.087)
hak (0.045)	dpr (0.114)	masyarakat (- (0.082)	dpr (0.141)	vaksinasi (- (0.094)	rakyat (- (0.086)

Penelitian yang menggunakan metode LDA atau LSA, serta kasus yang berkaitan dengan COVID-19 telah banyak dilakukan sebelumnya. Adapun perbedaan dari penelitian ini dengan penelitian yang dilakukan oleh Fajar Fathur Rachman, dan Setia Pramana [4] dengan judul “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter” pada tahun 2020. Penelitian ini menggunakan kata kunci “Vaksin Covid” dan “Vaksin Corona” untuk menjangkau data Twitter. Model yang digunakan dalam melakukan analisa

sentimen pada penelitian tersebut adalah *Lexicon Based*, dan pemodelan topik algoritma LDA. Penelitian ini menghasilkan 20 topik pembahasan dari pemodelan LDA dan analisa sentimen mampu membuat kluster *tweet* positif sejumlah 1461, *tweet* negatif 1167, dan *tweet* netral 2313.

Penelitian terhadap kasus seputar Covid – 19 ada pada penelitian yang dilakukan Kevin Rafi Adjie Putra Santoso, Asmaul Husna, Nadia Widyawati Putri, Nur Aini Rakhmawati [14], diketahui bahwa analisis jumlah topik yang mengandung *hashtag* covidindonesia mendapatkan 3 topik melalui LDA memiliki nilai *perplexity* kecil dengan nilai koherensi yang tinggi

Penelitian yang dilakukan oleh Herlina Jayadianti, Ruth Damayanti, dan Juwairiah [15] pada artikel yang bersumber dari *website* Kompas.com, Liputan6.com, Klikdokter.com, detikhealth.com menghasilkan model LSA yang dibantu dengan *cross method* memberi hasil ringkasan terbaik. Model ini mampu menampilkan 12 dokumen.

## 5. KESIMPULAN

Pada penelitian ini telah diaplikasikan pemodelan melalui LDA dan LSA terhadap 1797 data *tweet* dengan bantuan pengambilan topik yang memiliki skor terbaik menggunakan *coherence score*. Hasilnya terdapat 6 topik bernilai koherensi tinggi pada masing – masing metode pemodelan. LDA mampu meringkas topik yang berkaitan dengan tolak vaksinasi COVID-19 berupa kebijakan pemerintah mengenai vaksinasi, kontroversi Ribka Tjiptaning, anggapan masyarakat bahwa vaksinasi COVID-19 adalah paksaan, campur tangan Cina terhadap vaksin tersebut, denda dan hukuman ketidakpatuhan aturan kesehatan, dan sertifikasi halal vaksin COVID-19.

Sementara itu, pada pemodelan LSA telah menyimpulkan topik – topik yang berhubungan dengan kata kunci tolak vaksinasi COVID-19 diantaranya adalah upaya pemerintah meyakinkan masyarakat tentang vaksinasi, kontroversi Ribka Tjiptaning seputar vaksinasi COVID-19, persepsi organisasi politik seperti PDIP dan PKI terhadap vaksin COVID-19, tindakan pemerintah terhadap isu penolakan vaksin oleh masyarakat, hak individu masyarakat dalam menolak vaksin tersebut, efektivitas vaksin, dan paksaan vaksinasi COVID-19.

Kesimpulan dari eksperimen pemodelan pada penelitian ini adalah topik – topik yang dihasilkan LDA dan LSA memiliki beberapa perbedaan, baik dari bobot topik maupun pembahasan.

## DAFTAR PUSTAKA

- [1] P. A. Sumitro, Rasiban, D. I. Mulyana, and W. Saputro, “Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter

- Menggunakan Metode Lexicon Based,” *J-ICOM - J. Inform. dan Teknol. Komput.*, vol. 2, no. 2, pp. 50–56, 2021, doi: 10.33059/jicom.v2i2.4009.
- [2] Q. A. Chairunnisa, Y. Herdiyeni, M. K. D. Hardhienata, and J. Adisantoso, “Analisis Sentimen Pengguna Twitter Terhadap Program Vaksinasi Covid-19 di Indonesia Menggunakan Algoritme Support Vector Machine,” *J. Ilmu Komput. dan Agri-Informatika*, vol. 9, no. 1, pp. 79–89, 2022, doi: 10.29244/jika.9.1.79-89.
- [3] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, and T. Zhu, “Public discourse and sentiment during the COVID 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter,” *PLoS One*, vol. 15, no. 9 September, pp. 1–12, 2020, doi: 10.1371/journal.pone.0239441.
- [4] F. F. Rachman and S. Pramana, “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter,” *Heal. Inf. Manag. J.*, vol. 8, no. 2, pp. 100–109, 2020, [Online]. Available: <https://inohim.esaunggul.ac.id/index.php/INO/article/view/223/175>
- [5] A. Muzaki and A. Witanti, “Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier,” *J. Tek. Inform.*, vol. 2, no. 2, pp. 101–107, 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [6] S. Sarica and J. Luo, “Stopwords in technical language processing,” *PLoS One*, vol. 16, no. 8 August, pp. 1–13, 2021, doi: 10.1371/journal.pone.0254937.
- [7] A. Amalia, D. Gunawan, Y. Fithri, and I. Aulia, “Automated Bahasa Indonesia essay evaluation with latent semantic analysis,” *J. Phys. Conf. Ser.*, vol. 1235, no. 1, 2019, doi: 10.1088/1742-6596/1235/1/012100.
- [8] B. O. Karo Karo, D. S. Naga, and V. C. Mawardi, “Perancangan Aplikasi Pendeteksi Kemiripan Teks Dengan Menggunakan Metode Latent Semantic Analysis,” *Comput. J. Comput. Sci. Inf. Syst.*, vol. 4, no. 1, p. 1, 2020, doi: 10.24912/computatio.v4i1.7191.
- [9] H. J. Kang, C. Kim, and K. Kang, “Analysis of the trends in biochemical research using latent dirichlet allocation (LDA),” *Processes*, vol. 7, no. 6, pp. 1–14, 2019, doi: 10.3390/PR7060379.
- [10] L. W. Narendra, “Topic Modeling in Conversational Dialogs for Naming Intent Labels Using LDA,” *JTECS J. Sist. Telekomun. Elektron. Sist. Kontrol Power Sist. dan Komput.*, vol. 2, no. 1, p. 65, 2022, doi: 10.32503/jtecs.v2i1.1820.
- [11] D. Ridhwanulah and D. H. Fudholi, “Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode Latent Dirichlet Allocation,” *J. Ilm. SINUS*, vol. 20, no. 1, p. 11, 2022, doi: 10.30646/sinus.v20i1.589.
- [12] F. Alattar and K. Shaalan, “Emerging Research Topic Detection Using Filtered-LDA,” *Ai*, vol. 2, no. 4, pp. 578–599, 2021, doi: 10.3390/ai2040035.
- [13] A. H. Ardiansyah, K. P. Kartika, and S. N. Budiman, “Penerapan Latent Semantic Indexing Pada Sistem Temu Balik Informasi Pada Undang-Undang Pemilu Berdasarkan Kasus,” *J. Mnemon.*, vol. 4, no. 2, pp. 64–70, 2021.
- [14] K. R. A. P. Santoso, A. Husna, N. W. Putri, and N. A. Rakhmawati, “Analisis Topik Tagar Covidindonesia pada Instagram Menggunakan Latent Dirichlet Allocation,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 1–9, 2022, doi: 10.14421/jiska.2022.7.1.1-9.
- [15] H. Jayadianti, R. Damayanti, and ..., “Latent Semantic Analysis (Lsa) Dan Automatic Text Summarization (Ats) Dalam Optimasi Pencarian Artikel Covid 19,” *Semin. Nas. ....*, vol. 2020, no. Semnasif, pp. 52–59, 2020, [Online]. Available: <http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/4085>

# Cek Plagiasi Ulfah

## ORIGINALITY REPORT

7%

SIMILARITY INDEX

7%

INTERNET SOURCES

2%

PUBLICATIONS

3%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://www.genpi.co">www.genpi.co</a> Internet Source	2%
2	<a href="http://dspace.uii.ac.id">dspace.uii.ac.id</a> Internet Source	1%
3	Submitted to Ciputra University Student Paper	1%
4	<a href="http://jurnal.stikesyatsi.ac.id">jurnal.stikesyatsi.ac.id</a> Internet Source	1%
5	<a href="http://repository.uin-suska.ac.id">repository.uin-suska.ac.id</a> Internet Source	1%
6	<a href="http://etd.unsyiah.ac.id">etd.unsyiah.ac.id</a> Internet Source	1%
7	<a href="http://media.neliti.com">media.neliti.com</a> Internet Source	1%

Exclude quotes  On

Exclude bibliography  On

Exclude matches  < 1%