

TOPIC MODELING IN COVID-19 VACCINATION REFUSAL CASES USING LATENT DIRICHLET ALLOCATION AND LATENT SEMANTIC ANALYSIS

Oleh:

Ulfah Malihatin S

Yulian Findawati

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

April, 2023

Pendahuluan

Berawal dari ramainya kemunculan vaksin Covid – 19 untuk pertama kali di Indonesia menjadikan banyak pro dan kontra sehingga terdapat sejumlah masyarakat yang menolak Vaksinasi Covid – 19.

- Alasan menolak vaksinasi Covid – 19
- Meringkas topik yang sering dibahas oleh pengguna Twitter

Tujuan dan Manfaat

Tujuan :

1. Mendapatkan ringkasan topik utama yang tersirat tentang penolakan vaksin Covid – 19.
2. mengimplementasikan metode coherence score untuk mendapatkan nilai tertinggi setiap topik.
3. Mengetahui topik – topik yang tersembunyi dari kedua model tersebut.

Manfaat :

1. Bagi penulis, diharapkan dapat menambah pengetahuan mengenai pemodelan topik dan opini – opini masyarakat terhadap kasus tersebut.
2. Bagi masyarakat, diharapkan dapat memberi pemahaman terhadap pendapat penolakan vaksin baik negatif maupun positif yang ada di Twitter.

Penelitian Sebelumnya

- **Fajar Fathur Rachman, dan Setia Pramana (2020)**

Judul : Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter

Analisa sentimen yang didapat adalah sentimen positif berjumlah 1461 *tweet* (29,6%), netral 2313 *tweet* (46,8%), dan negatif 1167 *tweet* (23,6%) dan mendapatkan 20 topik pembahasan

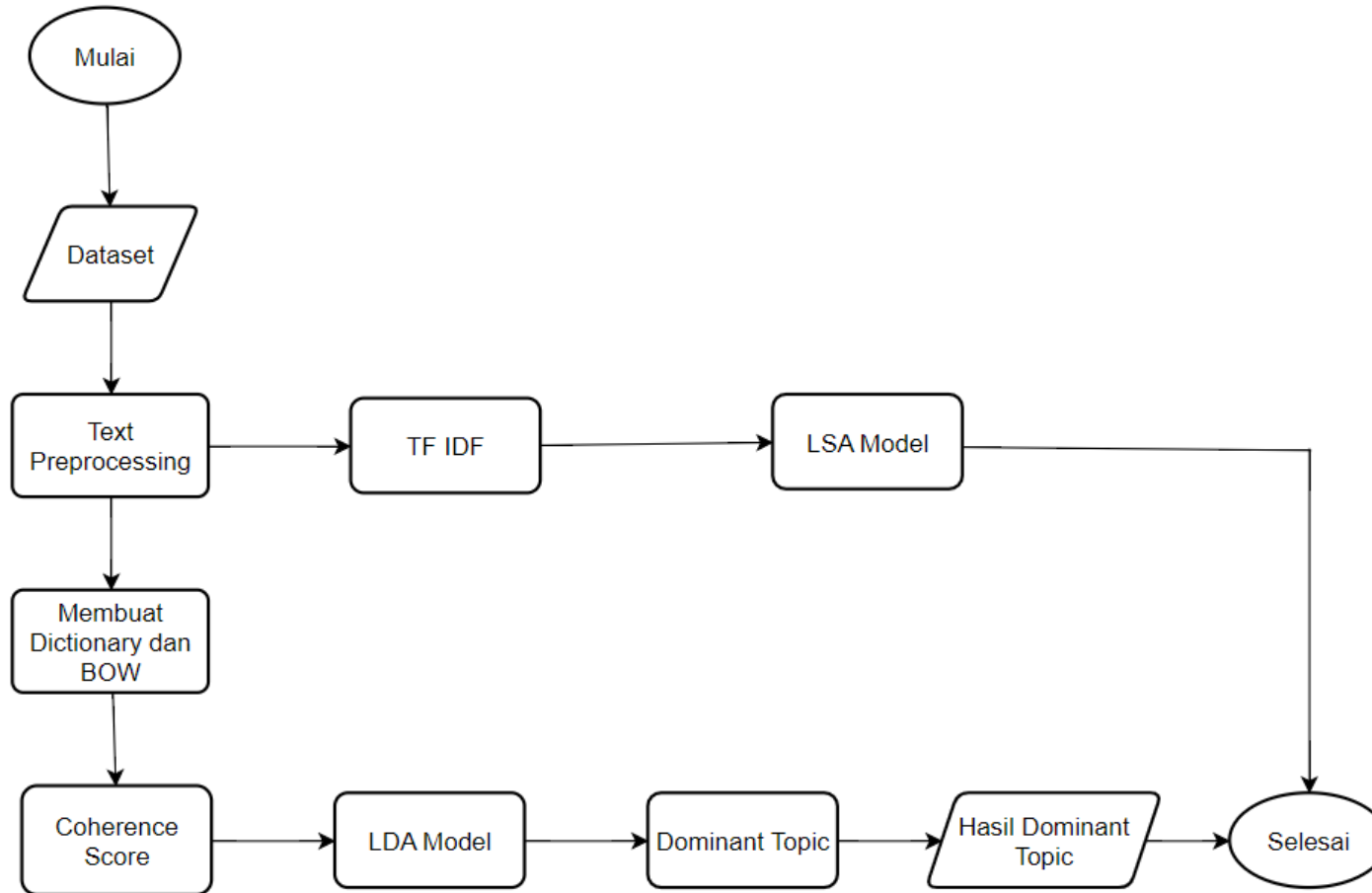
.

- **Novita Anggraini, Edi Surya Negara Harahap, Tri Basuki Kurniawan**

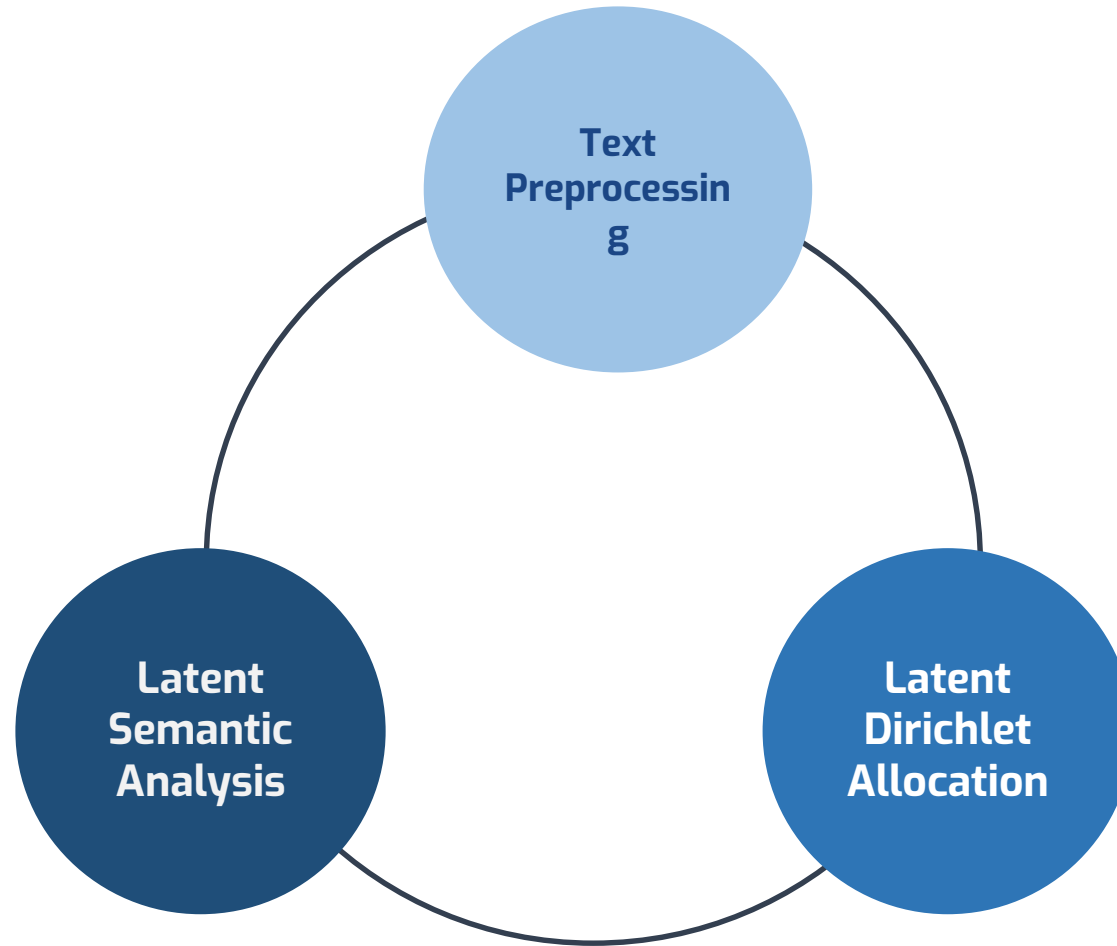
Judul : Text Mining - Analisis Teks Terkait Isu Vaksinasi COVID-19

Terdapat 3 kluster pengklasifikasian respon masyarakat, yaitu sebanyak jumlah masing – masing nilai setiap label negatif, positif, maupun netral.

Flowchart System



Metode



Metode

Text
Preprocessing

Lower Case Folding

Text Normalization

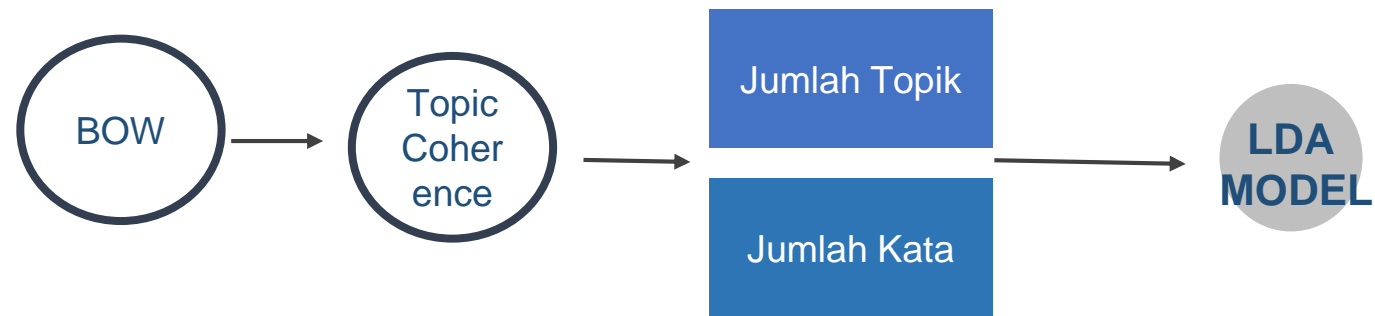
Tokenizing

Stemming

Stopwords

Metode

Latent
Dirichlet
Allocation

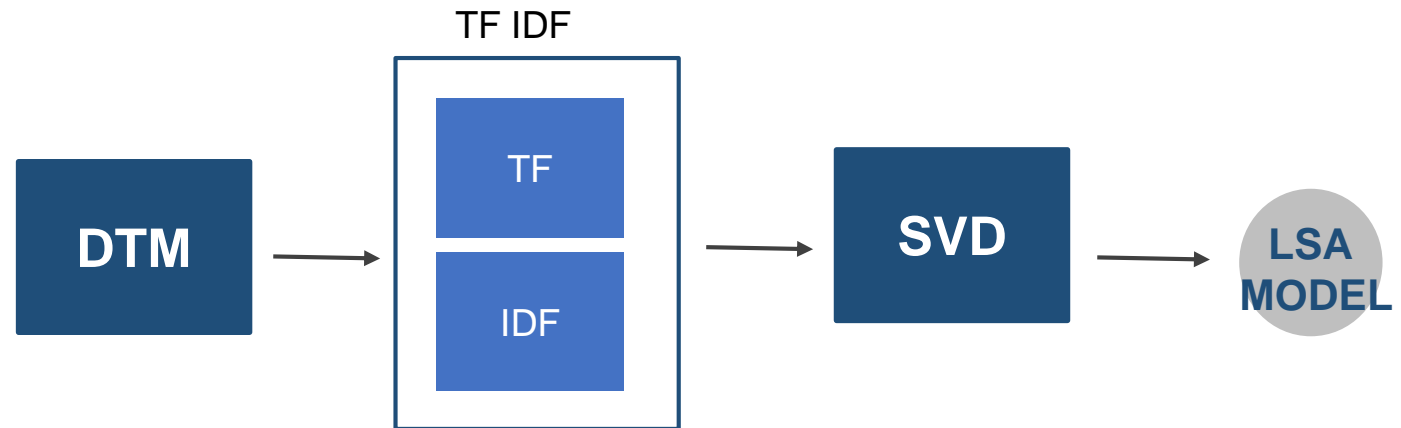


“Metode pemodelan **probabilistik** yang digunakan untuk mengidentifikasi topik yang mendasari kumpulan dokumen.”

- Mengidentifikasi topik-topik yang mendasari kumpulan dokumen
 - Mengetahui topik paling dominan

Metode

Latent
Semantic
Analysis



Hasil

Lower Case Folding

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 🤔👉👉👉 <https://t.co/JZL3oYy29f>

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 🤔🤔👉👉 <https://t.co/jzl3oYy29f>

Text Preprocessing

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 🤔👉👉👉 <https://t.co/jzl3oYy29f>

['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']

Tokenizing

Stopwords

['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']

Hasil

Text Normalization

['vaksin', 'lulus', 'uji', 'bp', 'pom', 'orang', 'divaksinanehnya', 'tuh', 'vaksin', 'distribusikan', 'daerah', 'menolak', 'divaksin', 'dikenakan', 'dendagagal', 'paham']

['vaksin', 'lulus', 'uji', 'bp', 'pom', 'orang', 'divaksinanehnya', 'tuh', 'distribusikan', 'daerah', 'menolak', 'divaksin', 'dikenakan', 'dendagagal', 'paham']

Text Preprocessing

Stemming

['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']

Hasil

Latent
Dirichlet
Allocation

Bag Of Words (BOW)

Contoh pada indeks data ke 1794

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

Hasil

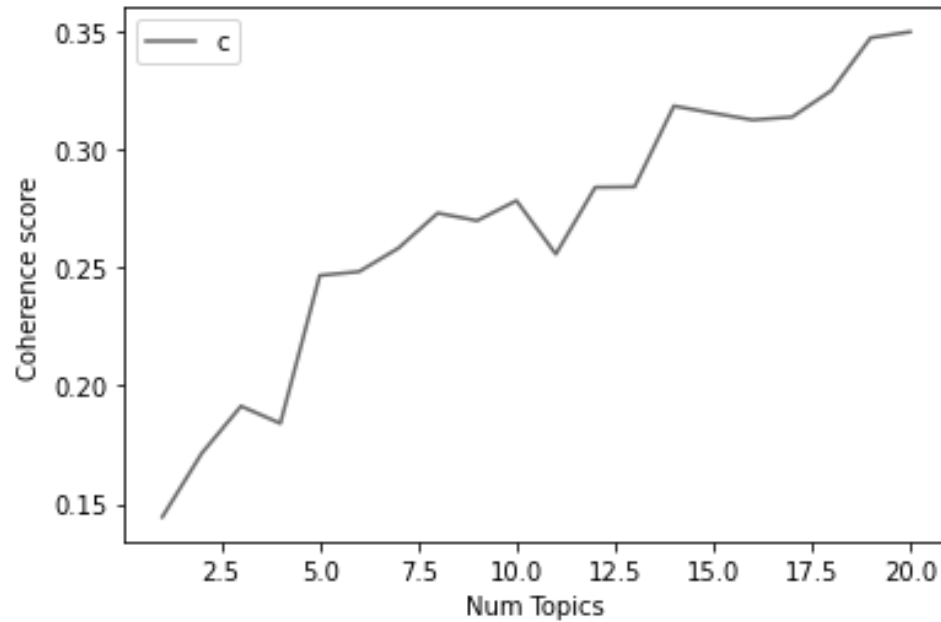
Term Frequency – Term Inverse Document (TF-IDF)

Contoh pada indeks data ke 1794

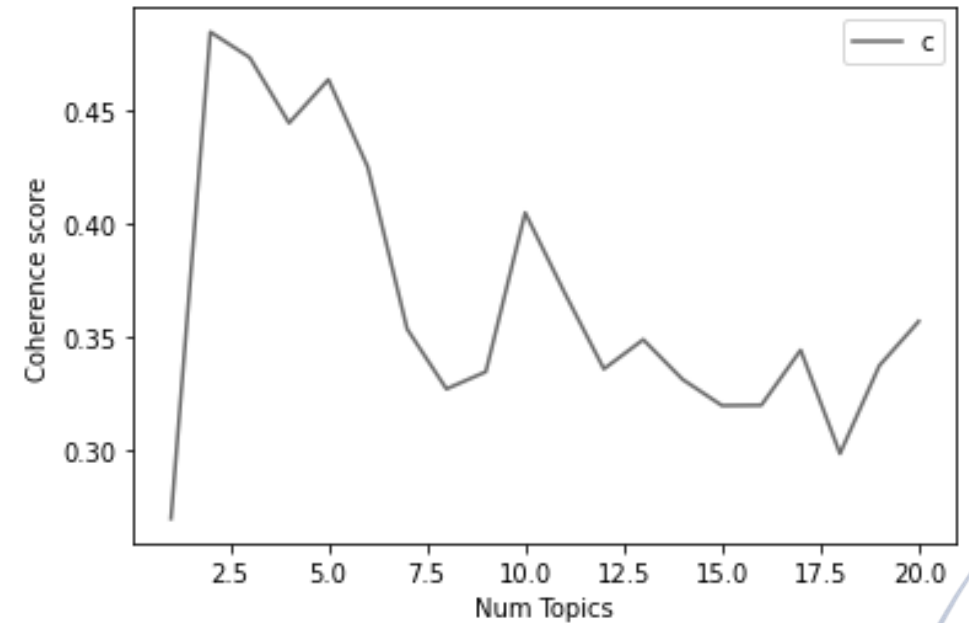
Latent Semantic Analysis

Kata	TF	IDF
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

Coherence Score



Latent Dirichlet Allocation



Latent Semantic Analysis

Hasil LDA

Topik 0

Kebijakan pemerintah dan organisasi politik di Indonesia terkait vaksinasi.

Topik 3

kepatuhan aturan kesehatan yang harus diikuti masyarakat dalam vaksinasi dan hukuman bagi ketidakpatuhan berupa denda

Topik 1

Pembahasan hukuman pidana yang berlaku.

Topik 4

kehalalan, dan keamanan vaksin COVID-19.

Topik 2

hak asasi manusia yang ada terkait beredarnya anggapan masyarakat menganggap vaksin COVID-19 adalah pemaksaan, dan vaksin bagian dari strategi Cina.

Topik 5

sanksi potensial bagi mereka yang tidak mematuhi kebijakan vaksinasi.

Hasil LDA

Topik 0

upaya-upaya pemerintah dalam meyakinkan masyarakat akan keamanan dan manfaat dari vaksinasi COVID-19.

Topik 1

persepsi masyarakat dan anggota DPR terhadap vaksinasi, serta pandangan partai politik seperti PDIP tentang vaksin.

Topik 2

tindakan tegas dari pihak berwenang dalam menangani situasi pandemi yang mungkin tidak disetujui oleh sebagian masyarakat.

Topik 3

sekelompok orang atau kelompok tertentu yang diwakili tokoh-tokoh seperti Ribka Tjiptaning, hak individu untuk menolak vaksinasi.

Topik 4

potensi campur tangan partai politik seperti PKI dalam penanganan pandemi.

Topik 5

pengaruh kebijakan pemerintah terhadap masyarakat, aksi tolak atau protes dari kelompok tertentu, atau peran individu tertentu dalam isu-isu sosial dan politik,

Referensi

S. S. Aljameel et al., "A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent covid-19 outbreaks in Saudi Arabia," *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, pp. 1–12, 2021, doi: 10.3390/ijerph18010218.

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/j-icom.v2i2.4009. The Oxford Dictionary of Computing, 5th ed. Oxford: Oxford University Press, 2003.

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. O. B. R. Strimpel, "Computer graphics," in *McGraw-Hill Encyclopedia of Science and Technology*, 8th ed., Vol. 4. New York: McGraw-Hill, 1997, pp. 279–283.

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. A. Altun, "Understanding hypertext in the context of reading on the web: Language learners' experience," *Current Issues in Education*, vol. 6, no. 12, July 2003. [Online]. Available: <http://cie.ed.asu.edu/volume6/number12/>. [Accessed Dec. 2, 2004].

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>. R. Beveridge and E. M. Riseman, "How easy is matching 2D line models using local search?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 564–579, June 1997.

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.

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. J. Lach, "SBFS: Steganography based file system," in *Proc. of the 2008 1st Int. Conf. on Information Technology, IT 2008, 19-21 May 2008, Gdansk, Poland* [Online]. Available: IEEE Xplore, <http://www.ieee.org>. [Accessed: 10 Sept. 2010].

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. T. J. van Weert and R. K. Munro, Eds., *Informatics and the Digital Society: Social, ethical and cognitive issues*: IFIP TC3/WG3.1&3.2 Open Conf. on Social, Ethical and Cognitive Issues of Informatics and ICT, July 22–26, 2002, Dortmund, Germany. Boston: Kluwer Academic, 2003.

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. European Telecommunications Standards Institute, "Digital Video Broadcasting (DVB): Implementation guidelines for DVB terrestrial services; transmission aspects," *European Telecommunications Standards Institute*, ETSI TR-101-190, 1997. [Online]. Available: <http://www.etsi.org>. [Accessed: Aug. 17, 1998].

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. G. Sussman, "Home page - Dr. Gerald Sussman," July 2002. [Online]. Available: <http://www.comm.pdx.edu/faculty/Sussman/sussmanpage.htm>. [Accessed: Sept. 12, 2004].

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. A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

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. *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.

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.

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.

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. December 2019, 2019, doi: 10.1063/1.5139825.

T. Williams and J. Betak, "A Comparison of LSA and LDA for the Analysis of Railroad Accident Text," vol. 11, no. 1, pp. 11–15, 2019, doi: 10.5383/JUSPN.11.01.002.

