

# Text Mining untuk Analisis Kasus Stunting di Nusa Tenggara Barat

Fahmi Syuhada<sup>a</sup>, Yuan Sa'adati<sup>\*b</sup>, Lia Arian Apriani<sup>c</sup>,

Syahrani Lonang<sup>d</sup>, Ahmad Fatoni Dwi Putra<sup>e</sup>

<sup>a,b,e</sup>Program Studi Ilmu Komputer, Universitas Qamarul Huda Badaruddin Bagu, Lombok Tengah, Indonesia

<sup>c</sup>Program Studi Kebidanan, Universitas Qamarul Huda Badaruddin Bagu, Lombok Tengah, Indonesia

<sup>d</sup>Program Studi Teknologi Informasi, Universitas Qamarul Huda Badaruddin Bagu, Lombok Tengah, Indonesia

email: <sup>a</sup>fahmisy@unihba.ac.id, <sup>b</sup>yuan@unihba.ac.id \*, <sup>c</sup>liaa@unihqba.ac.id, <sup>d</sup>lonang@unihba.ac.id,

<sup>e</sup>ahmadfatoni@unihba.ac.id

DOI: <https://doi.org/10.21107/edutic.v12i1.29522>

Diterima: 17 Maret 2025 | Direvisi: 01 April 2025 | Diterbitkan : 04 Mei 2025

## Abstrak

Pemberitaan stunting yang menjadi permasalahan nasional khususnya di provinsi Nusa Tenggara Barat (NTB) sudah masif tersedia pada dunia maya beberapa tahun terakhir. Oleh sebab itu analisis terhadap trend pemberitaan kasus ini sangat menarik dilakukan. Tujuannya yaitu melihat istilah-istilah kata yang berhubungan dengan kata pada pemberitaan stunting. Dengan ini akan diketahui istilah-istilah atau kata-kata dan pola komunikasi publik terkait isu stunting di NTB. Penelitian ini mengusulkan penerapan teknik Text Mining dalam menganalisis trend stunting NTB pada pemberitaan dunia maya. Data dikoleksi dari portal berita dengan query yang berkaitan dengan stunting NTB dari tahun 2018 hingga 2024. Penerapan metode text mining seperti preprocessing, eksplorasi data (EDA) dan Latent Dirichlet Allocation (LDA) hingga visualisasi hasil digunakan untuk reprints trend tersebut. Kontribusi penelitian difokuskan pada bagaimana analisis trend berdasarkan tiga kumpulan corpus yaitu trend Utama Stunting, Sebab, dan Dampak. Analisis tren NTB menunjukkan bahwa istilah stunting, gizi, dan air bersih mendominasi pemberitaan, mencerminkan fokus pada faktor kesehatan utama dalam pencegahan stunting. Faktor sosial seperti nikah muda dan pendidikan juga memiliki hubungan signifikan, menunjukkan perlunya pendekatan yang mencakup dimensi sosial dan budaya. Sementara itu, intervensi berbasis komunitas, seperti posyandu, berperan penting dalam mendukung edukasi dan pemantauan gizi anak. Namun, istilah geografis seperti Bima, meskipun sering muncul, tidak memiliki hubungan langsung dengan stunting, melainkan lebih terkait dengan konteks administratif. Keseluruhan analisis menegaskan perlunya pendekatan terintegrasi yang melibatkan faktor kesehatan, sosial, dan komunikasi publik yang efektif untuk menekan prevalensi stunting di NTB.

**Kata kunci:** NTB, Pola Komunikasi Publik, Stunting, Text Mining, Trend.

## Abstract

The widespread coverage of stunting as a national concern, particularly in West Nusa Tenggara (NTB) Province, has been increasingly prominent in online media in recent years. This study aims to analyze trends in news reporting on this issue, focusing on identifying key terms and phrases associated with stunting. By doing so, it seeks to uncover patterns in public communication regarding stunting in NTB. This research proposes the use of Text Mining techniques to examine stunting trends in online news. Data were collected from news portals using queries related to stunting in NTB from 2018 to 2024. Techniques such as preprocessing, exploratory data analysis (EDA), and Latent Dirichlet Allocation (LDA) were applied, along with data visualization, to map these trends. The study contributes by analyzing trends across three corpora: Main Stunting Trends, Causes, and Impacts. The findings indicate that terms like stunting, nutrition, and clean water dominate news coverage, emphasizing the focus on health-related factors in stunting prevention. Social factors, including early marriage and education, also play a significant role, highlighting the need for approaches that address social and cultural dimensions. Community-based interventions, such as posyandu (integrated health service posts), are crucial for supporting education and child nutrition monitoring. However, geographic terms like Bima, while frequently mentioned, are not directly linked to stunting but rather reflect administrative contexts. The analysis underscores the importance of an integrated approach that combines health, social, and effective public communication strategies to reduce stunting prevalence in NTB.

**Keywords:** NTB, Public Communication Patterns, Stunting, Text Mining, Trends.



© 2025 Author (s)

---

## PENDAHULUAN

Stunting merupakan permasalahan nasional yang mengistilahkan gizi kronis karna multifaktorial dan ditandai dengan ketidaksesuaian tinggi badan bayi dengan usianya (Apriani et al., 2024; Madani Pertiwi & Hendrati, 2023). Kasus ini bisa terjadi sejak anak masih dalam kandungan hingga usia dua tahun (Hamdi et al., 2023). Dampak yang ditimbulkan yaitu pertumbuhan fisik anak yang terhambat dan perkembangan otak hingga kemampuan dalam proses belajar. Malnutrisi kronis yang sering terjadi di negara berkembang, menjadi penyebab utama stunting. Anak-anak yang mengalami stunting akan kesulitan untuk mencapai potensi penuh mereka (Mustika & Syamsul, 2018). Kasus stunting telah diamati di seluruh Indonesia dan dianggap sebagai masalah besar di sektor kesehatan yang memerlukan penyelesaian dini. Nusa Tenggara Barat (NTB) merupakan salah satu daerah dengan angka prevalensi stunting tertinggi keempat di Indonesia sebesar 32,7%, setelah provinsi Papua sebesar 34,6% (Apriani et al., 2024). Berdasarkan Survei Status Gizi Indonesia (SSGI), prevalensi stunting di Nusa Tenggara Barat meningkat dari 31,4 persen pada tahun 2021 menjadi 32,7 persen pada tahun 2022 (Irwansyah et al., n.d.). Peningkatan prevalensi stunting jauh lebih tinggi dibandingkan rata-rata nasional. Nilai tersebut melebihi rata-rata nasional (21,6%) dan nilai standar WHO (kurang dari 20%), serta masih jauh dari target nasional sebesar 14% pada tahun 2024 (Wardani et al., 2021). Berdasarkan kondisi ini eksplorasi kasus stunting di NTB menjadi topik yang sangat penting dilakukan.

Salah satu hal yang menarik untuk diketahui dalam kasus ini yaitu bagaimana trend pemberitaan pada dunia maya. Tujuan yang diharapkan yaitu melihat istilah-istilah kata atau frasa yang berhubungan dengan stunting. Dengan ini akan diketahui istilah-istilah atau kata-kata yang sering dikaitkan dengan stunting di NTB dalam pemberitaan di dunia maya. Analisis ini bertujuan untuk mengidentifikasi pola komunikasi publik terkait isu stunting, sehingga dapat memberikan gambaran bagaimana isu ini dipersepsikan oleh masyarakat. Selain itu, analisis ini juga dapat mengungkap tema-tema utama yang sering dibahas, seperti gizi buruk, kemiskinan, kesehatan ibu dan anak, atau akses terhadap air bersih. Hasil analisis ini dapat dimanfaatkan untuk mendukung strategi komunikasi kebijakan, dengan menggunakan bahasa yang lebih efektif dan sesuai dengan isu yang sedang hangat dibicarakan. Dengan memahami istilah-istilah yang relevan, pemerintah atau lembaga terkait dapat menyusun pesan yang lebih tepat sasaran dalam kampanye pencegahan stunting.

Salah satu bidang penelitian Ilmu Komputer untuk menyelesaikan tugas ini yaitu pengolahan bahasa alami (Natural Language Processing) atau lebih spesifik Text Mining. Teks Mining merupakan teknik untuk mendapatkan informasi-informasi atau pengetahuan dari serangkaian data teks yang tidak terstruktur (Fanni et al., 2023; Greven et al., 2023; Hung & Zhang, 2012). Text Mining akan memperoleh data frekuensi istilah-istilah penting dalam satu atau kumpulan konten data tekstual pada sumber data seperti forum, artikel, atau halaman web kemudian mengklasifikasikan hubungannya antar fitur (Aggarwal & Aggarwal, 2015; Francis, 2006). Kim dkk. menjabarkan tentang teknik text mining yang diterapkan untuk memahami analisis trend kebijakan konsumen (Kim et al., 2017). Kemudian penelitian Amado et al menjelaskan bahwa text mining telah digunakan untuk mengidentifikasi trend utama Big Data dalam pemasaran (Amado et al., 2018) Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. Hasil penelitian membantu memajukan upaya yang lebih langsung ke arah bisnis Big Data di arena pemasaran. Penelitian tentang penambangan teks diterapkan untuk penemuan pengetahuan dalam penelitian akademis (Ojo & Adeyemo, 2013). Selanjutnya penelitian yang dilakukan oleh Sudigyo et al membuktikan bahwa teks mining berhasil dilakukan menggunakan dataset dari kumpulan studi literatur terkait intervensi stunting pada anak Indonesia (Sudigyo et al., 2023).

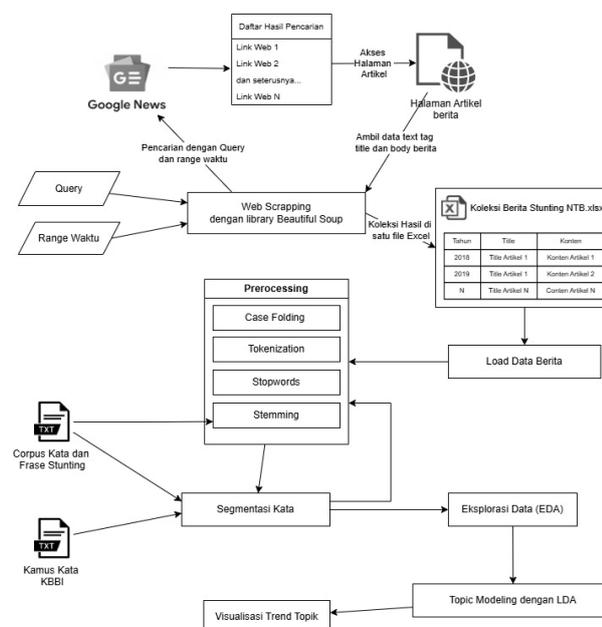
Exploratory Data Analysis (EDA) dan Latent Dirichlet Allocation (LDA) adalah metode yang sering digunakan untuk analisis tren karena kemampuan keduanya dalam menggali pola dari data yang kompleks. EDA sangat efektif dalam mengidentifikasi tren melalui eksplorasi awal data numerik dan temporal. Metode ini dapat memahami pola musiman, tren linear, atau bahkan anomali dalam data yang

dapat menjadi indikasi perubahan signifikan (Yuhefizar et al., 2024). Keterbisaannya terletak pada fleksibilitasnya untuk diterapkan pada berbagai jenis data terstruktur dan kemampuannya memberikan wawasan awal dengan cepat. Dalam tren yang melibatkan data kuantitatif, seperti angka prevalensi atau distribusi demografis, EDA menjadi alat utama untuk mengidentifikasi faktor pendorong di balik perubahan (Yuhefizar et al., 2024).

LDA juga memberikan keunggulan dalam menganalisis data teks yang besar dan tidak terstruktur, seperti artikel, laporan, atau media sosial. Dengan kemampuannya untuk mengidentifikasi tema utama dan mendeteksi perubahan narasi dari waktu ke waktu, LDA menjadi sangat bermanfaat dalam memahami perkembangan isu sosial, ekonomi, atau politik (Manuaba & Karim, 2024), (Batool & Byun, 2024). LDA memanfaatkan metode topic modeling yang mengelompokkan kata-kata ke dalam tema tertentu, sehingga memungkinkan analisis yang mendalam terhadap tren berbasis teks (Manuaba & Karim, 2024), (Sugiantoro et al., 2023). Kombinasi EDA dan LDA memberikan kemampuan yang saling melengkapi, memungkinkan analisis tren yang lebih mendalam dan menyeluruh, baik dari sisi data numerik maupun narasi publik yang relevan dengan tren tertentu (Yuhefizar et al., 2024), (Batool & Byun, 2024). Berdasarkan hal itu metode ini juga sangat cocok untuk digunakan untuk analisis trend stunting di NTB yang diusulkan pada paper ini.

Yang sangat penting untuk diperhatikan dalam analisis trend stunting selain dari istilahnya itu sendiri yaitu sebab dan dampak dari stunting tersebut. Trend tersebut perlu dilakukan analisis untuk menentukan keterkaitan mendalam bagaimana stunting itu terjadi dan bagaimana dampaknya. Paper ini lebih mengusulkan kontribusi mengenai hal tersebut yaitu bagaimana analisis trend Stunting, Sebab, dan Dampaknya. Kumpulan corpus yang berisi kata dan frase terkait ketiga hal tersebut dikumpulkan terlebih dahulu berdasarkan studi literatur yang dilakukan. Data Berita penelitian didapatkan dari hasil web scrapping dengan query yang berhubungan dengan stunting di NTB. Kemudian penerapan text mining seperti preprocessing, EDA dan LDA dilakukan untuk menganalisis trend topik sebenarnya yang terdapat pada teks tersebut. Pada akhirnya hasil analisis didapatkan berupa bagaimana istilah itu muncul dan saling berkaitan hingga rekomendasi langkah penting penekanan stunting berdasarkan penelitian ini.

## METODE PENELITIAN



Gambar 1. Metodologi penelitian analisis trend stunting NTB

## 1. WEB SCRAPING

Web scraping digunakan dalam penelitian ini untuk mengumpulkan data secara otomatis dari situs web berita yang relevan dengan topik stunting di NTB. Proses ini memanfaatkan pustaka Python Beautiful Soup untuk mengekstraksi elemen-elemen spesifik dari dokumen HTML. Teknik ini diterapkan untuk mendapatkan data berdasarkan query-query dan pengaturan lainnya (Manuaba & Karim, 2024), (Ishmael et al., 2023).

Gambar 1 merepresentasikan bagaimana kumpulan data artikel didapatkan. Pertama, query seperti yang ada di dalam Tabel 1 dimasukkan dengan pengaturan tahun dari 2018 hingga 2024 untuk pencarian dari Google dengan menggunakan library request. Beautiful soup digunakan untuk menangkap response dari hasil pencarian tersebut. Kemudian, kumpulan link artikel hasil pencarian akan didapatkan untuk diakses kembali satu per satu. Dengan Beautiful Soup, response berupa halaman HTML yang berisi artikel berita akan ditangkap. Selanjutnya, artikel tersebut diolah untuk mendapatkan data berupa judul (title) dan isi artikel (body). Pengaturan untuk beberapa halaman berita perlu dilakukan karena setiap halaman memiliki tag masing-masing untuk menandai bagian judul dan isi. Hasil akhir dari proses ini dikumpul menjadi satu file excel yang berisi kolom tahun, *title*, *text*.

Tabel 1. Query pencarian stunting pada proses web scrapping

No.	Query
1	stunting di Nusa Tenggara Barat
2	stunting di NTB
3	stunting di Kota Mataram
4	stunting di Lombok Barat
5	stunting di Lombok Tengah
6	stunting di Lombok Timur
7	stunting di Lombok Utara
8	stunting di Sumbawa
9	stunting di Sumbawa Barat
10	stunting di Dompu
11	stunting di Kabupaten Bima
12	stunting di Kota Bima

## 2. PREPROCESSING

Tahap preprocessing merupakan langkah awal yang penting dalam pengolahan data teks untuk memastikan data yang digunakan memiliki kualitas yang baik dan relevan. Pada tahap pertama ini, dilakukan serangkaian proses untuk membersihkan dan menyederhanakan teks agar sesuai untuk analisis lebih lanjut. Proses ini dirancang untuk menangani data mentah yang sering kali mengandung elemen tidak relevan, seperti simbol, angka, atau variasi kata, sehingga data yang dihasilkan lebih terstruktur dan dapat diandalkan.

Adapun langkah-langkah yang dilakukan yaitu mulai dari normalisasi teks untuk mengubah ke huruf kecil (lowercase) menghapus simbol, angka, dan karakter khusus. Setelah normalisasi, teks dipecah menjadi unit-unit kecil berupa kata-kata melalui proses tokenisasi. Dalam penelitian ini, tokenisasi dilakukan menggunakan pustaka Natural Language Toolkit (NLTK), yang memungkinkan setiap kata dalam teks dapat diproses secara individual. Langkah ini memberikan fleksibilitas untuk manipulasi lebih lanjut pada tingkat kata. Selanjutnya, dilakukan penghapusan stopword, yaitu kata-kata umum seperti "dan," "di," atau "yang" yang tidak memberikan kontribusi signifikan dalam analisis. Penghapusan dilakukan dengan menggunakan daftar stopwords bahasa Indonesia dari pustaka NLTK. Hal ini memastikan hanya kata-kata yang relevan dan memiliki makna penting yang dipertahankan dalam data.

Terakhir, proses stemming diterapkan untuk mengubah setiap kata ke bentuk dasarnya. Dengan menggunakan pustaka Sastrawi, yang dirancang khusus untuk bahasa Indonesia, variasi kata seperti "belajar" dan "belajarlaha" direduksi menjadi bentuk dasar "belajar." Proses ini penting untuk menyederhanakan data dan memastikan konsistensi dalam analisis. Proses tambahan yang penting dari proses stemming ini yaitu tidak diperkenankan untuk melakukan steaming istilah-istilah yang terdapat pada Tabel 2. Karena tabel 2 tersebut pada paper ini merupakan list yang memiliki makna terkait dengan stunting dan akan memiliki arti yang berbeda jika dimodifikasi.

### 3. SEGMENTASI KATA

Segmentasi teks merupakan langkah penting dalam analisis data teks, khususnya untuk memisahkan atau mengenali kata-kata dalam sebuah teks berdasarkan kamus yang telah ditentukan. Dalam penelitian ini, proses segmentasi diterapkan pada data teks yang telah melalui tahap preprocessing untuk menghasilkan teks yang lebih terstruktur dan relevan dengan kebutuhan analisis. Segmentasi dilakukan dengan menggunakan kamus utama dan tambahan yang memuat daftar kata-kata yang relevan.

Kamus utama yang digunakan dalam proses ini berasal dari file teks kamus\_data\_v\_1\_0\_0.txt, yang berisi corpus kata umum. Kamus ini kemudian diperluas dengan sesuai kontribusi paper ini yaitu penambahan istilah term corpus relevan, sebab, dan dampak stunting seperti pada Tabel 2. Kedua kamus tersebut digabungkan menjadi satu set kata unik untuk digunakan dalam proses segmentasi. Hal ini memastikan cakupan kata-kata yang relevan dengan konteks penelitian menjadi lebih luas.

Hasil akhir dari proses segmentasi adalah teks yang telah diproses dan dipisahkan sesuai dengan kamus yang digunakan. Setelah segmentasi dilakukan tentu teks akan menjadi tidak struktur lagi karena akan muncul sisa huruf atau angka yang tidak berarti. Oleh sebab itu reprocessing dilakukan kembali untuk membersihkan hasil segmentasi dan digunakan pada tahap selanjutnya.

Tabel 2. Corpus berisi kamus kata dan frase utama serta turunan yang berhubungan dengan Stunting

No	Kata Utama	Kata dan Frase Turunan Terkait Stunting
<b>Corpus Utama</b>		
1	Stunting	stunting, pendek, gagal tumbuh, pertumbuhan terhambat, tidak optimal, pencegahan, masalah stunting, faktor risiko, penyebab stunting, kebijakan stunting
2	gizi	nutrisi, kurang gizi, malnutrisi, defisiensi nutrisi, asupan gizi, gizi buruk, pola makan, pemenuhan gizi, kecukupan gizi, ketahanan pangan, gizi seimbang, intervensi gizi, program gizi, pemberian ASI eksklusif
3	kesehatan	akses kesehatan, layanan kesehatan, kesehatan masyarakat, pola hidup sehat, kesehatan ibu, kesehatan anak, kesehatan lingkungan, promosi kesehatan, pencegahan penyakit, pelayanan kesehatan dasar, status kesehatan, imunisasi, kesehatan reproduksi
4	anak	bayi, balita, usia dini, tumbuh kembang, perkembangan anak, anak usia dini, kesejahteraan anak, hak anak, pendidikan anak usia dini, anak sehat, anak tumbuh optimal, perlindungan anak, perkembangan kognitif anak
5	pertumbuhan	perkembangan, gangguan pertumbuhan, pertumbuhan fisik, tumbuh kembang, pertumbuhan anak, perkembangan motorik, pertumbuhan optimal, pengukuran pertumbuhan, status pertumbuhan, indikator pertumbuhan
6	tinggi	tinggi badan, indeks, standar tinggi, rasio tinggi, ukuran tinggi, grafik pertumbuhan tinggi, tinggi anak, tinggi optimal, indikator tinggi badan, tinggi ideal, tinggi proporsional

No	Kata Utama	Kata dan Frase Turunan Terkait Stunting
7	pendek	tubuh pendek, anak pendek, prevalensi pendek, indikator pendek, risiko pendek, pertumbuhan pendek, anak dengan tubuh pendek, masalah tubuh pendek, tubuh pendek kronis, kondisi pendek akibat gizi buruk
8	malnutrisi	kurang nutrisi, kekurangan gizi, malas gizi, gizi kurang kronis, penyebab malnutrisi, dampak malnutrisi
9	prevalensi	prevalensi, angka kejadian, frekuensi, kasus stunting, penyebaran stunting, angka prevalensi
10	intervensi	upaya pencegahan, strategi mitigasi, program kesehatan, penanganan stunting, intervensi berbasis masyarakat, intervensi spesifik, intervensi sensitif
<b>Corpus Sebab</b>		
1	air bersih	air bersih, ketersediaan air bersih, akses air bersih terbatas, kualitas air rendah, pencemaran air, ketersediaan air minum, air tidak layak konsumsi
2	akses kesehatan	akses kesehatan, fasilitas kesehatan, layanan kesehatan, akses kesehatan terbatas, biaya kesehatan tinggi, kurangnya tenaga medis, infrastruktur kesehatan yang buruk, lokasi fasilitas kesehatan yang jauh
3	gizi buruk	gizi buruk, kurang gizi, malnutrisi, kekurangan nutrisi, asupan gizi rendah, gizi tidak seimbang, defisiensi gizi, kelaparan kronis, kekurangan kalori
4	hukum_norma	pernikahan anak, perkawinan anak, pelanggaran usia pernikahan, batas usia menikah, undang-undang pernikahan, dispensasi menikah, kawin kontrak
5	kemiskinan	kemiskinan, miskin, ekonomi lemah, pendapatan rendah, pengangguran, ketidakstabilan ekonomi, kemiskinan ekstrem
6	ketahanan pangan	ketahanan pangan, kelangkaan pangan, akses pangan terbatas, produksi pangan rendah, harga pangan tinggi, ketergantungan pada impor pangan, distribusi pangan tidak merata
7	nikah_muda	nikah muda, pernikahan dini, kawin muda, pernikahan remaja, kawin remaja
8	pendidikan	pendidikan rendah, pendidikan ibu, kurangnya edukasi, tingkat literasi rendah, kesadaran akan gizi, pendidikan kesehatan, pengetahuan tentang pola makan sehat
9	perubahan iklim	perubahan iklim, cuaca ekstrem, bencana alam, kerawanan pangan akibat iklim, curah hujan rendah, kekeringan, banjir, pola cuaca tidak menentu
10	sanitasi	sanitasi, sanitasi buruk, akses sanitasi, lingkungan tidak sehat, air kotor, limbah tidak terkelola, kekurangan fasilitas sanitasi, kondisi sanitasi tidak memadai
11	usia	usia muda, usia dini, remaja awal, usia sekolah, anak usia dini, di bawah umur, belum cukup umur, tidak dewasa secara hukum

No	Kata Utama	Kata dan Frase Turunan Terkait Stunting
<b>Corpus Dampak</b>		
1	pertumbuhan	pertumbuhan, pertumbuhan terhambat, gangguan tumbuh, perkembangan fisik, pertumbuhan tidak optimal, pertumbuhan anak terganggu, indikator pertumbuhan rendah
2	pendidikan	pendidikan, rendahnya pendidikan, putus sekolah, tingkat literasi rendah, kesempatan pendidikan terbatas, prestasi akademik rendah, akses pendidikan terganggu
3	ekonomi	ekonomi, ekonomi lemah, daya saing menurun, penghasilan rendah, produktivitas rendah, ketergantungan ekonomi, kerugian ekonomi jangka panjang
4	kesehatan anak	kesehatan anak, imunisasi rendah, penyakit anak, daya tahan tubuh rendah, risiko penyakit menular, masalah kesehatan kronis, kematian anak, malnutrisi pada anak
5	kesehatan masyarakat	kesehatan masyarakat, penurunan kualitas kesehatan, peningkatan beban kesehatan, penyebaran penyakit, akses layanan kesehatan terganggu
6	psikososial	psikososial, stres keluarga, tekanan psikologis, kesejahteraan mental rendah, dampak emosional pada anak, hubungan sosial terganggu
7	produktivitas	produktivitas rendah, kinerja buruk, kemampuan kerja terganggu, efisiensi berkurang, kontribusi ekonomi menurun

#### 4. EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) adalah langkah penting dalam penelitian ini untuk memahami karakteristik awal data yang digunakan. EDA bertujuan untuk mengeksplorasi pola, hubungan, dan distribusi data, serta mengidentifikasi potensi anomali atau data yang tidak konsisten. Dengan pendekatan ini, data dapat dianalisis secara menyeluruh sebelum memasuki tahap pemrosesan atau analisis lebih lanjut.

Pada penelitian ini, EDA dilakukan pada dataset yang telah melalui tahap preprocessing dan segmentasi teks. Proses ini mencakup beberapa langkah utama. Langkah yang dilakukan yaitu menghitung distribusi frekuensi kata dalam dataset untuk mengidentifikasi kata-kata yang paling sering muncul, baik dari kategori Corpus Utama, Sebab, maupun Dampak. Frekuensi kata ini memberikan gambaran awal tentang tema atau isu utama yang mendominasi dataset. Data teks divisualisasikan menggunakan diagram batang dan tabel untuk menampilkan kata-kata yang paling sering muncul secara visual. Teknik ini membantu dalam memahami pola distribusi kata yang relevan dengan topik penelitian, seperti kata-kata yang berkaitan dengan stunting, gizi, atau pertumbuhan.

#### 5. TOPIK MODELLING dengan LDA

Hasil dari EDA memberikan wawasan awal yang esensial untuk menentukan arah sebaran istilah kata yang muncul. Namun bisal langsung dikatakan sebagai bentuk trend topik yang terjadi. Topik ternd sebenarnya yang terjadi dapat menggunakan metode Latent Dirichlet Allocation (LDA) yang akan menunjukkan bagaimana sebaran dan relevansi istilah dari topik-topik yang dibahas.

LDA adalah metode topic modeling yang digunakan dalam penelitian ini untuk mengeksplorasi tema-tema utama dalam data teks yang berkaitan dengan isu stunting di Nusa Tenggara Barat (NTB).

Metode ini bertujuan untuk mengidentifikasi tren sebaran topik stunting serta relevan term yang terjadi berdasarkan text yang diberikan. Proses penerapan LDA diawali dengan pembentukan korpus teks dari data yang telah melalui tahap preprocessing dan segmentasi. Korpus ini berisi kumpulan teks yang relevan dengan isu stunting, seperti laporan, berita, atau diskusi publik yang mencakup periode waktu antara tahun 2018 hingga 2024. Data ini kemudian diproses menjadi representasi numerik menggunakan pendekatan bag-of-words atau term frequency-inverse document frequency (TF-IDF), sehingga dapat diolah oleh model LDA.

Tujuan utama LDA dalam penelitian ini adalah untuk mengidentifikasi tema-tema yang mendominasi pembahasan stunting di NTB. Tema yang dihasilkan oleh model dikategorikan berdasarkan relevansinya dengan Corpus Utama, Sebab, dan Dampak yang telah dirumuskan sebelumnya (lihat Tabel 2). Misalnya, topik yang terkait dengan "kemiskinan," "akses sanitasi," atau "gizi buruk" akan diidentifikasi sebagai tema penyebab, sedangkan topik seperti "pertumbuhan terhambat" atau "pendidikan rendah" akan diklasifikasikan sebagai tema dampak. Setelah model LDA dilatih, analisis dilakukan untuk mengidentifikasi tren stunting di NTB selama periode waktu yang diteliti. Analisis temporal memungkinkan pemahaman tentang bagaimana tema-tema tertentu berkembang atau berubah dalam rentang waktu tersebut. Misalnya, model dapat mengungkapkan peningkatan pembahasan terkait kebijakan intervensi stunting pada tahun tertentu atau mengidentifikasi tren peningkatan fokus pada isu gizi dan kesehatan anak.

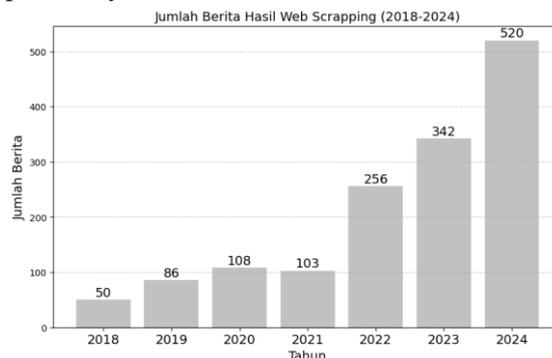
Hasil pemodelan LDA juga divisualisasikan menggunakan alat seperti pyLDAvis untuk mempermudah interpretasi topik dan memahami distribusi kata dalam setiap topik. Visualisasi ini membantu mengungkap hubungan antar topik, dominasi tema tertentu, serta keterkaitan antara tema penyebab dan dampak dalam data stunting di NTB. Dengan menggunakan LDA, penelitian ini memberikan wawasan yang mendalam tentang tren stunting di NTB. Temuan dari analisis ini diharapkan dapat mendukung pengambilan keputusan yang lebih tepat sasaran, baik dalam merancang kebijakan intervensi maupun dalam menyusun program pencegahan stunting yang lebih efektif di wilayah tersebut.

## HASIL DAN PEMBAHASAN

Implementasi penelitian analisis trend kasus stunting di NTB ini dilakukan menggunakan perangkat komputer dengan Spesifikasi prosesor CPU Intel(R) Core(TM) i5-6300U CPU @ 2.40GHz 2.50 GHz RAM 8 GB. Seluruh proses dilakukan menggunakan program Visual Studio Code dengan Jupyter-Lab yang terinstal didalamnya.

### 1. KOLEKSI DATA WEB SCRAPPING

Berdasarkan hasil web scrapping yang dilakukan dengan query yang dibuat, maka didapatkan 1.465 total berita tentang stunting di NTB dari tahun 2018 hingga 2024. Gambaran sebaran jumlah berita dari setiap tahunnya seperti pada Gambar 2. Berdasarkan diagram secara umum pemberitaan stunting di NTB terjadi peningkatan setiap tahunnya.



**Gambar 2.** Sebaran jumlah berita stunting NTB dari tahun 2018-2024

Gambaran sebaran jumlah berita dari setiap tahunnya seperti pada Gambar 2. Berdasarkan diagram secara umum pemberitaan stunting di NTB terjadi peningkatan setiap tahunnya.

## 2. HASIL PREPROCESSING DAN SEGMENTASI

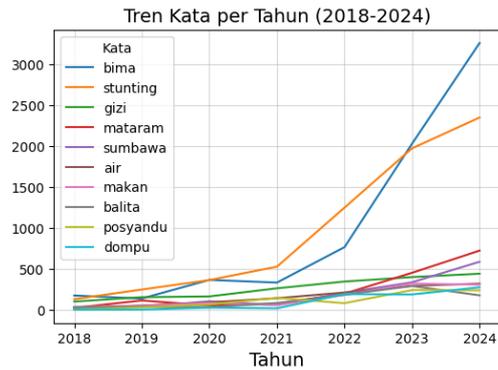
Berdasarkan hasil web scrapping yang dilakukan dengan query yang dibuat, maka didapatkan 1.465 total berita tentang stunting di NTB dari tahun 2018 hingga 2024.

**Tabel 3. Hasil Preprocessing dan Segmentasi Berita**

<b>Sample Original Text</b>
Kompas.idKasus Kurang Gizi di Nusa Tenggara Barat TinggiMATARAM, KOMPAS—Angka kasus anak berusia di bawah lima tahun atau balita yang mengalami kurang gizi dan stunting atau tubuh pendek masih tinggi di Nusa....May 26, 2018 Para ibu, warga Lingkungan Karang Seme, KelurahanKarang Pule. Mataram, Nusa Tenggara Barat, antre menimbang berat badan bayinya, sebagai upaya inventarisasi dan menjangring penderita gizi buruk.
<b>Preprocessing tahap 1</b>
kompasidkasus gizi nusa tenggara barat tinggimataran kompasangka anak usia balita alami gizi stunting tubuh pendek nusamay warga lingkung karang seme kelurahankarang pule mataram nusa tenggara barat antre timbang berat badan bayi upaya inventarisasi jaring derita gizi buruk
<b>Segmentasi Kata</b>
kompas id kasus gizi nusa tenggara barat tinggi mataram kompas angka anak usia balita alami gizi stunting tubuh pendek nusa ma y warga lingkung karang se me kelurahan karang pul e mataram nusa tenggara barat antre timbang berat badan bayi upaya inventarisasi jaring derita gizi buruk
<b>Preprocessing Tahap 2</b>
kompas gizi nusa tenggara barat mataram kompas angka anak usia balita alami gizi stunting tubuh pendek nusa warga lingkung karang kelurahan karang pul mataram nusa tenggara barat antre timbang berat badan bayi upaya inventarisasi jaring derita gizi buruk

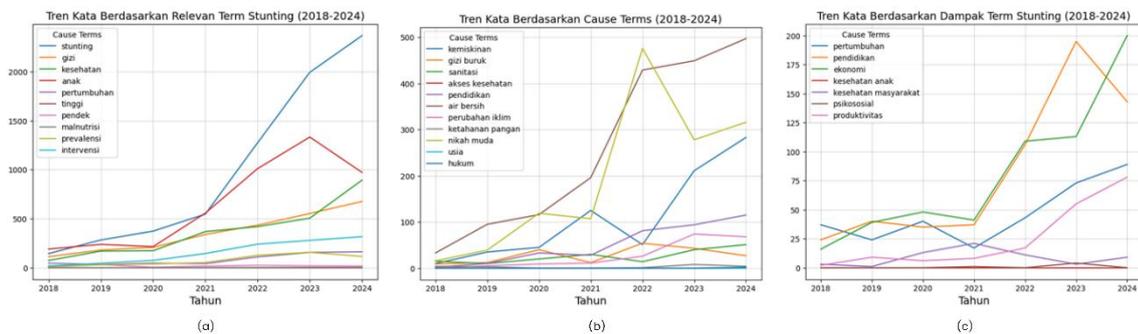
## 3. HASIL EDA

Exploratory Data Analysis (EDA) adalah langkah penting dalam penelitian ini untuk memahami karakteristik awal data yang digunakan. EDA bertujuan untuk mengeksplorasi pola, hubungan, dan distribusi data, serta mengidentifikasi potensi anomali atau data yang tidak konsisten. Dengan pendekatan ini, data dapat dianalisis secara menyeluruh sebelum memasuki tahap pemrosesan atau analisis lebih lanjut.



**Gambar 3.** Tren kata utama berita stunting

Berdasarkan analisis EDA pada Gambar 3 terhadap frekuensi kata kunci dalam pemberitaan stunting di NTB dari tahun 2018 hingga 2024, teridentifikasi tren peningkatan signifikan pada sejumlah istilah kritis. Kata “bima” menempati posisi tertinggi dengan peningkatan drastis dari 179 kemunculan (2018) menjadi 3.260 (2024), diikuti oleh kata “stunting” yang meningkat dari 135 menjadi 2.351 kemunculan. Kedua kata ini menunjukkan dominasi dalam diskursus pemberitaan, terutama pada periode 2022–2024. Selain itu, istilah terkait lokasi seperti “mataram” dan “sumbawa” serta aspek penanganan seperti “gizi”, “air”, dan “posyandu” juga mengalami kenaikan frekuensi, meskipun dengan pola yang lebih fluktuatif. Misalnya, kata “posyandu” mengalami puncak pada tahun 2021, kemudian menurun sebelum kembali naik pada 2023–2024, mengindikasikan fokus periodik pada program kesehatan masyarakat. Sementara itu, kata “dompu” dan “balita” menunjukkan lonjakan tertunda, dengan peningkatan tertinggi terjadi pada 2022–2023, kemungkinan terkait intensifikasi pelaporan kasus stunting di wilayah tersebut. Meskipun frekuensi kata menggambarkan intensitas pemberitaan, hasil ini belum secara langsung membuktikan korelasi dengan kasus stunting di lapangan, melainkan merefleksikan dinamika komunikasi media dalam mengangkat isu tersebut.



**Gambar 4.** Tren kata berdasarkan, (a) relevansi, (b) penyebab, (c) dampak terhadap kata stunting

Berdasarkan Gambar 4 (a), analisis tren kata yang secara spesifik relevan dengan kasus stunting di NTB menunjukkan pola pemberitaan yang erat terkait dengan determinan kesehatan dan intervensi. Kata kunci “stunting” mengalami peningkatan signifikan kemunculannya pada 2018-2024, mencerminkan intensifikasi diskusi media tentang isu ini seiring waktu. Istilah penyebab utama seperti “gizi” dan “kesehatan” juga mengalami kenaikan bertahap. Hal ini mengindikasikan fokus media pada faktor nutrisi dan layanan kesehatan sebagai akar masalah stunting. Namun pola unik terlihat pada kata “tinggi” yang tidak muncul sama sekali, mengisyaratkan kurangnya framing solutif atau pencapaian positif dalam pemberitaan. Sebaliknya, istilah “pendek” (stunting) hanya muncul secara sporadis, dengan frekuensi tertinggi 31 kemunculan (2019).

Berdasarkan Gambar 4 (b) dapat dilihat hasil eksplorasi data mengungkapkan pemberitaan stunting di NTB lebih terfokus pada faktor socio-ekonomi dan lingkungan. Kata “kemiskinan” meningkat signifikan menegaskan kemiskinan sebagai akar masalah. “Nikah muda” melonjak pada

tahun 2022, merefleksikan kepedulian media terhadap perkawinan dini sebagai pemicu stunting. “Air bersih” juga konsisten naik, mengaitkan sanitasi dengan risiko stunting. Faktor “gizi buruk” dan “sanitasi” fluktuatif, dengan penurunan drastis “gizi buruk” pasca-2020. Istilah seperti “akses kesehatan”, “usia”, dan “hukum” minim muncul, mengindikasikan kurangnya pembahasan disparitas layanan dan aspek legal. “perubahan iklim” mulai menonjol pada 2023, diduga terkait dampak bencana iklim terhadap ketahanan pangan.

Berdasarkan Gambar 4 (c) menunjukkan bahwa pemberitaan tentang dampak stunting di NTB lebih banyak menyoroti aspek pertumbuhan, pendidikan, dan ekonomi. Kata “pertumbuhan” mengalami peningkatan konsisten, mencerminkan fokus media pada dampak fisik stunting terhadap perkembangan anak. Kata “pendidikan” juga menonjol dengan lonjakan signifikan, mengindikasikan kesadaran media terhadap keterkaitan stunting dengan penurunan kualitas pendidikan. Sementara itu, “ekonomi” meningkat pesat, menegaskan dampak stunting terhadap produktivitas dan kesejahteraan masyarakat. Di sisi lain, istilah seperti “kesehatan anak” dan “psikososial” hampir tidak muncul, menunjukkan kurangnya pembahasan tentang dampak kesehatan mental dan perkembangan psikologis anak. “Kesehatan masyarakat” juga minim terekspos, sementara “produktivitas” mulai mendapat perhatian, diduga terkait narasi tentang dampak jangka panjang stunting terhadap tenaga kerja. Temuan ini menggarisbawahi dominasi pemberitaan pada dampak fisik dan ekonomi, sementara aspek kesehatan mental dan sosial masih kurang mendapat perhatian.

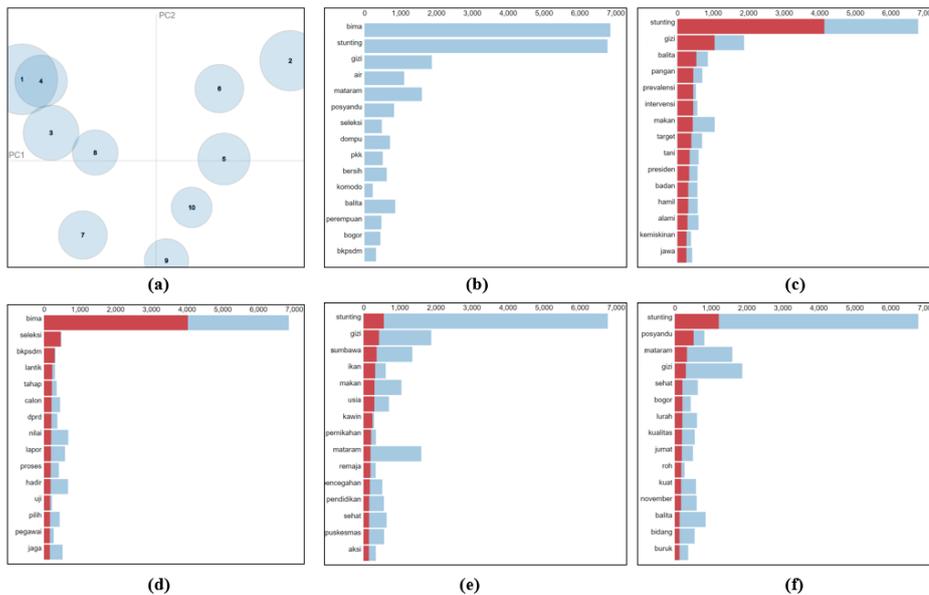
#### 4. HASIL LDA

Pada bagian sebelumnya, telah dilakukan EDA untuk menganalisis sebaran kata dalam dataset tren stunting di NTB, dengan menampilkan istilah-istilah yang paling sering muncul. Namun, hasil tersebut belum merepresentasikan tren kasus secara akurat karena hanya berbasis frekuensi kata. Oleh karena itu, analisis LDA diperlukan untuk mengidentifikasi pola statistik, sebagaimana dijelaskan dalam Persamaan 1 dan 2. LDA bekerja dengan mengelompokkan teks ke dalam beberapa topik berdasarkan distribusi kata secara probabilistik, sehingga dapat mengungkap pola tersembunyi dan menghasilkan topik-topik utama yang lebih representatif terkait isu stunting.

Gambar 5 merupakan reprints bagaimana data yang digunakan pada paper ini dikelompokkan menjadi 10 topik menggunakan LDA. Hasil analisis LDA yang disajikan pada Gambar (a) dan (b) memberikan pandangan mendalam terkait distribusi dan dominasi istilah dalam pembahasan stunting di NTB. Pada Gambar (b), dominasi istilah dalam pembahasan lebih jelas terlihat. Istilah seperti Bima, Stunting, dan Gizi menjadi yang paling sering muncul dalam corpus. Bima, sebagai salah satu istilah geografis, memiliki frekuensi yang tinggi karena sering menjadi fokus pemberitaan lokal. Stunting, di sisi lain, muncul sebagai istilah dominan karena isu ini menjadi perhatian utama dalam narasi berita yang dianalisis. Istilah Gizi juga menonjol, menyoroti kaitan erat antara stunting dan asupan gizi yang memadai. Selain itu, istilah seperti air dan posyandu muncul dalam frekuensi tinggi, mencerminkan fokus pada kebutuhan dasar dan layanan kesehatan komunitas dalam upaya pencegahan stunting.

Namun, meskipun istilah Bima mendominasi dalam hal frekuensi, seperti yang ditunjukkan pada Gambar (d), analisis LDA menunjukkan bahwa istilah ini memiliki korelasi yang lemah dengan istilah lain yang relevan dengan stunting, seperti gizi, air, atau intervensi. Hal ini mengindikasikan bahwa istilah Bima lebih sering muncul dalam konteks administratif atau lokasi geografis, dan tidak terkait langsung dengan diskusi utama mengenai penyebab atau dampak stunting. Gambar (d) memperjelas bahwa meskipun Bima sering muncul pada analisis eksplorasi data, istilah ini tidak memiliki hubungan kuat dengan topik yang relevan dengan stunting, seperti yang ditunjukkan oleh korelasi antar-istilah dalam topik LDA.

Sebaliknya, Gambar (c), (e), dan (f) memberikan gambaran relevansi topik yang memiliki korelasi erat dengan stunting di NTB. Pada Gambar (c), istilah seperti stunting, gizi, pangan, dan intervensi menunjukkan fokus utama pada penyebab dan upaya pencegahan stunting, dengan perhatian khusus pada ketahanan pangan, pengelolaan gizi, dan intervensi berbasis komunitas. Gambar (e) menyoroti hubungan antara isu sosial dan budaya dengan stunting, di mana istilah seperti nikah muda, pendidikan, dan remaja menggarisbawahi bagaimana rendahnya tingkat pendidikan dan praktik pernikahan dini dapat berkontribusi terhadap peningkatan risiko stunting. Sementara itu, Gambar (f) lebih menekankan aspek kesehatan masyarakat, dengan istilah seperti posyandu, sehat, dan kualitas mencerminkan peran penting layanan kesehatan berbasis komunitas dalam pencegahan stunting, khususnya di wilayah perkotaan seperti Mataram. Ketiga gambar ini secara keseluruhan menunjukkan bahwa pembahasan stunting di NTB melibatkan berbagai aspek kesehatan, sosial, dan budaya yang saling terkait, menegaskan pentingnya pendekatan multidimensi dalam menangani masalah ini.



## KESIMPULAN

Stunting merupakan permasalahan gizi kronis yang berdampak pada pertumbuhan fisik dan perkembangan kognitif anak, dengan prevalensi yang tinggi di Indonesia, khususnya di Nusa Tenggara Barat (NTB). Pemberitaan di media masa internet sudah banyak dipublikasi dari tahun 2018 sampai 2024. Analisis dilakukan untuk mendapatkan gambaran term istilah yang berkaitan untuk mendukung strategi komunikasi kebijakan, dengan menggunakan bahasa yang lebih efektif dan sesuai dengan isu yang sedang hangat dibicarakan. Setelah dilakukan proses scrapping dan preprocessing, Segmentasi dilakukan dengan kamus data dan kumpulan corpus kata utama, sebab, dan dampak terkait stunting. Kemudian Exploratory Data Analysis (EDA) dilakukan untuk memberikan gambaran frekuensi term yang sering muncul pada data berita yang digunakan. Bima, Stunting, Gizi, Mataram, dan Sumbawa merupakan trend istilah yang sering muncul pada tahapan ini. Namun, istilah tersebut belum dapat dikatakan sebagai bentuk trend Stunting yang dibicarakan pada data berita karena prinsip EDA hanya menganalisis frekuensi kemunculan. Selanjutnya LDA dilakukan guna mengelompokkan kumpulan istilah kedalam beberapa topik sehingga didapatkan istilah yang dibahas beserta relevansi dengan istilah lainnya.

Hasil analisis tren stunting di NTB menunjukkan bahwa istilah stunting, gizi, dan air bersih mendominasi pemberitaan, menyoroti fokus pada faktor kesehatan seperti asupan gizi dan akses air bersih sebagai isu utama dalam pencegahan stunting. Selain itu, faktor sosial seperti nikah muda dan pendidikan juga muncul sebagai penyebab signifikan, menunjukkan pentingnya pendekatan holistik yang mencakup aspek sosial dan budaya. Intervensi berbasis komunitas, seperti posyandu dan program

kesehatan, menjadi sorotan penting, mencerminkan peran layanan kesehatan lokal dalam mendukung upaya pencegahan stunting. Namun, analisis juga mengungkap bahwa istilah seperti Bima, meskipun sering muncul, lebih terkait dengan konteks geografis dan administratif, tanpa hubungan langsung dengan isu stunting. Keseluruhan hasil menegaskan bahwa pemberitaan tentang stunting di NTB mencerminkan kompleksitas isu ini, yang melibatkan faktor kesehatan, sosial, dan budaya. Untuk menekan prevalensi stunting, diperlukan pendekatan terintegrasi yang tidak hanya fokus pada intervensi kesehatan, tetapi juga mengatasi penyebab sosial, dengan strategi komunikasi yang tepat sasaran untuk meningkatkan kesadaran masyarakat.

#### UCAPAN TERIMAKASIH

Penelitian ini didanai oleh Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat, Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia pada program Bima Skema Penelitian Dosen Pemula (PDP-Afirmasi) tahun 2024.

#### DAFTAR PUSTAKA

- Aggarwal, C. C., & Aggarwal, C. C. (2015). Mining text data. Springer.
- Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics*, 24(1), 1–7.
- Apriani, L. A., Ismiati, I., Apriani, N., & Malika, R. (2024). Gambaran Pengetahuan dan Sikap Ibu Tentang Stunting pada Balita Usia 24-59 Bulan. *Jurnal Kesehatan Qamarul Huda*, 12(1), 69–74.
- Batool, A., & Byun, Y. C. (2024). Enhanced Sentiment Analysis and Topic Modeling During the Pandemic Using Automated Latent Dirichlet Allocation. *IEEE Access*, 12, 81206–81220. <https://doi.org/10.1109/ACCESS.2024.3411717>
- Fanni, S. C., Febi, M., Aghakhanya, G., & Emanuele Neri. (2023). Natural language processing. In *Introduction to Artificial Intelligence*. Cham: Springer International Publishing, 87–99.
- Francis, L. A. (2006). Taming text: An introduction to text mining. *Casualty Actuarial Society Forum*, 51–88.
- Greven, A., Kruse, S., Vos, A., Strese, S., & Brettel, M. (2023). Achieving product ambidexterity in new product development: The role of middle managers' dynamic managerial capabilities. *Journal of Management Studies*, 60(7), 1786–1818.
- Hamdi, S., Elmiana, D. S., Maulana, I., Haromain, N., Rahmawadi, I., Malik, F. A., Mataram, U., Barat, N. T., Tengah, L., Desa, P., & Sosial, M. (2023). STRATEGI PEMERINTAH DESA DALAM PENANGANAN STUNTING BERBASIS MODAL SOSIAL DI DESA BILEBANTE , LOMBOK TENGAH VILLAGE GOVERNMENT STRATEGY IN HANDLING SOCIAL CAPITAL BASED STUNTING IN BILEBANTE VILLAGE , CENTRAL LOMBOK. 18, 209–224. <https://doi.org/10.47441/jkp.v18i2.346>
- Hung, J.-L., & Zhang, K. (2012). Examining mobile learning trends 2003--2008: A categorical meta-trend analysis using text mining techniques. *Journal of Computing in Higher education*, 24, 1–17.
- Irwansyah, I., Ismail, D., & Hakimi, M. (n.d.). Kehamilan remaja dan kejadian stunting anak usia 6 – 23 bulan di Lombok Barat.
- Ishmael, O., Kiely, E., Quigley, C., & McGinty, D. (2023). Topic modelling using latent dirichlet allocation (LDA) and analysis of students sentiments. 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE), 1–6.
- Kim, M.-J., Ohk, K., Moon, C.-S., & others. (2017). Trend analysis by using text mining of journal articles regarding consumer policy. *New Physics: Sae Mulli*, 67(5), 555–561.
- Madani Pertiwi, A. N. A., & Hendrati, L. Y. (2023). Literature Review: Analysis of the Causes of

Stunting in Toddlers in East Java Province. *Amerta Nutrition*, 7.

- Manuaba, I. B. K., & Karim, M. F. (2024). Latent Dirichlet Allocation (LDA) Topic Modeling and Sentiment Analysis for Myanmar Coup Tweets. *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 161–166.
- Mustika, W., & Syamsul, D. (2018). Analisis permasalahan status gizi kurang pada balita di puskesmas teupah selatan kabupaten simeuleu. *Jurnal Kesehatan Global*, 1(3), 127.
- Ojo, A. K., & Adeyemo, A. B. (2013). Knowledge discovery in academic electronic resources using text mining. *International Journal of Computer Science and Information Security*, 11(2), 10–19.
- Sudigyo, D., Hidayat, A. A., Nirwantono, R., Rahutomo, R., Trinugroho, J. P., & Pardamean, B. (2023). Literature study of stunting supplementation in Indonesian utilizing text mining approach. *Procedia Comput. Sci*, 216(2022), 722–729.
- Sugiantoro, B., Humam, A. I., Fitriyani, N. L., Alfian, G., Maarif, M. R., & Syafrudin, M. (2023). Utilizing Latent Dirichlet Allocation for Analyzing Topics in Undergraduate Theses. *2023 10th International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2023*, 121–126. <https://doi.org/10.1109/ICITACEE58587.2023.10277046>
- Wardani, Z., Sukandar, D., Baliwati, Y. F., Riyadi, H., & others. (2021). Sebuah alternatif: indeks stunting sebagai evaluasi kebijakan intervensi balita stunting di Indonesia. *Gizi indonesia*, 44(1), 21–30.
- Yuhefizar, Y., Watrianthos, R., & Komalasari, R. (2024). An LDA Analysis for Topic Modeling of the RESTI Journal. *2024 2nd International Symposium on Information Technology and Digital Innovation (ISITDI)*, July.