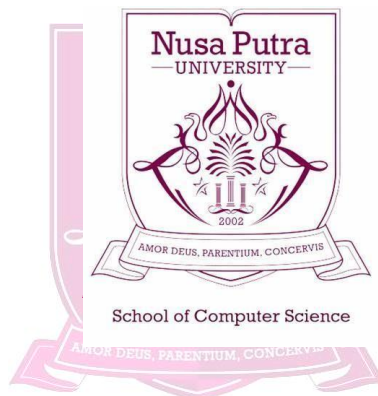


**Pemodelan Topik Pemilu 2024 Menggunakan Metode BERTopic  
pada Artikel Berita Detik.com**

**SKRIPSI**

DINI ARYANI

20200040030



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK, KOMPUTER DAN DESAIN  
SUKABUMI  
JUNI 2024**

**Pemodelan Topik Pemilu 2024 Menggunakan Metode BERTopic  
pada Artikel Berita Detik.com**

**SKRIPSI**

*Diajukan Untuk Memenuhi Salah Satu Syarat Dalam Menempuh  
Gelar Sarjana Komputer*

Dini Aryani

20200040030



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK, KOMPUTER DAN DESAIN  
SUKABUMI  
JUNI 2024**

## PERNYATAAN PENULIS

**JUDUL** : Pemodelan Topik Pemilu 2024 Menggunakan Metode BERTopic  
pada Artikel Berita Detik.com  
**NAMA** : Dini Aryani  
**NIM** : 20200040030

"Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut"

Sukabumi, 21 Juni 2024

  
  
**Dini Aryani**

Penulis


## PENGESAHAN SKRIPSI

Judul : PEMODELAN TOPIK PEMILU 2024 MENGGUNAKAN  
METODE BERTOPIC PADA ARTIKEL BERITA DETIK.COM  
Nama : DINI ARYANI  
NIM : 20200040030

Skripsi ini telah diujikan dan dipertahankan didepan Dewan Penguji pada Sidang Skripsi tanggal 21 Juni 2024. Menurut pandangan kami, Skripsi ini memadai dari segi kualitas untuk tujuan penganugrahan gelar Sarjana Komputer (S.Kom).

Sukabumi, 21 Juni 2024

Pembimbing I



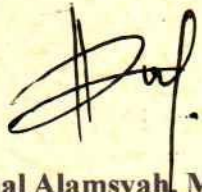
Ivana Lucia Kharisma, M.Kom  
NIDN. 0429038002

Pembimbing II



Alun Sujjada, S.Kom., M.T  
NIDN. 0718108001

Ketua Penguji



Zaenal Alamsyah, M.Kom  
NIDN. 0409069602

Ketua Program Studi Teknik Informatika



Ir. Somantri S.T., M.Kom  
NIDN. 0419128801

Plh. Dekan Fakultas Teknik, Komputer dan Desain

Ir. Paikun, S.T., M.T., IPM., ASEAN Eng  
NIDN. 0402037401



***Skripsi ini ku tujukan kepada:***

***Ayah dan Ibu tercinta  
Kakak dan keluarga besar serta kerabat  
Universitas Nusa Putra Sukabumi  
Rektor dan Wakil Rektor Universitas Nusa Putra  
Seluruh Civitas Akademik Universitas Nusa Putra  
Seluruh Jajaran dan Dosen Prodi Teknik Informatika beserta jajarannya  
Ibu Ivana Lucia Kharisma, M.Kom  
Bapak Alun Sujjada, S.Kom., M.T  
Pujaan Hati  
Teman-teman yang saat ini sedang berjuang menempuh gelar sarjana***

.....

## ABSTRAK

Pada tahun 2024, Indonesia mengadakan pemilihan umum serentak yang didominasi oleh partisipasi generasi muda, terutama generasi Z dan milenial. Generasi ini lebih sering mencari informasi politik melalui internet, menunjukkan peran penting media digital dalam membentuk opini publik. Penelitian menunjukkan bahwa paparan media digital yang positif dapat meningkatkan dukungan dan partisipasi politik generasi milenial. Penelitian ini menggunakan teknik *Topic Modelling* untuk menganalisis topik-topik terkait pemilu 2024 dari artikel berita Kompas.com, dengan metode BERTopic. *Dataset* berasal dari portal berita Detik.com pada sub kanal pemilu, diambil dengan metode *scraping* dari 1 September 2023 hingga 14 Februari 2024, total 15.019 artikel. Proses *text preprocessing* melibatkan empat tahap: *case folding*, *cleaning*, *tokenizing*, dan *stopword removal*. Pemodelan topik menggunakan BERTopic melalui lima tahapan: *embeddings*, *dimensionality reduction*, *clustering*, *tokenizer*, dan *weighting scheme*. Hasil pemodelan menunjukkan nilai *topic coherence* sebesar 0.0885752. Penelitian ini berhasil mengidentifikasi lima topik utama pada pemberitaan pemilu 2024 di portal berita Detik.com. Topik-topik tersebut mencakup: Topic 0, berfokus pada calon-calon presiden dan wakil presiden yang mencerminkan tingginya perhatian terhadap figur-figur utama dalam pemilu, dengan jumlah 6211 berita dengan representasi kata 'ganjar', 'prabowo', 'anies', 'imin'; Topic 1, berfokus pada nomor urut pasangan capres-cawapres dalam pemilihan umum di Indonesia, terutama menyebutkan tokoh seperti Gibran Rakabuming Raka, dengan jumlah artikel 2661 artikel berita dengan representasi kata 'nomor', 'urut', 'gibran', 'indonesia'; Topic 2, berita tentang Presiden Joko Widodo serta dinamika politiknya, dengan jumlah 2615 artikel berita dengan representasi kata 'presiden', 'politik', 'jokowi', 'survei' ; Topic 3, berkaitan dengan pemilihan umum dan pemilihan presiden, termasuk suara dan survei terkait 2349 artikel berita dengan representasi kata '2024', 'pemilu', 'pilpres', 'kpu'; Topic 4, tentang debat capres dan cawapres yang terjadi, dengan jumlah 1183 artikel berita 'debat', 'pilpres', 'cawapres', '2024'.

Kata Kunci: *Topic modelling*, BERTopic, pemilu 2024, artikel berita, Detik.com

## ABSTRACT

*In 2024, Indonesia held simultaneous general elections dominated by the participation of young generations, especially Generation Z and millennials. This generation more often seeks political information through the internet, highlighting the important role of digital media in shaping public opinion. Research shows that positive exposure to digital media can increase support and political participation among the millennial generation. This study uses Topic Modelling techniques to analyze topics related to the 2024 election from news articles on Kompas.com, using the BERTopic method. The dataset comes from the Detik.com news portal in the "pemilu" sub-channel, collected through scraping from September 1, 2023, to February 14, 2024, totaling 15,019 articles. The text preprocessing process involves four stages: case folding, cleaning, tokenizing, and stopword removal. Topic modeling uses BERTopic through five stages: embeddings, dimensionality reduction, clustering, tokenizer, and weighting scheme. The modeling results show a topic coherence value of 0.0885752. This study successfully identified five main topics in the news coverage of the 2024 "pemilu" on the Detik.com news portal. These topics include: Topic 0, focusing on "calon-calon presiden" and "wakil presiden" reflecting high attention to key figures in the "pemilu", with a total of 6211 news articles with representative words 'ganjar', 'prabowo', 'anies', 'imin'; Topic 1, focusing on "nomor urut" of "pasangan capres-cawapres" in the general election in Indonesia, particularly mentioning figures such as "Gibran Rakabuming Raka", with a total of 2661 news articles with representative words 'nomor', 'urut', 'gibran', 'indonesia'; Topic 2, about news on "Presiden Joko Widodo" and his political dynamics, with a total of 2615 news articles with representative words 'presiden', 'politik', 'jokowi', 'survei'; Topic 3, related to the general election and "pemilihan presiden", including votes and related surveys, with 2349 news articles with representative words '2024', 'pemilu', 'pilpres', 'kpu'; Topic 4, about news on "debat capres" and "cawapres" that occurred, with a total of 1183 news articles with representative words 'debat', 'pilpres', 'cawapres', '2024'.*

*Keywords: Topic modelling, BERTopic, 2024 election, news, Detik.com*

## KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Allah SWT, berkat rahmat dan karunia-Nya akhirnya penulis dapat menyelesaikan skripsi ini dengan judul " Pemodelan Topik Pemilu 2024 Menggunakan Metode BERTopic pada Artikel Berita Detik.com " tepat pada waktunya. Tujuan penulisan skripsi ini adalah untuk memenuhi salah satu syarat memperoleh gelar Sarjana pada Fakultas Teknik, komputer dan Desain Universitas Nusa Putra Sukabumi. Selain itu, skripsi ini diharapkan dapat memberikan kontribusi pengetahuan dalam bidang Machine Learning dan memberikan manfaat bagi pembaca serta pihak-pihak terkait. Sehubungan dengan itu penulis menyampaikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada :

1. Bapak Rektor Universitas Nusa Putra Sukabumi.
2. Wakil Rektor I Bidang Akademik Universitas Nusa Putra Sukabumi.
3. Kepala Program Studi Universitas Nusa Putra Sukabumi.
4. Dosen Pembimbing I Universitas Nusa Putra Sukabumi Ibu Ivana Lucia Kharisma, M.Kom yang selalu menyempatkan waktunya untuk memberi arahan kepada penulis selama penelitian, bahkan pernah memberi bimbingan di kantin.
5. Dosen Pembimbing II Universitas Nusa Putra Sukabumi Bapak Alun Sujjada, S.Kom.,M.T yang juga memberi arahan kepada penulis.
6. Dosen Penguji Zaenal Alamsyah, M.Kom yang juga memberi masukan dan saran kepada penulis. Juga para dosen Program Studi Teknik Informatika Universitas Nusa Putra Sukabumi yang telah memberikan dan berbagi ilmu dengan penulis, sehingga pada akhirnya penulis dapat memahami apa yang telah dipelajari sebelumnya. Terutama terkait dengan materi machine learning.
7. Orang tua dan keluarga yang telah memberikan dukungan tanpa henti, baik secara moral maupun material, serta doa dan kasih sayang yang tiada henti, sehingga penulis dapat menyelesaikan skripsi ini. Terima kasih atas segala pengorbanan dan cinta yang selalu menguatkan langkah penulis.
8. Kepada teman lelaki terbaik yang selalu memberikan dukungan dan kesabaran tanpa batas. Terima kasih telah menjadi sumber inspirasi dan semangat dalam menyelesaikan skripsi ini. Kehadirannya telah memberikan kekuatan dan motivasi yang luar biasa dalam

setiap langkah perjalanan ini. Terima kasih telah menemani dari awal kuliah hingga saat ini, berbagi suka dan duka dalam mengejar cita-cita bersama. Semoga keberhasilan ini menjadi awal yang baik untuk masa depan.

9. Rekan-rekan mahasiswa yang selalu memberikan dukungan, semangat, dan kebersamaan selama masa studi. Terima kasih atas kerjasama, persahabatan, dan kebersamaan yang telah terjalin, yang semuanya berperan besar dalam membantu penulis menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, oleh karena itu kritik dan saran yang membangun dari berbagai pihak sangat kami harapkan demi perbaikan. Amin Yaa Rabbal 'Alamiin.

Sukabumi, 21 Juni 2024



Dini Aryani  
Penulis

## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik UNIVERSITAS NUSA PUTRA , saya yang bertanda tangan di bawah ini:

Nama : Dini Aryani  
NIM : 20200040030  
Program Studi : Teknik Informatika  
Jenis karya : Tugas Akhir

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Nusa Putra **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty- Free Right*)** atas karya ilmiah saya yang berjudul :

**“Pemodelan Topik Pemilu 2024 Menggunakan Metode BERTopic pada Artikel Berita Detik.com”**

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Nusa Putra berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Sukabumi  
Pada tanggal : 21 Juni 2024

Yang menyatakan

  
28BD3ALX205663653  
**Dini Aryani**

## DAFTAR ISI

HALAMAN JUDUL.....	i
PERNYATAAN PENULIS .....	ii
PENGESAHAN SKRIPSI .....	iii
ABSTRAK.....	v
ABSTRACT .....	vi
KATA PENGANTAR .....	vii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terkait.....	6
2.2 Landasan Teori .....	11
2.2.1 Topik.....	11
2.2.2 <i>Natural Language Processing</i> .....	11
2.2.3 <i>Topic Modelling</i> .....	12
2.2.4 <i>Clustering</i> .....	12
2.2.5 Artikel Berita .....	13
2.2.6 <i>Bidirectional Encoder Representations Transformers (BERT)</i> .....	13
2.2.7 <i>Scraping</i> .....	14
2.2.8 <i>Preprocessing</i> .....	15
2.2.9 BERTopic .....	16
2.2.10 <i>K-means</i> .....	18
2.2.11 Evaluasi .....	20

2.2.12	Streamlit .....	26
2.3	Kerangka Pemikiran .....	26
<b>BAB III METODOLOGI PENELITIAN.....</b>		<b>28</b>
3.1	Identifikasi Masalah .....	28
3.2	Pengumpulan Data.....	29
3.2.1	Studi Literatur .....	29
3.2.2	Scraping .....	29
3.3	Preprocessing Data .....	31
3.3.1	Case Folding.....	31
3.3.2	Cleaning.....	31
3.3.3	Tokenizing.....	32
3.3.4	Stopword Removal.....	33
3.4	Implementasi Model .....	33
3.4	Evaluasi Model.....	34
3.6	Deployment.....	35
3.7	Alat dan Bahan .....	35
3.8	Perancangan Sistem.....	36
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>39</b>
4.1	Implementasi Pengumpulan <i>Dataset</i> .....	39
4.2	Implementasi <i>Text Preprocessing</i> .....	45
4.3	Implementasi Pemodelan Topik BERTopic .....	49
4.3.1	Embeddings.....	49
4.3.2	Dimentionality Reduction .....	51
4.3.3	Clustering.....	52
4.3.4	Tokenizer .....	54
4.4	Evaluasi Pemodelan Topik BERTopic .....	58
4.5	Implementasi Sistem .....	65
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>67</b>
4.1	Kesimpulan.....	67
4.2	Saran.....	68
<b>DAFTAR PUSTAKA .....</b>		<b>69</b>

## DAFTAR TABEL

Tabel 2.1 Penelitian terkait.....	6
Tabel 2.2 Interpretasi nilai Silhouette Coefficient.....	22
Tabel 3.1 Contoh hasil <i>case folding</i> .....	31
Tabel 3.2 Contoh hasil <i>cleaning</i> .....	32
Tabel 3.3 Contoh hasil <i>tokenizing</i> .....	32
Tabel 3.4 Contoh hasil <i>stopword removal</i> .....	33
Tabel 3.5 Contoh hasil <i>topic modelling</i> BERTopic.....	34
Tabel 3.6 Contoh hasil evaluasi <i>topic coherence</i> BERTopic .....	35
Tabel 4.1 Hasil <i>scraping link index</i> .....	39
Tabel 4.2 Hasil <i>scraping</i> tahap pertama .....	40
Tabel 4.3 Hasil <i>scraping</i> tahap kedua .....	41
Tabel 4.4 Hasil penggabungan data <i>scraping</i> .....	43
Tabel 4.5 Hasil <i>case folding</i> .....	45
Tabel 4.6 Hasil <i>cleaning</i> .....	46
Tabel 4.7 Hasil <i>tokenizing</i> .....	47
Tabel 4.8 Hasil <i>stopword removal</i> .....	48
Tabel 4.9 Hasil <i>postprocessing</i> .....	49
Tabel 4.10 Hasil <i>embeddings</i> .....	50
Tabel 4.11 Hasil <i>dimentionality reduction</i> .....	51
Tabel 4.12 Hasil <i>tokenizer</i> .....	55
Tabel 4.13 Hasil <i>weighting scheme</i> .....	55
Tabel 4.14 Hasil evaluasi <i>topic coherence</i> .....	64



## DAFTAR GAMBAR

Gambar 2.1 Proses BERTopic .....	17
Gambar 2.2 Ilustrasi klasterisasi algoritma K-means .....	19
Gambar 2.3 Kerangka berpikir .....	26
Gambar 3.1 Tahap penelitian.....	28
Gambar 3.2 Desain menu <i>home</i> .....	36
Gambar 3.3 Desain menu data mentah.....	37
Gambar 3.4 Desain menu informasi artikel.....	38
Gambar 4.1 Hasil evaluasi <i>Elbow</i> .....	53
Gambar 4.2 Grafis <i>Silhouette plot</i> .....	53
Gambar 4.3 Hasil <i>clustering</i> .....	54
Gambar 4.4 Grafis <i>intertopic distance map</i> .....	59
Gambar 4.5 Grafis <i>hierarchical clustering</i> .....	59
Gambar 4.6 Grafis <i>term score decline per topic</i> \ .....	60
Gambar 4.7 Grafis <i>similarity matrix</i> .....	61
Gambar 4.8 Grafis <i>topic word scores</i> .....	62
Gambar 4.9 <i>Word cloud</i> .....	62
Gambar 4.10 Implementasi menu <i>home</i> .....	65
Gambar 4.11 Implementasi menu BERTopic.....	66
Gambar 4.12 Implementasi menu informasi artikel .....	66



# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pada tanggal 14 Februari 2024, Indonesia menyelenggarakan pemilihan umum serentak untuk menentukan pemimpin serta anggota legislatif di tingkat kabupaten/kota, provinsi, nasional, dan juga wakil-wakil di DPD-RI. Dari sekitar 204,8 juta pemilih tetap yang terdaftar, lebih dari 116,5 juta atau sekitar 56 persen adalah generasi Z dan milenial. Generasi Z, kelahiran tahun 1995 hingga 2000-an, memiliki hak pilih sekitar 22,8 persen. Sedangkan milenial, yang lahir antara tahun 1980 hingga 1994, memiliki suara sekitar 33,6 persen dari total [1]. Hal tersebut menunjukkan bahwa pemilu 2024 akan diwarnai oleh partisipasi yang dominan dari kalangan pemilih muda yaitu generasi Z dan milenial.

Kedua generasi ini telah tumbuh dalam era digital, di mana akses ke *internet* dan media sosial menjadi bagian penting dari kehidupan sehari-hari generasi tersebut. Dukungan dan partisipasi politik generasi milenial ini akan terbentuk jika terus menerus terpapar media digital yang berisikan sosialisasi atau informasi/berita tentang kampanye mendukung kebijakan-kebijakan pemerintah yang positif. Mengingat generasi ini adalah mayoritas pengguna media digital (*internet*, media sosial dan lain sebagainya) [2].

Generasi Z dan Milenial aktif dalam berdiskusi, berbagi artikel berita, dan mencari opini dari berbagai sumber yang tersedia secara daring. Dengan demikian, *internet* menjadi sumber utama bagi generasi milenial dan Z untuk membentuk pandangan politik dan membuat keputusan dalam pemilihan umum. Interaksi ini kemudian dipermudah dengan adanya media massa *online* yang menyajikan berita-berita aktual yang bisa diakses semua kalangan. Ketersediaan dan kemudahan dalam mengakses media massa secara *online* membuat media ini menjadi pilihan yang sangat diminati masyarakat [3].

Detik.com adalah situs berita daring di Indonesia yang didirikan pada 9 Juli 1998 oleh Budiono Darsono, Yayan Sopyan, Abdul Rahman, dan Didi Nugraha. Sejak tanggal 3 Agustus 2011, detik.com menjadi bagian dari PT Trans Corporation, anak perusahaan CT Corp. Nama "detik.com" diambil dari tabloid DeTik yang didirikan pada tahun 1977. Awalnya fokus pada politik, ekonomi, dan teknologi informasi, tetapi kemudian juga menyajikan berita hiburan dan olahraga. Detik.com dikenal dengan pembaruan berita secara cepat dan merupakan situs informasi digital paling populer di Indonesia. Menurut data dari

*SimilarWeb* pada Januari 2024, Detik.com menduduki peringkat ke 3 teratas sebagai penerbit berita dan media terkemuka.

Detik.com secara aktif memberitakan mengenai pemilu 2024, terbukti dengan adanya sub kanal khusus yang memberitakan pemilu. Hal ini menyebabkan melimpahnya informasi pada media tersebut. Tidak seperti sub kanal pemilu lainnya (seperti sub kanal pemilu Kompas dan CNN), pada sub kanal pemilu Detik belum ada pengkategorian berita berdasarkan topik tertentu. Hal ini menyulitkan pembaca untuk mencari informasi yang lebih mendalam pada sub kanal tersebut. Namun, melacak dan menganalisis topik-topik yang muncul dalam cakupan berita menjadi tantangan, terutama mengingat volume besar artikel berita yang dipublikasikan setiap hari. Oleh karena itu, diperlukan pendekatan yang efisien dan efektif untuk memodelkan topik-topik terkait Pemilu 2024 dari artikel berita detik.

*Topic Modelling* adalah salah satu teknik di dalam *Natural Language Processing* (NLP) untuk menganalisis teks [4]. Algoritma *topic modelling* digunakan untuk mengidentifikasi pola tersembunyi dalam rangkaian kata yang terdistribusi di dalam sebuah kumpulan dokumen. Hasil dari teknik ini adalah sekumpulan topik yang terdiri dari beberapa kelompok kata yang muncul bersamaan dalam dokumen berdasarkan pola tertentu. Dalam konteks *topic modelling*, "topik" merujuk pada sekumpulan kata atau frasa yang sering muncul bersama dalam dokumen dan membentuk tema atau konsep tertentu yang dapat dikenali [18]. Penelitian ini akan menggunakan Meta Deskripsi sebagai pengukurannya. Meta Deskripsi adalah keterangan yang digunakan untuk menjelaskan apa maksud dari konten yang dibuat [5].

Pada penelitian sebelumnya yang membandingkan beberapa metode *Natural Language Processing*, *Bidirectional Encoder Representations Transformers* (BERT) dianggap lebih baik dibandingkan model lainnya, termasuk LDA [6]. BERT adalah model representasi bahasa baru yang dirancang untuk dapat melakukan pelatihan representasi dua arah dari teks yang tidak berlabel [7]. Kemudian, salah satu yang memanfaatkan BERT sebagai model untuk pemodelan topik adalah BERTopic [8]. BERTopic merupakan model *transformer* BERT yang dikhususkan untuk pemodelan topik dengan menggunakan pendekatan c-TF-IDF. Sehingga pemodelan topik dapat diinterpretasikan dengan mudah tanpa mengurangi kata – kata kunci dalam deskripsi kata yang dilakukan topik modeling [9]

Untuk itu, penulis merasa tertarik untuk melakukan penelitian dengan judul Pemodelan Topik Pemilu 2024 Menggunakan Metode BERTopic pada Artikel Berita Detik.com

## **1.2 Rumusan Masalah**

Rumusan masalah yang sesuai untuk penelitian ini dapat dirumuskan sebagai berikut:

- a. Bagaimana menerapkan BERTopic untuk pemodelan topik pada meta deskripsi artikel berita terkait Pemilu 2024?
- b. Apa saja topik-topik utama yang mendominasi cakupan berita Pemilu 2024 pada portal berita Detik.com?

## **1.3 Batasan Masalah**

Dengan merujuk pada konteks masalah yang telah diuraikan sebelumnya, penulis memberikan batasan penulisan untuk memastikan bahwa pembahasan dan isi tulisan tetap sesuai dengan judul penelitian. Batasan-batasan yang diberlakukan melibatkan hal-hal berikut:

- a. Data yang dikumpulkan adalah data dalam Bahasa Indonesia.
- b. Data yang digunakan merupakan data meta deskripsi artikel berita pada sub kanal pemilu dari tanggal 1 September 2023 hingga 14 Februari 2024 pada portal berita Detik.com.
- c. Menggunakan bahasa pemrograman Python.
- d. Penggunaan pustaka BERTopic untuk pemodelan topik

## **1.4 Tujuan**

Tujuan yang hendak dicapai dalam penelitian ini adalah:

- a. Menerapkan metode BERTopic untuk memodelkan topik pada meta deskripsi artikel berita Pemilu 2024.
- b. Mengidentifikasi dan menganalisis topik-topik utama yang mendominasi cakupan berita Pemilu 2024 di portal berita Detik.com.
- c. Memberikan informasi mengenai tren topik pemilu 2024 pada artikel berita di portal berita Detik.com.
- d. Mengelompokkan berita Pemilu 2024 dengan topik yang serupa untuk mempermudah pemahaman.

## 1.5 Manfaat Penelitian

Manfaat dari penelitian ini di antaranya sebagai berikut:

- a. Manfaat Bagi Penulis
  - 1) Memberikan pengetahuan lebih kepada penulis dalam menganalisis *topic modelling* terhadap meta deskripsi berita pemilu 2024 menggunakan BERTopic.
  - 2) Sebagai salah satu syarat dalam penyelesaian gelar strata (S1) program studi Teknik Informatika Universitas Nusa Putra.
- b. Manfaat Bagi Universitas
  - 1) Menjadi sebuah tolak ukur bagi Universitas dalam menentukan keberhasilan dan kemampuan penulis dalam mengimplementasikan ilmu yang sudah didapatkan selama menempuh pendidikan perkuliahan di universitas.
  - 2) Sebagai referensi untuk penelitian selanjutnya.
- c. Manfaat Bagi Masyarakat
  - 1) Informasi mengenai hasil *topic modelling* terhadap topik pemilu 2024.
  - 2) Dapat dijadikan studi literatur untuk penelitian selanjutnya.

## 1.6 Sistematika Penulisan

Sistematika penulisan dalam tugas akhir ini terdiri dari beberapa bab yang dirancang secara terstruktur untuk menyajikan isi penelitian dengan jelas. Berikut adalah gambaran garis besar mengenai konten masing-masing bab:

### **BAB I: PENDAHULUAN**

Merupakan bab awal yang memberikan latar belakang, permasalahan, tujuan, manfaat, dan batasan penelitian.

### **BAB II: TINJUAN PUSTAKA**

Berisi *review* literatur dan penelitian terkait untuk membangun landasan teoritis dan kerangka konseptual.

### **BAB III: METODE PENELITIAN**

Pada bab 3, yaitu metodologi penelitian ini membahas mengenai metode serta langkah-langkah yang digunakan penulis dalam melakukan penelitian yang dilakukan. Bab ini mencakup objek penelitian, analisis, tahap penelitian, alat dan bahan serta perancangan sistem.

### **BAB IV: HASIL**

Bab ini mencakup penjelasan tentang penerapan hasil dan evaluasi dari uji coba penelitian mulai dari tahap analisis, perancangan, implementasi perancangan, hasil pengujian, dan penerapannya. Penjelasan ini mencakup aspek teoritis baik secara kualitatif, kuantitatif, maupun statistik.

#### **BAB V: PENUTUP**

Bab ini merupakan penutup dari laporan atau tesis yang menyajikan rangkuman dari keseluruhan penelitian serta memberikan kesimpulan, implikasi praktis, saran untuk penelitian masa depan, dan pesan terakhir dari penulis.





## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **4.1 Kesimpulan**

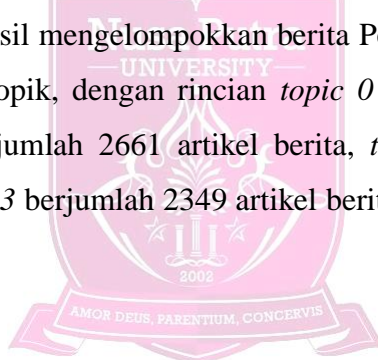
Berdasarkan hasil penelitian pemodelan topik pemilu 2024 menggunakan metode BERTopic pada artikel berita Detik.com, dapat diambil beberapa kesimpulan, yaitu:

- a. Penerapan BERTopic untuk pemodelan topik pemilu dilakukan menggunakan meta deskripsi artikel berita terkait Pemilu 2024 menggunakan *dataset* hasil *scraping* pada rentang waktu tanggal 1 September 2023 hingga 14 Februari 2024 dengan total data 15.019 artikel. Data tersebut kemudian dilakukan *text preprocessing* dengan melewati 4 proses yaitu *case folding*, *cleaning*, *tokenizing* dan *stopword removal* agar data bersih dan menjadi lebih terstruktur. Kemudian menggunakan *dataset* tersebut dilakukan pemodelan topik menggunakan BERTopic yang melewati lima tahapan yaitu *embeddings* menggunakan *sentence-transformers* "distiluse-base-multilingual-cased-v1", *dimensionality reduction* menggunakan UMAP, *clustering* menggunakan K-Means dengan nilai k optimal menggunakan hasil evaluasi Elbow dengan k sama dengan 5, *tokenizer* menggunakan *CountVectorizer* dan *weighting scheme* menggunakan c-TF-IDF. Pemodelan topik menggunakan BERTopic menghasilkan nilai *topic coherence* sebesar 0.0885752;
- b. Berdasarkan hasil pemodelan topik yang dilakukan topik – topik utama yang mendominasi cakupan berita Pemilu 2024 pada portal berita Detik.com, yaitu lima topik. Dari kelima topik tersebut, topik terkait dengan capres dan cawapres menjadi topik yang paling sering muncul pada portal berita Detik.com mengenai Pemilu 2024.
- c. Penelitian ini berhasil mengidentifikasi lima topik utama yang mendominasi pemberitaan pemilu 2024 di portal berita Detik.com. Topik-topik tersebut mencakup: Topic 0, berfokus pada calon-calon presiden dan wakil presiden yang mencerminkan tingginya perhatian terhadap figur-figur utama dalam pemilu, dengan jumlah 6211 artikel berita dengan representasi kata 'ganjar', 'prabowo', 'anies', 'imin'; Topic 1, berfokus pada

Topik ini berfokus pada nomor urut pasangan capres-cawapres dalam pemilihan umum di Indonesia, terutama menyebutkan tokoh seperti Gibran Rakabuming Raka, dengan jumlah artikel 2661 artikel berita dengan representasi kata 'nomor', 'urut', 'gibran', 'indonesia'; Topic 2, tentang berita tentang Presiden Joko Widodo serta dinamika politiknya, dengan jumlah 2615 artikel berita dengan representasi kata 'presiden', 'politik', 'jokowi', 'survei'; Topic 3, berkaitan dengan pemilihan umum dan pemilihan presiden, termasuk suara dan survei terkait 2349 artikel berita dengan representasi kata '2024', 'pemilu', 'pilpres', 'kpu'; Topic 4, tentang berita tentang debat capres dan cawapres yang terjadi, dengan jumlah 1183 artikel berita 'debat', 'pilpres', 'cawapres', '2024'.

Dapat disimpulkan berdasarkan Tabel 4.13, topic 0 topik terkait capres dan cawapres merupakan topik yang sering diberitakan oleh portal berita Detik.com pada sub kanal pemilu dengan jumlah 6211 artikel berita.

- d. Penelitian ini berhasil mengelompokkan berita Pemilu 2024 yang serupa ke dalam 5 jenis topik, dengan rincian *topic 0* berjumlah 6211 artikel berita, *topic 1* berjumlah 2661 artikel berita, *topic 2* berjumlah 2615 artikel berita, *topic 3* berjumlah 2349 artikel berita dan *topic 4* berjumlah 1183 artikel berita.



## 4.2 Saran

Berdasarkan hasil penelitian pemodelan topik pemilu 2024 menggunakan metode BERTopic pada artikel berita Detik.com, dapat diambil beberapa saran untuk penelitian di masa depan, yaitu:

- a. Melakukan perbandingan metode pemodelan topik BERTopic dengan model lainnya seperti *Correlated Topic Models* (CTM), *Gibbs Sampling Dirichlet Multinomial Mixture* (GSDMM), *Non-Negative Matrix Factorization* (NMF), dan lain sebagainya;
- b. Menggunakan metode *embeddings* yang berbeda;
- c. Menambah *dataset* dari portal berita lainnya terkait dengan Pemilu 2024;
- d. Membandingkan beberapa portal berita untuk mengetahui berita yang mendominasi dari setiap portal berita mengenai Pemilu 2024.

## DAFTAR PUSTAKA

- [1] Ahmad Arif, “Memperebutkan Gen Z dan Milenial,” Kompas.id.
- [2] Christiany Juditha dan Josep J. Darmawan, “PENGUNAAN MEDIA DIGITAL DAN PARTISIPASI POLITIK GENERASI MILENIAL,” vol. 22, 2018.
- [3] Emilsyah Nur, “PERAN MEDIA MASSA DALAM MENGHADAPI SERBUAN MEDIA ONLINE,” 2021.
- [4] C. C. Aggarwal dan C. C. Aggarwal, *Mining Text Data*. Springer, 2015.
- [5] Putri Septiani dan Hendra Kurniawan, “Analisa Penggunaan Keyword Untuk Implementasi Search Engine Optimization (SEO),” vol. XV, 2020.
- [6] Roman Egger dan Joanne Yu, “A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts,” 2022.
- [7] J. Devlin, M.-W. Chang, K. Lee, dan K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Okt 2018, [Daring]. Tersedia pada: <http://arxiv.org/abs/1810.04805>
- [8] G. Malki, “Efficient Sentiment Analysis and Topic Modeling in NLP using Knowledge Distillation and Transfer Learning,” 2023.
- [9] Maarten Grootendorst, “BERTopic.” Diakses: 27 Maret 2024. [Daring]. Tersedia pada: <https://maartengr.github.io/BERTopic>
- [10] Enny Fitriani Nasution, “KOMPARASI *K-MEANS* CLUSTERING DAN LATENT DIRICHLET ALLOCATION PADA KASUS SHORT TEXT TOPIC MODELLING,” 2023.
- [11] S. E. Uthirapathy dan D. Sandanam, “Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model,” *Procedia Comput Sci*, vol. 218, hlm. 908–917, 2023.
- [12] A. Abuzayed dan H. Al-Khalifa, “BERT for Arabic topic modeling: An experimental study on BERTopic technique,” *Procedia Comput Sci*, vol. 189, hlm. 191–194, 2021.
- [13] I. K. T. Mertayasa dan I. D. M. B. A. Darmawana, “Pemodelan Topik Pada Ulasan Hotel Menggunakan Metode BERTopic Dengan Prosedur c-TF-IDF,” *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 1, no. 1, hlm. 307–316, 2022.

- [14] E. Puspita, D. F. Shiddieq, dan F. F. Roji, “Pemodelan Topik pada Media Berita Online Menggunakan Latent Dirichlet Allocation (Studi Kasus Merek Somethinc),” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, hlm. 481–489, 2024.
- [15] M. Amien, “Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia,” *arXiv preprint arXiv:2304.02746*, 2023.
- [16] S. D. Lestari dan A. Z. Falani, “Implementasi Fitur Assistance Virtual Sebagai Media Pembelajaran Interaktif,” *Jurnal RESTIKOM: Riset Teknik Informatika dan Komputer*, vol. 5, no. 1, hlm. 46–54, 2023.
- [17] Y. Matira dan I. Setiawan, “Pemodelan Topik pada Judul Berita Online Detikcom Menggunakan Latent Dirichlet Allocation,” *Estimasi: Journal of Statistics and Its Application*, hlm. 53–63, 2023.
- [18] I. Vayansky dan S. A. P. Kumar, “A review of topic modeling methods,” *Inf Syst*, vol. 94, hlm. 101582, 2020.
- [19] D. Ridhwanullah, “Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode Latent Dirichlet Allocation,” 2022.
- [20] D. Yolanda, M. H. Hersyah, dan E. Marozi, “Implementasi Metode Unsupervised Learning Pada Sistem Keamanan Dengan Optimalisasi Penyimpanan Kamera IP,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, hlm. 1099–1105, 2021.
- [21] D. P. Sari, F. Sembiring, D. P. Sulisdianto, Y. Jentiner, dan E. Suratno, “Implementasi Algoritma Fuzzy C-Means Untuk Memprediksi Faktor Siswa Membolos (Studi Kasus: Faktor Siswa Membolos Di Smpn 1 Parakansalak),” *Jurnal RESTIKOM: Riset Teknik Informatika dan Komputer*, vol. 2, no. 1, hlm. 1–7, 2020.
- [22] N. Wakhidah, “Clustering menggunakan *k-means* algorithm,” *Jurnal Transformatika*, vol. 8, no. 1, hlm. 33–39, 2010.
- [23] S. Setyaningtyas, B. I. Nugroho, dan Z. Arif, “Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma *K-means*,” *Jurnal*

*Teknoif Teknik Informatika Institut Teknologi Padang*, vol. 10, no. 2, hlm. 52–61, 2022.

- [24] D. N. Chandra, G. Indrawan, dan I. N. Sukaraja, “Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram,” *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 10, no. 1, hlm. 11–19, 2016.
- [25] A. M. Mansah, “Tren Pergeseran Media Konvensional Ke Era Digitalisasi (Studi Kasus Konvergensi Media Di Lembaga Kantor Berita Nasional Antara Biro Sulawesi Selatan-Sulawesi Barat),” *Al-MUNZIR*, vol. 12, no. 1, hlm. 121–130, 2019.
- [26] B. Kurniawan, A. A. Aldino, dan A. R. Isnain, “Sentimen Analisis Terhadap Kebijakan Penyelenggara Sistem Elektronik (Pse) Menggunakan Algoritma Bidirectional Encoder Representations From Transformers (Bert),” *J. Teknol. dan Sist. Inf*, vol. 3, no. 4, hlm. 98–106, 2022.
- [27] S. Nabilah, “Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Novel Online Di Media Sosial Menggunakan Latent Dirichlet Allocation Dan Bidirectional Encoder Representations From Transformers,” Perpustakaan Fakultas Sains dan Teknologi UIN Jakarta, 2022.
- [28] A. Josi dan L. A. Abdillah, “Penerapan teknik web scraping pada mesin pencari artikel ilmiah,” *arXiv preprint arXiv:1410.5777*, 2014.
- [29] V. A. Flores, P. A. Permatasari, dan L. Jasa, “Penerapan web scraping sebagai media pencarian dan menyimpan artikel ilmiah secara otomatis berdasarkan keyword,” *Maj. Ilm. Teknol. Elektro*, vol. 19, no. 2, hlm. 157, 2020.
- [30] D. Normawati dan S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, vol. 5, no. 2, hlm. 697–711, 2021.
- [31] J. Supriyanto, D. Alita, dan A. R. Isnain, “Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring,” *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, hlm. 74–80, 2023.

- [32] M. Nashrullah dan D. A. Efrilianda, "Sentiment Analysis of Independent Campus Policy on Twitter Using Support Vector Machine and Naïve Bayes Classifier," *Journal of Advances in Information Systems and Technology*, vol. 4, no. 1, hlm. 13–23, 2022.
- [33] A. Sujjada dan A. Fergina, "Implementasi Metode Vector Space Model Untuk Deteksi Emosi Menggunakan Data Teks Twitter," *Jurnal RESTIKOM: Riset Teknik Informatika dan Komputer*, vol. 3, no. 3, hlm. 116–129, 2021.
- [34] Y. An, D. Kim, J. Lee, H. Oh, J.-S. Lee, dan D. Jeong, "Topic Modeling-Based Framework for Extracting Marketing Information From E-Commerce Reviews," *IEEE Access*, vol. 11, hlm. 135049–135060, 2023.
- [35] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," *arXiv preprint arXiv:2203.05794*, 2022.
- [36] A. Alamsyah dan N. D. Girawan, "Improving clothing product quality and reducing waste based on consumer review using RoBERTa and BERTopic language model," *Big Data and Cognitive Computing*, vol. 7, no. 4, hlm. 168, 2023.
- [37] R. N. Fahmi, M. Jajuli, dan N. Sulistiyowati, "Analisis pemetaan tingkat kriminalitas di kabupaten Karawang menggunakan Algoritma *K-means*," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 4, no. 1, hlm. 67–79, 2021.
- [38] R. B. Trianto, A. S. Nugroho, dan E. Supriyadi, "Klasterisasi Menggunakan Algoritma *K-means* Dan *Elbow* Pada Opini Masyarakat Tentang Kebijakan Sekolah Luring Tahun 2022," *Jurnal Inovtek Polbeng Seri Informatika*, vol. 8, no. 1, hlm. 1–13, 2023.
- [39] M. S. Pangestu dan M. A. Fitriani, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma *K-means*," *Sainteks*, vol. 19, no. 2, hlm. 141–155, 2022.
- [40] H. Fitriyah, E. M. Safitri, N. Muna, M. Khasanah, D. A. Aprilia, dan D. Nurdiansyah, "IMPLEMENTASI ALGORITMA CLUSTERING DENGAN MODIFIKASI METODE *ELBOW* UNTUK MENDUKUNG

STRATEGI PEMERATAAN BANTUAN SOSIAL DI KABUPATEN BOJONEGORO,” *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 4, no. 3, hlm. 1598–1607, 2023.

- [41] D. Adimanggala, F. A. Bachtiar, dan E. Setiawan, “Evaluasi Topik Tersembunyi Berdasarkan Aspect Extraction menggunakan Pengembangan Latent Dirichlet Allocation,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, hlm. 511–519, 2021.
- [42] Made Anditya Mahesastra dan Dewa Made Bayu Atmaja Darmawana, “Pemodelan Topik Teks Berita Menggunakan DistilBERT,” vol. 1, no. 1, 2022.
- [43] F. Nuraeni, D. Kurniadi, dan G. F. Dermawan, “Pemetaan Karakteristik Mahasiswa Penerima Kartu Indonesia Pintar Kuliah (KIP-K) menggunakan Algoritma *K-means++*,” *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, vol. 11, no. 3, hlm. 437–443, 2023.
- [44] R. Hidayati, A. Zubair, A. H. Pratama, dan L. Indana, “Analisis *Silhouette Coefficient* pada 6 Perhitungan Jarak *K-means* Clustering,” *Techno. Com*, vol. 20, no. 2, hlm. 186–197, 2021.
- [45] D. E. Herwindiati dan T. Handhayani, “Clustering Data Covid-19 Di Indonesia Menggunakan Intelligent *K-means*,” *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 10, no. 2, 2022.
- [46] L. Qadrini, “Metode *K-means* dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017,” *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 13, no. 2, hlm. 5–11, 2020.
- [47] A. P. Riani, A. Voutama, dan T. Ridwan, “Penerapan *K-means* clustering dalam pengelompokan hasil belajar peserta didik dengan metode *Elbow*,” *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, no. 1, hlm. 164–172, 2023.
- [48] N. T. Hartanti, “Metode *Elbow* dan *K-means* Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” *J. Nas. Teknol. dan Sist. Inf*, vol. 6, no. 2, hlm. 82–89, 2020.

- [49] N. Azmi, H. S. Hafsah, Y. Yuyun, dan H. Hazriani, “Penerapan Metode *K-means* Clustering Dalam Mengelompokkan Data Penjualan Obat pada Apotek M23,” *Prosiding SISFOTEK*, vol. 7, no. 1, hlm. 244–248, 2023.
- [50] R. Dwirahmanto dan A. Bisri, “Menentukan Nilai K Pada Metode *K-means* Menggunakan Teknik Grid Search Untuk Strategi Produk Pakaian Medis,” *Jurnal Informatika Multi*, vol. 1, no. 2, hlm. 93–103, 2023.
- [51] I. T. Umagapi, B. Umaternate, H. Hazriani, dan Y. Yuyun, “Uji Kinerja *K-means* Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa,” *Prosiding SISFOTEK*, vol. 7, no. 1, hlm. 303–308, 2023.
- [52] A. M. Sikana dan A. W. Wijayanto, “Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering,” *J. Ilmu Komput*, vol. 14, no. 2, hlm. 66–78, 2021.
- [53] D. A. Saidah, R. Santoso, dan T. Widiari, “Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Kesehatan Lingkungan Menggunakan Metode Partitioning Around Medoids Dengan Validasi Indeks Internal,” *Jurnal Gaussian*, vol. 11, no. 2, hlm. 302–312, 2022.
- [54] D. R. S. Saputro, “Algoritme Partitioning Around Medoid (Pam) Dengan Calinski-Harabasz Index Untuk Clustering Data Outlier,” *UNEJ e-Proceeding*, hlm. 22–29, 2022.
- [55] A. Putranto, N. L. Azizah, dan I. R. I. Astutik, “Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit,” *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 4, no. 2, hlm. 442–452, 2023.
- [56] I. Amal, E. W. Pamungkas, S. Kom, dan M. Kom, “Aplikasi Pendeteksi Berita Palsu Bahasa Indonesia Menggunakan Framework Flask dan Streamlit serta Algoritma Machine Learning,” Universitas Muhammadiyah Surakarta, 2023.