

# TOPIC MODELING MENGGUNAKAN BERTOPIC DENGAN LLAMA2 SEBAGAI TOPIC REPRESENTATION TUNING

Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata-1 Pada  
Jurusan Teknik Informatika



Oleh:

Alpian Khairi  
NIM: 09021282025044

**Jurusan Teknik Informatika**  
**FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**  
**2024**

LEMBAR PENGESAHAN SKRIPSI

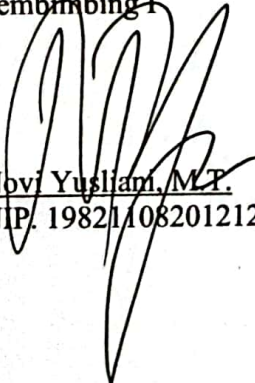
TOPIC MODELING MENGGUNAKAN BERTOPIC DENGAN  
LLAMA2 SEBAGAI TOPIC REPRESENTATION TUNING

Oleh:


Alpian Khairi  
NIM: 09021282025044

Indralaya, 17-05-2024

Pembimbing I,

  
Novi Yuslimi, M.T.  
NIP. 198211082012122001

Pembimbing II,

  
Danny Matthew Saputra, M.Sc.  
NIP. 198505102015041002

Mengetahui,

Ketua Jurusan Teknik Informatika  
  
  
Dr. M. Fachrurrozi, S.Si., M.T.  
NIP. 198005222008121002

## TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Jumat tanggal 17 Mei 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Alpian Khairi  
NIM : 09021282025044  
Judul : *Topic Modeling Menggunakan Bertopic Dengan LLAMA2 Sebagai Topic Representation Tuning*

Dan dinyatakan **LULUS**.

1. Ketua

Mastura Diana Marieska, M.T  
NIP. 198603212018032001



2. Penguji I

Desty Rodiah, S.Kom., M.T.  
NIP 198912212020122011



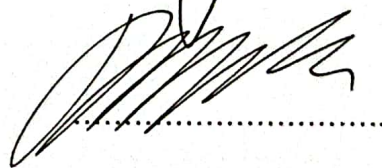
3. Pembimbing I

Novi Yusliani, M.T.  
NIP 198211082012122001

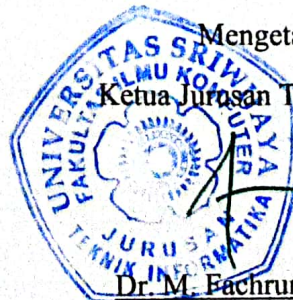


4. Pembimbing II

Danny Matthew Saputra, M.Sc.  
NIP 198505102015041002



Mengetahui,  
Ketua Jurusan Teknik Informatika



Dr. M. Fachrurrozi, S.Si., M.T.  
NIP 198005222008121002



## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Alpian Khairi

NIM : 09021282025044

Program Studi : Teknik Informatika

Judul Skripsi : *Topic Modeling Menggunakan Bertopic Dengan LLAMA2 Sebagai Topic Representation Tuning*

Hasil pengecekan *Software iThenticate/Turnitin*: 19%

Menyatakan bahwa laporan proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak mana pun.

Indralaya, 17-05-2024



Alpian Khairi  
NIM. 09021282025044

## **MOTTO DAN PERSEMBAHAN**

Motto:

Ketika kita melewati kesulitan, kita mendapati pintu-pintu kemudahan yang terbuka. Berhenti mengeluh, mulailah berkembang.

“Karena sesungguhnya sesudah kesulitan itu ada kemudahan. Sesungguhnya sesudah kesulitan itu ada kemudahan.”

(QS. Al-Insyirah ayat 5-6)

Kupersembahkan Karya tulis ini kepada:

- Allah SWT
- Kedua Orang Tua
- Keluarga Besar
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

## ABSTRACT

*The growing use of social media, particularly Twitter, has generated large amounts of information in the form of text data covering a wide range of topics and issues. Analyzing Twitter data has the potential to find important insights into topics relevant to society. Topic modeling is one of the latest innovations in text data processing to find topics in a set of text. This research aims to perform topic modeling on Indonesian tweets using BERTopic with LLAMA2 as topic representation tuning. LLAMA2 is used to generate labels from a set of keywords generated from c-TF-IDF calculations. The dataset used consists of 10,000 Indonesian tweets taken from the Twitter account @detikcom. The data is divided into 2 parts, 8,000 tweets are used for training data and 2,000 tweets are used for testing. Based on the results of topic modeling with BERTopic, 49 total topics were obtained. Topic Modeling Evaluation is done using coherence score cv, obtained an average coherence score cv of 0.86 on training data and 0.73 on testing data.*

**Keywords:** *Topic Modeling, BERTopic, Large Language Models, Coherence Score cv, Tweet*

## ABSTRAK

Pertumbuhan penggunaan media sosial, khususnya Twitter telah menghasilkan informasi dalam bentuk data teks dalam jumlah besar yang mencakup berbagai topik dan isu. Melakukan analisis data Twitter berpotensi menemukan wawasan penting tentang topik-topik yang relevan dengan masyarakat. Pemodelan topik adalah salah satu inovasi terbaru dalam pengolahan data teks untuk menemukan topik pada sekumpulan teks. Penelitian ini bertujuan untuk melakukan pemodelan topik pada *tweet* berbahasa Indonesia dengan menggunakan BERTopic dengan LLAMA2 sebagai *topic representation tuning*. LLAMA2 digunakan untuk menghasilkan label dari sekumpulan kata kunci yang dihasilkan dari perhitungan c-TF-IDF. Dataset yang digunakan terdiri dari 10.000 *tweets* berbahasa Indonesia yang diambil dari akun Twitter @detikcom. Data dibagi menjadi 2 bagian, 8.000 *tweets* digunakan untuk data *training* dan 2.000 *tweets* digunakan untuk *testing*. Berdasarkan hasil pemodelan topik dengan BERTopic, diperoleh 49 total topik. Evaluasi Pemodelan Topik dilakukan menggunakan *coherence score cv*, diperoleh rata-rata *coherence score cv* sebesar 0,86 pada data *training* dan 0,73 pada data *testing*.

**Kata Kunci:** BERTopic, *Cohrence Score cv*, *Large Language Models*, Pemodelan Topik, *Tweet*

## KATA PENGANTAR

Puji syukur kepada Allah SWT atas rahmat dan nikmat-Nya yang telah diberikan kepada penulis sehingga dapat menyelesaikan skripsi ini dengan baik.

Skripsi ini disusun sebagai salah satu syarat menyelesaikan pendidikan program Strata-1 di Fakultas Ilmu Komputer Universitas Sriwijaya. Dalam menyelesaikan skripsi ini, penulis menerima bantuan, bimbingan dan dukungan dari banyak pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Allah SWT atas rahmat dan nikmat-Nya penulis dapat menyelesaikan skripsi ini dengan baik.
2. Kedua orang tua, saudara dan teman yang telah mendoakan, memberi semangat, motivasi, dan nasihat untuk menyelesaikan skripsi ini.
3. Bapak Dr. M. Fachrurrozi, S.Si., M.T. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya.
4. Ibu Novi Yusliani, M.T. selaku Dosen Pembimbing I dan Bapak Danny Matthew Saputra, M.Sc. selaku dosen Pembimbing II yang telah membimbing, memberikan motivasi serta arahan kepada penulis dalam proses pengerjaan skripsi.
5. Ibu Desty Rodiah, M.T. selaku Dosen Penguji Tugas Akhir yang telah memberikan ilmu, nasihat serta saran yang membangun.
6. Ibu Mastura Diana Marieska, M.T. selaku Pembimbing Akademik selama di Universitas Sriwijaya.



7. Seluruh dosen, staf dan pegawai Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Teman-teman penulis yang telah memberikan saran, motivasi, dan, semangat selama mengerjakan skripsi ini
9. Pihak-pihak lain yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak sekali kekurangan dikarenakan kurangnya pengalaman dan pengetahuan penulis. Oleh karena itu penulis mengharapkan saran dan kritik yang membangun guna kemajuan penelitian selanjutnya. Semoga tugas akhir ini dapat bermanfaat. Terima kasih.

Indralaya, 27 Maret 2024

Penulis

Alpian Khairi

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	i
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	ii
HALAMAN PERNYATAAN .....	iii
MOTTO DAN PERSEMBAHAN .....	iv
ABSTRACT.....	v
ABSTRAK.....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan .....	I-1
1.2 Latar Belakang Masalah .....	I-1
1.3 Rumusan Masalah .....	I-4
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-6
BAB II KAJIAN LITERATUR.....	II-1
2.1 Pendahuluan .....	II-1
2.2 Landasan Teori .....	II-1
2.2.1 Pemodelan Topik .....	II-1
2.2.2 BERTopic.....	II-3
2.2.3 <i>Sentence-BERT</i> .....	II-4
2.2.4 <i>Uniform Manifold Approximation and Projection</i> .....	II-5
2.2.5 HDBSCAN.....	II-6
2.2.6 <i>CountVectorizer</i> .....	II-8
2.2.7 <i>Class Based TF-IDF</i> .....	II-9
2.2.8 LLAMA 2-7b .....	II-10

2.2.9	<i>Coherence Score</i> .....	II-11
2.2.10	Rational Unified Process.....	II-12
2.2.11	Penelitian Lain yang Relevan.....	II-14
2.3	Kesimpulan.....	II-15
<b>BAB III</b>	<b>METODOLOGI PENELITIAN</b> .....	<b>III-1</b>
3.1	Pendahuluan .....	III-1
3.2	Pengumpulan Data .....	III-1
3.3	Tahapan Penelitian .....	III-2
3.3.1	Menentukan Kerangka Kerja Penelitian .....	III-3
3.3.2	Menentukan Kriteria Pengujian .....	III-4
3.3.3	Format Data Pengujian.....	III-5
3.3.4	Alat yang Digunakan Dalam Pelaksanaan Penelitian .....	III-6
3.3.5	Pengujian Penelitian.....	III-7
3.3.6	Analisis Hasil Pengujian dan Sintesis Kesimpulan.....	III-7
3.4	Metode Pengembangan Perangkat Lunak .....	III-8
3.4.1	Fase Insepsi .....	III-8
3.4.2	Fase Elaborasi .....	III-8
3.4.3	Fase Konstruksi.....	III-9
3.4.4	Fase Transisi .....	III-9
3.5	Kesimpulan.....	III-9
<b>BAB IV</b>	<b>PENGEMBANGAN PERANGKAT LUNAK</b> .....	<b>IV-1</b>
4.1	Pendahuluan .....	IV-1
4.2	Fase Insepsi .....	IV-1
4.2.1	Pemodelan Bisnis .....	IV-1
4.2.2	Kebutuhan Sistem .....	IV-2
4.2.3	Analisis dan Perancangan .....	IV-3
4.2.3.1	Analisis Kebutuhan Perangkat Lunak.....	IV-3
4.2.3.2	Analisis Proses Pemodelan Topik.....	IV-3
4.2.3.3	Analisis Hasil Pemodelan Topik.....	IV-7
4.2.4	Implementasi .....	IV-7
4.3	Fase Elaborasi.....	IV-11
4.3.1	Pemodelan Bisnis .....	IV-11

4.3.1.1	Perancangan Data .....	IV-11
4.3.1.2	Perancangan User Interface .....	IV-11
4.3.2	Kebutuhan .....	IV-13
4.3.3	Analisis dan Perancangan .....	IV-13
4.3.3.1	Diagram Aktivitas .....	IV-13
4.3.3.2	Diagram Alur .....	IV-15
4.4	Fase Konstruksi .....	IV-17
4.4.1	Kebutuhan Sistem .....	IV-18
4.4.2	Implementasi .....	IV-18
4.4.2.1	Implementasi Kelas .....	IV-19
4.4.2.2	Implementasi Antarmuka .....	IV-19
4.5	Fase Transisi .....	IV-20
4.5.1	Pemodelan Bisnis .....	IV-21
4.5.2	Kebutuhan .....	IV-21
4.5.3	Analisis dan Perancangan .....	IV-21
4.5.3.1	Rencana Pengujian .....	IV-22
4.5.3.2	Implementasi .....	IV-23
4.6	Kesimpulan .....	IV-24
BAB V	HASIL DAN ANALISIS .....	V-1
5.1	Pendahuluan .....	V-1
5.2	Hasil Penelitian .....	V-1
5.3	Analisis Hasil Penelitian .....	V-4
5.4	Kesimpulan .....	V-6
BAB VI	KESIMPULAN DAN SARAN .....	VI-1
6.1	Pendahuluan .....	VI-1
6.2	Kesimpulan .....	VI-1
6.3	Saran .....	VI-1
DAFTAR PUSTAKA	.....	iii
LAMPIRAN	.....	vi

## DAFTAR GAMBAR

Gambar II-1 Arsitektur BERTopic.....	II-3
Gambar II-2 Diagram proses <i>Sentence-BERT</i> .....	II-4
Gambar II-3 Arsitektur RUP .....	II-13
Gambar III-1 Rincian Kegiatan Penelitian.....	III-3
Gambar III-2 Diagram Alur Proses Umum Perangkat Lunak .....	III-4
Gambar IV-1. Output Proses Document Embedding Menggunakan SBERT ...	IV-4
Gambar IV-2. <i>Output</i> Proses <i>Dimensionality Reduction</i> Menggunakan UMAP.	IV-5
Gambar IV-3. Visualisasi Output Proses Topic Representation Menggunakan c-TF-IDF .....	IV-6
Gambar IV-4. <i>Use Case</i> Pemodelan Topik Menggunakan BERTopic.....	IV-7
Gambar IV-5. Rancangan Antarmuka Pra-Pengolahan Data.....	IV-12
Gambar IV-6. Rancangan Antarmuka Hasil Pemodelan Topik.....	IV-12
Gambar IV-7. Diagram Aktivitas Melakukan Pra-Pengolahan Data Pada Sistem .....	IV-14
Gambar IV-8. Diagram Aktivitas Melakukan Pemodelan Topik .....	IV-15
Gambar IV-9. Diagram Alur Proses Pra-Pengolahan Data.....	IV-16
Gambar IV-10. Diagram Alur Proses Pemodelan Topik Menggunakan BERTopic .....	IV-17
Gambar IV-11. Diagram Kelas Perangkat Lunak .....	IV-18
Gambar IV-12. Implementasi Antarmuka Pra-pengolahan Data.....	IV-20
Gambar IV-13. Implementasi Antarmuka Hasil Pemodelan Topik .....	IV-20
Gambar V-1. Persebaran <i>cluster</i> data pada setiap topik .....	V-5

## DAFTAR TABEL

Tabel III-1 Sampel <i>tweet</i> yang digunakan .....	III-1
Tabel III-2. Hasil Pemodelan Topik Menggunakan BERTopic pada data <i>training</i> .....	III-5
Tabel III-3. Hasil Pengujian .....	III-6
Tabel IV-1. Kebutuhan Fungsional Perangkat Lunak.....	IV-2
Tabel IV-2. Kebutuhan Non-Fungsional Perangkat Lunak .....	IV-3
Tabel IV-3. Tabel Definisi Aktor .....	IV-8
Tabel IV-4. Definisi Use Case .....	IV-8
Tabel IV-5. Skenario Use Case Melakukan Proses Pre-processing Data Masukan .....	IV-9
Tabel IV-6. Skenario <i>Use Case</i> Melakukan Pemodelan Topik .....	IV-10
Tabel IV-7. Keterangan Implementasi Kelas.....	IV-19
Tabel IV-8. Rencana Pengujian <i>Use Case</i> Proses Pra-Pengolahan Data.....	IV-22
Tabel IV-9. Rencana Pengujian <i>Use Case</i> Proses Pemodelan Topik Menggunakan BERTopic.....	IV-22
Tabel IV-10. Pengujian <i>Use Case</i> Proses Pemodelan Topik Menggunakan BERTopic.....	IV-23
Tabel IV-11. Pengujian <i>Use Case</i> Proses Pemodelan Topik Menggunakan BERTopic.....	IV-23
Tabel V-1. Hasil Pemodelan Topik Menggunakan BERTopic pada data <i>training</i> dengan distribusi jumlah <i>tweet</i> terbanyak .....	V-1
Tabel V-2. Hasil evaluasi pemodelan topik menggunakan BERTopic pada data <i>training</i> dengan 5 <i>coherence score cv</i> tertinggi .....	V-3
Tabel V-3. Hasil evaluasi pemodelan topik menggunakan BERTopic pada data <i>testing</i> dengan 5 <i>coherence score cv</i> tertinggi .....	V-3

# **BAB I PENDAHULUAN**

## **1.1 Pendahuluan**

Dalam bab ini, akan dibicarakan mengenai beberapa hal penting terkait penelitian, seperti latar belakang permasalahan, perumusan masalah, tujuan penelitian, manfaat penelitian, dan juga batasan-batasan yang ada. Selain itu, bab ini juga akan memuat penjelasan secara umum mengenai keseluruhan isi penelitian.

## **1.2 Latar Belakang Masalah**

Pada era penyebaran informasi yang semakin maju dan berkembang, kebutuhan untuk mengelola dan menganalisis teks yang besar dan kompleks semakin mendesak. Data teks yang berasal dari berbagai sumber seperti media sosial, berita online, riset akademik, dan lainnya, memiliki potensi besar untuk memberikan wawasan yang berharga. Pertumbuhan penggunaan media sosial, khususnya Twitter telah menghasilkan informasi dalam bentuk data teks dalam jumlah besar yang mencakup berbagai topik dan isu.

Pemodelan topik adalah salah satu inovasi terbaru dalam pengolahan data teks untuk menemukan topik pada sekumpulan teks. Inovasi ini digunakan dalam pendekatan penambangan dan analisis teks untuk menemukan informasi tersembunyi dan hubungan antar teks (Jelodar et al., 2018). Salah satu metode yang dapat digunakan untuk pemodelan topik adalah BERTopic. BERTopic

merupakan suatu pendekatan pemodelan topik yang memanfaatkan integrasi BERT dan pengukuran TF-IDF berbasis kelas untuk menciptakan representasi topik yang koheren. Pendekatan ini juga memanfaatkan teknik *Uniform Manifold Approximation and Projection* (UMAP) untuk mengurangi dimensi penyematan sebelum melakukan pengelompokan dokumen-dokumen. BERTopic merupakan metode pemodelan topik yang menggunakan *transformer* dan c-TF-IDF untuk membentuk cluster padat yang memudahkan interpretasi topik sambil tetap mempertahankan kata-kata kunci dalam deskripsi topik (Grootendorst, 2022). Penggunaan metode c-TF-IDF pada BERTopic, dapat memfasilitasi berbagai pendekatan pemodelan topik, seperti pemodelan topik yang terpadu, dinamis, atau berbasis kelas (Abuzayed and Al-Khalifa, 2021). Keunggulan utama BERTopic adalah kemampuannya untuk memberikan performa yang baik pada sebagian besar aspek domain pemodelan topik, sementara algoritma lainnya cenderung unggul dalam satu aspek tertentu (Egger and Yu, 2022).

Baru-baru ini BERTopic melakukan peningkatan lebih lanjut pada topik guna menciptakan label, ringkasan dari topik, dan hal-hal lainnya. *Large Language Models* (LLMs) adalah algoritma deep learning yang mampu mengidentifikasi, merangkum, menerjemahkan, memprediksi, dan menciptakan konten menggunakan dataset yang sangat besar. *Large Language Models* (LLMs) adalah bentuk kecerdasan buatan (AI) yang sangat efektif untuk berbagai tugas, seperti *machine translation* and *question answering* (Hadi et al., 2023). Manfaat utama dari pendekatan ini adalah dapat menjelaskan suatu topik hanya dengan



menggunakan sedikit dokumen, sehingga tidak perlu meneruskan seluruh dokumen ke model pembuatan teks.

Berdasarkan data dari [www.databoks.katadata.co.id](http://www.databoks.katadata.co.id), hingga April 2023, pengguna Twitter di dunia mencapai 372,9 juta, termasuk 14,75 juta pengguna berasal dari Indonesia. Indonesia menempati posisi ke-6 sebagai negara dengan pengguna Twitter terbanyak. Sehingga melakukan analisis data Twitter warga Indonesia berpotensi menemukan wawasan penting tentang topik-topik yang relevan dengan masyarakat Indonesia. Namun, untuk memanfaatkan potensi ini, diperlukan alat dan teknik yang canggih untuk mengidentifikasi pola-pola tersembunyi dan mengelompokkan teks berdasarkan topiknya.

Berdasarkan permasalahan yang telah disampaikan sebelumnya bahwa kebutuhan untuk pengambilan topik pada teks yang ada, dapat dilakukan dengan menggunakan model BERTopic dengan LLAMA2-7b sebagai *topic representation tuning*. LLAMA2-7b digunakan untuk menghasilkan deskripsi label dari sekumpulan kata kunci yang dihasilkan dari perhitungan c-TF-IDF (Grootendorst, 2022). LLAMA2-7b merupakan *open source pretrained and fine-tuned Large Language Models* (LLMs) yang dikembangkan oleh META dengan 7 miliar parameter (Touvron et al., 2023). LLAMA2-7b dipilih karena memiliki kemampuan menghasilkan teks yang berkualitas tinggi dengan ukuran yang cukup kecil dibanding *Large Language Models* (LLMs) serupa dengan ukuran parameter yang lebih besar. Oleh karena itu, peneliti bertujuan untuk menyelesaikan masalah ini melalui penelitian yang berjudul “**TOPIC MODELING MENGGUNAKAN**

## **BERTOPIC DENGAN LLAMA2 SEBAGAI TOPIC REPRESENTATION TUNING”.**

### **1.3 Rumusan Masalah**

Berikut adalah rumusan masalah dalam penelitian ini:

1. Bagaimana mengembangkan sistem topic modeling menggunakan BERTopic dengan LLAMA2 sebagai *topic representation tuning*?
2. Bagaimana kinerja model BERTopic dengan LLAMA2 sebagai *topic presentation tuning* dalam menemukan topik berdasarkan *coherence score*?

### **1.4 Tujuan Penelitian**

Berikut adalah tujuan dari penelitian ini:

1. Menghasilkan sistem yang dapat memodelkan topik pada teks menggunakan model BERTopic dengan LLAMA2 sebagai *topic presentation tuning*.
2. Mengetahui kinerja model BERTopic dengan LLAMA2 sebagai *topic presentation tuning*.

### **1.5 Manfaat Penelitian**

Berikut adalah manfaat dari penelitian ini:

1. Sistem dapat digunakan untuk menemukan topik pada teks menggunakan model BERTopic dengan LLAMA2 sebagai *topic presentation tuning*.

2. Hasil penelitian dapat digunakan sebagai acuan untuk penelitian yang terkait.

### **1.6 Batasan Masalah**

Penelitian ini memiliki batasan sebagai berikut:

1. Data yang digunakan adalah *tweet* atau cuitan para pengguna Twitter.
2. Data *tweet* terbatas pada akun berita dari akun berita detikcom.
3. Data yang digunakan merupakan *tweet* atau cuitan yang menggunakan Bahasa Indonesia.

### **1.7 Sistematika Penulisan**

Sistematika penulisan tugas akhir mengikuti standar penulisan tugas akhir

Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut:

## **BAB I. PENDAHULUAN**

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah/ruang lingkup, metodologi, penelitian, dan sistematika penulisan yang akan dijadikan sebagai pokok pikiran penelitian ini.

## **BAB II. KAJIAN LITERATUR**

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi *topic MODELING* dan metode BERTopic dan beberapa literatur yang relevan dengan penelitian ini.

### **BAB III. METODOLOGI PENELITIAN**

Pada bab ini akan dibahas mengenai proses yang akan dilakukan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja.

### **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Bab ini membahas mengenai analisis dan rancangan perangkat lunak yang akan dibangun. Tahapan pengembangan perangkat lunak dimulai dengan analisis kebutuhan, perancangan perangkat lunak dan pengujian.

### **BAB V. HASIL DAN ANALISIS PENELITIAN**

Dalam bagian ini, hasil pengujian akan dipaparkan sesuai dengan rencana langkah-langkah yang telah disusun sebelumnya. Analisis akan disajikan sebagai dasar untuk menarik kesimpulan dari penelitian ini.

### **BAB VI. KESIMPULAN DAN SARAN**

Bab ini berfokus pada rangkuman dari uraian yang telah dijelaskan pada bab sebelumnya dan memberikan rekomendasi berdasarkan hasil penelitian yang telah dilakukan.

#### **1.8 Kesimpulan**

Pada bab ini telah dijelaskan latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penelitian yang akan dijadikan sebagai pokok pikiran peneliti.

## DAFTAR PUSTAKA

- Abuzayed, A., & Al-Khalifa, H. (2021). BERT for Arabic topic modeling: an experimental study on BERTopic technique. *Procedia Computer Science*, 189, 191-194.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, 7.
- Gornik, D. (2004). IBM Rational Unified Process: Best practices for software development teams. *Rational Software White Paper TP026B*, Rev, 11(01).
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. <http://arxiv.org/abs/2203.05794>
- Hadi, M. U., Al-Tashi, Q., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Al-Garadi, M. A., Wu, J., & Mirjalili, S. (2023). Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects. <https://doi.org/10.36227/techrxiv.23589741>
- Hendry, D., Darari, F., Nurfadillah, R., Khanna, G., Sun, M., Condylis, P. C., & Taufik, N. (2021, October). Topic modeling for customer service chats.

- In 2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS) (pp. 1-6). IEEE.
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169-15211.
- Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. <http://arxiv.org/abs/1312.6114>
- McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. <http://arxiv.org/abs/1802.03426>
- Mertayasa, K. T., , I.D.M.B. Darmawan. (2022). Pemodelan Topik Pada Ulasan Hotel Menggunakan Metode BERTopic Dengan Prosedur c-TF-IDF. (n.d.).
- Nathiya, G., Punitha S. C., Punithavalli, M. (2010). An Analytical Study on Behavior of Clusters Using K Means, EM and K\* Means Algorithm. <http://sites.google.com/site/ijcsis/>
- Putra, I. M. K. B. (2017). Analisis topik informasi publik media sosial di surabayamenggunakan pemodelan latent dirichlet allocation (LDA) (Doctoral dissertation, Institut Teknologi Sepuluh Nopember)
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. <http://arxiv.org/abs/1908.10084>

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., ... Scialom, T. (2023). *Llama 2: Open Foundation and Fine-Tuned Chat Models*. <http://arxiv.org/abs/2307.09288>

Wu, X., Nguyen, T., & Luu, A. T. (2023). *A Survey on Neural Topic Models: Methods, Applications, and Challenges*. <https://doi.org/10.21203/rs.3.rs-3049182/v1>

Zakeshia Tiara Kannitha, D., & Kartikasari, P. (2022). *PEMODELAN TOPIK PADA KELUHAN PELANGGAN MENGGUNAKAN ALGORITMA LATENT DIRICHLET ALLOCATION DALAM MEDIA SOSIAL TWITTER*. *11(2)*, 266–277. <https://ejournal3.undip.ac.id/index.php/gaussian/>