

PEMODELAN TOPIK MENGGUNAKAN *BERTopic* DENGAN
KeyBERT UNTUK EKSTRAKSI KATA KUNCI SEBAGAI TOPIC
REPRESENTATION TUNING

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



Oleh:

Erlangga Nicholas Purba
NIM: 09021282025049

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024

LEMBAR PENGESAHAN SKRIPSI

PEMODELAN TOPIK MENGGUNAKAN *BERTopic* DENGAN *KeyBERT* UNTUK EKSTRAKSI KATA KUNCI SEBAGAI TOPIC REPRESENTATION TUNING

Oleh:

Erlangga Nicholas Purba
NIM: 09021282025049

Indralaya, 31- Mei-2024

Pembimbing I



Novi Yushiana, M.T.
NIP. 198211082012122001

Pembimbing II,



Anggina Pramanita, S.Kom., M.IT.
NIP. 198908062015042002

Mengetahui,

Ketua Jurusan Teknik Informatika



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

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Jumat tanggal 31 Mei 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Erlangga Nicholas Purba

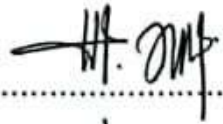
NIM : 09021282025049

Judul : PEMODELAN TOPIK MENGGUNAKAN *BERTopic* DENGAN *KeyBERT* UNTUK EKSTRAKSI KATA KUNCI SEBAGAI TOPIC REPRESENTATION TUNING

Dan dinyatakan **LULUS**.


1. Ketua

Annisa Darmawahyuni, M.Kom.
NIP. 199006302023212044



2. Penguji I

Dr. Abdiansah, S.Kom., M.Cs.
NIP 198410012009121005



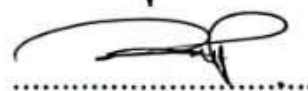
3. Pembimbing I

Novi Yusliani, M.T.
NIP 198211082012122001



4. Pembimbing II

Anggina Pramanita, S.Kom., M.IT.
NIP 198908062015042002

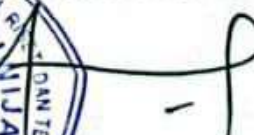


Mengetahui,

Ketua Jurusan Teknik Informatika



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



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Erlangga Nicholas Purba

NIM : 09021282025049

Program Studi : Teknik Informatika

Judul Skripsi :PEMODELAN TOPIK MENGGUNAKAN *BERT*Topic
DENGAN *KeyBERT* UNTUK EKSTRAKSI KATA
KUNCI SEBAGAI *TOPIC REPRESENTATION TUNING*

Hasil pengecekan *Software iThenticate/Turnitin*: 13%

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.

di Palembang, 31-05-2024



Erlangga Nicholas Purba
NIM. 09021282025049

MOTTO DAN PERSEMBAHAN

Motto:

Bersabarlah dalam kesakitan dalam hidup, tetapi yang paling penting bersabar
dengan emosi diri sendiri.

“Sabarmaho disasude gulmit ni ngolumu , alai na paling petting ima bersabar
dohot emosi na adong didirim.”

Kupersembahkan Karya tulis ini kepada:

- Tuhan Yang Maha Esa
- Kedua orang tua dan adik saya
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

In the era of technological advancement, the use of social media such as Twitter has become commonplace as a medium for online interaction. Every day, a vast number of tweets are generated by Twitter users from around the world. To determine which topics are trending, reading all the tweets on Twitter would take an extremely long time due to the sheer volume of tweets. One method used to efficiently extract information from Twitter tweets is topic modeling. Topic modeling is a method for discovering topics from various texts. This research aims to perform topic modeling on Indonesian-language tweets using BERTopic with KeyBERT for keyword extraction in each topic. KeyBERT will generate keywords for each topic cluster and will be used by BERTopic to enrich the results of the topic modeling. The dataset used consists of 10,000 Indonesian-language tweets taken from the Twitter account @detikcom. The data is divided into two parts: 8,000 tweets are used for training data and 2,000 tweets are used for testing. Based on the topic modeling results with BERTopic, a total of 50 topics were obtained. Topic Modeling evaluation was conducted using coherence score, yielding an average of 0.765 on the training data and 0.675 on the testing data.

Keywords: *Topic Modeling, BERTopic, KeyBERT, Coherence Score , Twitter*

ABSTRAK

Pada era kemajuan teknologi, penggunaan media sosial seperti twitter menjadi hal biasa sebagai media untuk berinteraksi secara *online*. Setiap harinya, cuitan dengan jumlah yang sangat besar dihasilkan oleh pengguna twitter dari seluruh dunia. Untuk mengetahui topik apa saja yang sedang hangat diperbincangkan, membaca keseluruhan cuitan yang ada di twitter akan membutuhkan waktu yang sangat lama karena jumlah cuitan yang sangat besar. Salah satu metode yang digunakan untuk mendapatkan informasi dari cuitan twitter dengan lebih efisien adalah pemodelan topik. Pemodelan topik adalah sebuah metode untuk menemukan topik dari berbagai teks. Penelitian ini bertujuan untuk melakukan pemodelan topik pada *tweet* berbahasa Indonesia dengan menggunakan BERTopic dengan *KeyBERT* untuk ekstraksi kata kunci di setiap topik. *KeyBERT* akan menghasilkan kata kunci pada setiap kelompok topik dan akan digunakan oleh *BERTopic* untuk memperkaya hasil dari pemodelan topik. 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 50 total topik. Evaluasi Pemodelan Topik dilakukan menggunakan *coherence score*, diperoleh rata-rata 0.765 pada data *training* dan 0.675 pada data *testing*.

Kata Kunci: Pemodelan Topik, *BERTopic*, *KeyBERT*, *Coherence Score*, *Twitter*

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa 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. Tuhan Yang Maha Esa 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 Ibu Anggina Primanita, S.Kom., M.IT. selaku dosen Pembimbing II yang telah membimbing, memberikan motivasi serta arahan kepada penulis dalam proses pengerjaan skripsi.
5. Bapak Abdiansah, S.Kom., M.Cs. selaku Dosen Penguji Tugas Akhir yang telah memberikan ilmu, nasihat serta saran yang membangun.

6. Seluruh dosen, staf dan pegawai Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Teman-teman penulis yang telah memberikan saran, motivasi, dan, semangat selama mengerjakan skripsi ini
8. 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.

Inderalaya, 27 Mei 2024

Penulis

Erlangga Nicholas Purba

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
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-3
1.4 Tujuan Penelitian	I-3
1.5 Manfaat Penelitian	I-3
1.6 Batasan Masalah	I-3
1.7 Sistematika Penulisan	I-4
1.8 Kesimpulan	I-5
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-2
2.2.3 <i>Sentence-BERT</i>	II-4
2.2.4 <i>Principal Component Analysis (PCA)</i>	II-5
2.2.5 K-Means	II-6
2.2.6 <i>CountVectorizer</i>	II-7
2.2.7 <i>Class Based TF-IDF (c-TF-IDF)</i>	II-8
2.2.8 <i>KeyBERT</i>	II-9
2.2.9 <i>Coherence Score</i>	II-10
2.2.10 <i>Rational Unified Process (RUP)</i>	II-11

2.2.11	Penelitian Lain yang Relevan.....	II-13
2.3	Kesimpulan.....	II-14
BAB III	METODOLOGI PENELITIAN	III-1
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-4
3.3.2	Menentukan Kriteria Pengujian	III-5
3.3.3	Format Data Pengujian.....	III-5
3.3.4	Alat yang Digunakan Dalam Pelaksanaan Penelitian	III-7
3.3.5	Membangun Perangkat Lunak Penelitian	III-7
3.3.6	Melakukan Pengujian Penelitian.....	III-7
3.3.7	Melakukan Analisis dan Menarik Kesimpulan Penelitian.....	III-8
3.4	Metode Pengembangan Perangkat Lunak	III-8
3.4.1	Fase Insepsi	III-8
3.4.2	Fase Elaborasi	III-9
3.4.3	Fase Konstruksi.....	III-9
3.4.4	Fase Transisi	III-10
3.5	Manajemen Proyek Penelitian.....	III-11
3.6	Kesimpulan.....	III-15
BAB IV	PENGEMBANGAN PERANGKAT LUNAK.....	IV-1
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-2
4.2.3.1	Analisis Kebutuhan Perangkat Lunak.....	IV-2
4.2.3.2	Analisis Proses Pemodelan Topik.....	IV-3
4.2.3.3	Analisis Hasil Pemodelan Topik.....	IV-5
4.2.4	Implementasi	IV-6
4.3	Fase Elaborasi.....	IV-10
4.3.1	Pemodelan Bisnis	IV-10

4.3.1.1	Perancangan Data	IV-10
4.3.1.2	Perancangan User Interface	IV-10
4.3.2	Kebutuhan	IV-12
4.3.3	Analisis dan Perancangan	IV-12
4.3.3.1	Diagram Aktivitas	IV-12
4.3.3.2	Diagram Alur	IV-14
4.4	Fase Konstruksi	IV-15
4.4.1	Kebutuhan Sistem	IV-16
4.4.2	Implementasi	IV-16
4.4.2.1	Implementasi Kelas	IV-17
4.4.2.2	Implementasi Antarmuka	IV-17
4.5	Fase Transisi	IV-19
4.5.1	Pemodelan Bisnis	IV-20
4.5.2	Kebutuhan	IV-20
4.5.3	Analisis dan Perancangan	IV-20
4.5.3.1	Rencana Pengujian	IV-21
4.5.3.2	Implementasi	IV-21
4.6	Kesimpulan	IV-22
BAB V	HASIL DAN ANALISIS	V-1
5.1	Pendahuluan	V-1
5.2	Hasil Penelitian	V-1
5.3	Analisis Hasil Penelitian	V-6
5.4	Kesimpulan	V-11
BAB VI	KESIMPULAN DAN SARAN	VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan	VI-1
6.3	Saran	VI-2
DAFTAR PUSTAKA	iii
LAMPIRAN	v

DAFTAR GAMBAR

Gambar II-1 Arsitektur <i>BERTopic</i> dan <i>KeyBERT</i> (Grootendorst,2022).....	II-3
Gambar II-2 Diagram proses <i>Sentence-BERT</i> (Reimers and Gurevych, 2019) ..	II-4
Gambar II-3 Proses <i>CountVectorizer</i> (Shah et al., 2020).....	II-8
Gambar II-4 Diagram Proses <i>KeyBERT</i> (Grootendorst, 2022).....	II-10
Gambar II-5 Arsitektur RUP (Anwar, 2014)	II-12
Gambar III-1 Rincian Kegiatan Penelitian.....	III-3
Gambar III-2 Diagram Alur Proses Umum Perangkat Lunak	III-4
Gambar IV-1. Keluaran Proses <i>Document Embedding</i> Menggunakan SBERT	IV-3
Gambar IV-2. Hasil Proses <i>Dimensionality Reduction</i> Menggunakan PCA	IV-4
Gambar IV-3. Visualisasi Keluaran Proses <i>Topic Representation</i> Menggunakan c-TF-IDF	IV-5
Gambar IV-4. <i>Use Case</i> Pemodelan Topik.....	IV-6
Gambar IV-5. Rancangan Antarmuka Halaman Utama Pemodelan Topik	IV-11
Gambar IV-6. Rancangan Antarmuka Halaman Hasil Pemodelan Topik	IV-11
Gambar IV-7. Diagram Aktivitas Melakukan <i>Pre-processing</i> Data Pada Sistem	IV-13
Gambar IV-8. Diagram Aktivitas Melakukan Pemodelan Topik	IV-13
Gambar IV-9. Diagram Alur Proses <i>Pre-processing</i> Data.....	IV-14
Gambar IV-10. Diagram Alur Proses Pemodelan Topik	IV-15
Gambar IV-11. Diagram Kelas Perangkat Lunak	IV-16
Gambar IV-12. Implementasi Antarmuka Pra-pengolahan Data.....	IV-18
Gambar IV-13. Implementasi Antarmuka Hasil Pemodelan Topik.....	IV-19

DAFTAR TABEL

Tabel III-1 Sampel <i>tweet</i> yang digunakan	III-1
Tabel III-2 Hasil Pengujian dengan <i>BERTopic</i> dan <i>KeyBERT</i>	III-6
Tabel III-3 Hasil evaluasi dengan <i>BERTopic</i> dan <i>KeyBERT</i>	III-6
Tabel III-4 Work Breakdown Structure	III-11
Tabel IV-1. Kebutuhan Fungsional Sistem.....	IV-2
Tabel IV-2. Kebutuhan Non-Fungsional Sistem.....	IV-2
Tabel IV-3. Tabel Definisi Aktor.....	IV-7
Tabel IV-4. Definisi <i>Use Case</i>	IV-7
Tabel IV-5. Skenario <i>Use Case</i> Melakukan Proses <i>Pre-processing</i> Data Masukan	IV-7
Tabel IV-6. Skenario <i>Use Case</i> Melakukan Pemodelan Topik	IV-8
Tabel IV-7. Keterangan Implementasi Kelas.....	IV-17
Tabel IV-8. Rencana Pengujian <i>Use Case</i> Proses Pra-Pengolahan Data.....	IV-21
Tabel IV-9. Rencana Pengujian <i>Use Case</i> Proses Pemodelan Topik	IV-21
Tabel IV-10. Pengujian <i>Use Case</i> Proses Pemodelan Topik	IV-22
Tabel IV-11. Pengujian <i>Use Case</i> Proses Pemodelan Topik	IV-22
Tabel V-1. Contoh Hasil Pemodelan Topik Pada Data <i>Training</i> Berdasarkan 5 Jumlah Tweet Terbanyak	V-1
Tabel V-2. Contoh Hasil Evaluasi 5 <i>Coherence Score</i> Tertinggi Pada Pemodelan Topik Dengan Data <i>Training</i>	V-2
Tabel V-3. Contoh Hasil Evaluasi 5 <i>Coherence Score</i> Terendah Pada Pemodelan Topik Dengan Data <i>Training</i>	V-3
Tabel V-4. Contoh Hasil Evaluasi 5 <i>Coherence Score</i> Tertinggi Pada Pemodelan Topik Dengan Data <i>Testing</i>	V-4
Tabel V-5. Contoh Hasil Evaluasi 5 <i>Coherence Score</i> Terendah Pada Pemodelan Topik Dengan Data <i>Testing</i>	V-5
Tabel V-6. Contoh 5 Dokumen Dengan Topik 17 Yang Memiliki Nilai <i>Coherence Score</i> Tertinggi Pada Data <i>Testing</i>	V-6
Tabel V-7. Contoh 5 Dokumen Dengan Topik 8 Yang Memiliki Nilai <i>Coherence Score</i> Terendah Pada Data <i>Testing</i>	V-8
Tabel V-8. Penelitian Tentang Pemodelan Topik Yang Menggunakan <i>Coherence Score</i> Pada Tahapan Evaluasi	V-10

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 kemajuan teknologi, penggunaan media sosial seperti twitter menjadi hal biasa sebagai media untuk berinteraksi secara online. Setiap harinya, cuitan dengan jumlah yang sangat besar dihasilkan oleh pengguna twitter dari seluruh dunia. Untuk mengetahui topik apa saja yang sedang hangat diperbincangkan, membaca keseluruhan cuitan yang ada di twitter akan membutuhkan waktu yang sangat lama karena jumlah cuitan yang sangat besar.

Salah satu metode yang digunakan untuk mendapatkan informasi dari cuitan twitter dengan lebih efisien adalah pemodelan topik. Pemodelan topik adalah sebuah metode untuk menemukan topik dari berbagai teks. Metode ini digunakan dalam pendekatan dan analisis teks untuk mencari informasi tersembunyi dan hubungan antar teks (Jelodar et al., 2018). Salah satu metode pemodelan topik yang bisa digunakan untuk pemodelan topik adalah *BERTopic*. *BERTopic* merupakan suatu metode baru yang diciptakan oleh Maarten Grootendorst pada tahun 2022.

Pemodelan topik yang dilakukan oleh *BERTopic* menggunakan integrasi dari BERT dengan c-TF-IDF yang berbasis sistem kelas untuk mendapatkan topik yang koheren. Pada prosesnya, *BERTopic* menggunakan *Principal Component Analysis* (PCA) dan K-Means. PCA berfungsi untuk mengurangi dimensi penyematanan dan K-Means untuk mengklusterisasi dokumen. *BERTopic* menggunakan c-TF-IDF dan transformator untuk membentuk kluster data sehingga mempermudah proses mempertahankan kata kunci pada saat pencarian topik berlangsung (Grootendorst, 2022). *BERTopic* bekerja sangat baik dengan *embedding* yang telah dilatih sebelumnya karena adanya pemisahan antara pengelompokan dokumen dan penggunaan c-TF-IDF untuk mengekstrak representasi topik. (Egger and Yu, 2022).

BERTopic dapat dikombinasikan dengan metode ekstraksi kata kunci. Ekstraksi kata kunci adalah teknik analisis teks yang secara otomatis mengekstraksi kata-kata dan ungkapan yang paling sering digunakan dan penting dari suatu teks. Teknik ini membantu merangkum isi teks dan mencari topik utama yang dibahas (Hasan et al., 2017). *KeyBERT* adalah teknik ekstraksi kata kunci yang sederhana dan mudah digunakan karena teknik ini menggunakan *BERT* untuk membuat kata kunci dan frasa kunci yang paling mirip dengan dokumen (Grootendorst 2020). Karena itu, penelitian ini bertujuan untuk mengembangkan “**PEMODELAN TOPIK MENGGUNAKAN *BERTopic* DENGAN *KeyBERT* UNTUK EKSTRAKSI KATA KUNCI SEBAGAI *TOPIC REPRESENTATION TUNING*”.**

1.3 Rumusan Masalah

Berikut adalah rumusan masalah dalam penelitian ini:

1. Bagaimana mengembangkan pemodelan topik menggunakan *BERTopic* dan *KeyBERT*?
2. Bagaimana kinerja dari *BERTopic* dan *KeyBERT* dalam membuat topik berdasarkan *coherence score*?

1.4 Tujuan Penelitian

Berikut adalah tujuan dari penelitian ini:

1. Menghasilkan pemodelan topik dengan *BERTopic* dan *KeyBERT*.
2. Mengetahui kinerja model *BERTopic* dan *KeyBERT* berdasarkan *coherence score*.

1.5 Manfaat Penelitian

Berikut adalah manfaat dari penelitian ini:

1. Sistem dapat dipakai untuk memodelkan topik pada teks menggunakan *BERTopic*.
2. Sistem dapat menghasilkan kata kunci di setiap topik yang dihasilkan.

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 pemodelan topik dan metode *BERTopic* serta 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

- Anwar, A. (2014). A Review of RUP (Rational Unified Process). In *Ashraf Anwar International Journal of Software Engineering (IJSE)* (Issue 5).
- Arora, S., Ge, R., & Moitra, A. (2012). *Learning Topic Models - Going beyond SVD*. <http://arxiv.org/abs/1204.1956>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Fina Nasari, C. J. (2016). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat. *Cogito Smart Journal/VOL. 2/NO. 2/DESEMBER 2016*, 108-119.
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-
- Hadiat, A. R. (n.d.). *Topic Modeling Evaluations: The Relationship Between Coherency and Accuracy*.
- Hasan, H. M. M., Sanyal, F., Chaki, D., & Ali, M. H. (2017). An empirical study of important keyword extraction techniques from documents. *Proceedings - 1st International Conference on Intelligent Systems and Information Management, ICISIM 2017, 2017-January*, 91–94. <https://doi.org/10.1109/ICISIM.2017.8122154>
- Koloski, B., Pollak, S., Škrlić, B., & Martinc, M. (2022). *Out of Thin Air: Is Zero-Shot Cross-Lingual Keyword Detection Better Than Unsupervised?* <http://arxiv.org/abs/2202.06650>
- Kruchten, P. (2001). *The Rational Unified Process-An Introduction*. <https://www.researchgate.net/publication/220018149>
- Medvecki, D. (n.d.). *Multilingual transformer and BERTopic for short text topic modeling: The case of Serbian*.
- Mertayasa, K. T., , I.D.M.B. Darmawan. (2022). Pemodelan Topik Pada Ulasan Hotel Menggunakan Metode BERTopic Dengan Prosedur c-TF-IDF. (n.d.).
- Mishra, S., Sarkar, U., Taraphder, S., Datta, S., Swain, D., & Saikhom, R. et al. (2017). Multivariate Statistical Data Analysis- Principal Component Analysis (PCA). *International Journal of Livestock Research*, 7(5), 60-78. <http://dx.doi.org/10.5455/ijlr.20170415115235>
- Putra, I. M. K.B. (2017). Analisis topik informasi publik media sosial di surabaya menggunakan pemodelan latent dirichlet allocation (LDA) (Doctoral dissertation, Institut Teknologi Sepuluh Nopember)

- Reimers, N., & Gurevych, I. (2020). *Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation*. <http://arxiv.org/abs/2004.09813>
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, 399–408. <https://doi.org/10.1145/2684822.2685324>
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., & Lin, C. T. (2017). A review of clustering techniques and developments. *Neurocomputing*, 267, 664–681. <https://doi.org/10.1016/j.neucom.2017.06.053>
- Shah, S. R., Kaushik, A., Sharma, S., & Shah, J. (2020). Opinion-mining on marglish and devanagari comments of youtube cookery channels using parametric and non-parametric learning models. *Big Data and Cognitive Computing*, 4(1), 1–19. <https://doi.org/10.3390/bdcc4010003>
- Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, 94. <https://doi.org/10.1016/j.is.2020.101582>