

PEMODELAN TOPIK MENGGUNAKAN METODE *LATENT*
DIRICHLET ALLOCATION DAN *GIBBS SAMPLING*

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



Oleh :

Rizki Ramadandi
NIM : 09021381722133

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021

PEMODELAN TOPIK MENGGUNAKAN METODE *LATENT*
DIRICHLET ALLOCATION DAN *GIBBS SAMPLING*

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



Oleh :

Rizki Ramadandi
NIM : 09021381722133

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021

LEMBAR PENGESAHAN SKRIPSI

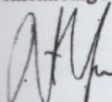
PEMODELAN TOPIK MENGGUNAKAN METODE *LATENT DIRICHLET ALLOCATION* DAN *GIBBS SAMPLING*

Oleh :

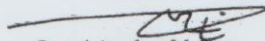
Rizki Ramadandi
NIM : 09021381722133

Palembang, Desember 2021

Pembimbing I


Novi Yusliani, M. T.
NIP 198211082012122001

Pembimbing II,


Osvari Arsalan, M. T.
NIP 198806282018031001

Mengetahui,
Ketua Jurusan Teknik Informatika




Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Jum'at tanggal 17 Desember 2021 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Rizki Ramadandi
NIM : 09021381722133
Judul : *Pemodelan Topik Menggunakan Metode Latent Dirichlet Allocation dan Gibbs Sampling*

1. Pembimbing I

Novi Yustiani, M. T.
NIP 198211082012122001

2. Pembimbing II

Osvari Arsalan, M. T.
NIP 198806282018031001

3. Penguji I

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

4. Penguji II

Mastura Diana Marieska, M.T.
NIP 198603212018032001

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

HALAMAN PERNYATAAN

Yang Bertanda tangan dibawah ini :

Nama : Rizki Ramadandi

NIM : 09021381722133

Program Studi : Teknik Informatika

Judul Skripsi : Pemodelan Topik Menggunakan Metode Latent Dirichlet Allocation dan Gibbs Sampling

Checking Result iThenticate / Turnitin Software : 18%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek 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 oleh siapapun.

Palembang, Desember 2021



Rizki Ramadandi

NIM. 09021381722133

MOTTO DAN PERSEMBAHAN

Motto

"Jika kamu tidak menghargai waktu kamu, begitu juga dengan hal hal yang lain dalam hidupmu. Jadi berhentilah untuk membuang buang waktu dan bakat kamu, mulailah untuk mengisinya."

-Kim Garst

"You can have the results you want, or you can have excuses. You can't have both."

-Clyde Lee Dennis

Kupersembahkan karya tulis ini kepada :

- Allah SWT. & Nabi Muhammad SAW.
- Ibu & Ayahku tercinta
- Keluarga besarku
- Dosen pembimbing & dosen penguji
- Sahabat & teman seperjuanganku
- Almamater

TOPIC MODELING WITH LATENT DIRICHLET ALLOCATION AND GIBBS SAMPLING

By :
Rizki Ramadandi
09021381722133

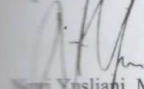
ABSTRACT

Topic modeling is a technique for identifying latent topics in a collection of texts. Topic modeling was done with the Latent Dirichlet Allocation and Gibbs Sampling methods, while software development was done with the Extreme Programming method. Six Indonesian news stories were chosen from thousands of articles collected using the Web Scraper technique from the detiknews news portal. Drugs and COVID-19 are the two primary categories in which news stories are classified. The topic coherence technique of evaluating UCI scores was used to analyze the LDA model, with the study's findings showing that five ideal topics were found in each test configuration.

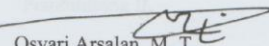
Key Word : Topic Modeling, Latent Dirichlet Allocation, Gibbs Sampling, Topic Coherence, Extreme Programming, Web Scraper, UCI

Palembang, December 2021

Supervisor I

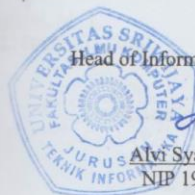

Nuri Yustiani, M. T.
NIP 198211082012122001

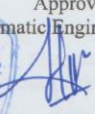
Supervisor II,


Osvari Arsalan, M. T.
NIP 198806282018031001

Approve,

Head of Informatic Engineering Department




Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

PEMODELAN TOPIK MENGGUNAKAN METODE *LATENT DIRICHLET ALLOCATION* DAN *GIBBS SAMPLING*

Oleh :

Rizki Ramadandi

09021381722133

ABSTRAK

Pemodelan topik adalah suatu alat yang digunakan untuk menemukan topik laten pada sekelompok dokumen. Pada penelitian ini dilakukan pemodelan topik dengan menggunakan metode *Latent Dirichlet Allocation* dan *Gibbs Sampling* dan pengembangan perangkat lunak dengan menggunakan metode *Extreme Programming*. Enam artikel berita Bahasa Indonesia dipilih dari ribuan artikel berita yang telah dikumpulkan dari portal berita detiknews dengan menggunakan metode *Web Scraper*. Artikel berita dibagi menjadi dua kategori utama yaitu, narkoba dan COVID-19. Analisis model LDA dilakukan dengan menggunakan metode koherensi topik pengukuran skor UCI dengan hasil penelitian menyebutkan diperoleh lima buah topik optimal pada masing-masing pengujian.

Kata Kunci : Pemodelan Topik, *Latent Dirichlet Allocation*, *Gibbs Sampling*, *Extreme Programming*, *Web Scraper*, Koherensi Topik, UCI

Palembang, Desember 2021

Pembimbing II,

Pembimbing I

Novi Yusliani, M. T.

NIP 198211082012122001

Osvari Arsalan, M. T.

NIP 198806282018031001

Mengetahui,

Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.

NIP 197812222006042003

KATA PENGANTAR

Puji dan syukur kepada Allah atas berkat dan rahmat-Nya yang diberikan kepada penulis sehingga dapat menyelesaikan tugas akhir yang berjudul “PEMODELAN TOPIK MENGGUNAKAN METODE *LATENT DIRICHLET ALLOCATION* DAN *GIBBS SAMPLING*” dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya

Untuk selanjutnya penyusun mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian skripsi ini, yaitu :

1. Ibu Siti Hawa, S.Pd. dan Bapak Wardi Danwidiq, S.Pd., M.Si., selaku orang tua tercinta dan Edward Ade Afriansyah S. Kom., selaku saudara yang selalu memberi dukungan penuh, motivasi, dan do’a tanpa henti.
2. Bapak Jaidan Jauhari, M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Alvi Syahrini Utami, M. Kom., selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya sekaligus ketua penguji sidang komprehensif.
4. Ibu Novi Yusliani, M.T., selaku pembimbing Tugas Akhir dan Bapak Osvari Arsalan, M.T., selaku pembimbing akademik sekaligus pembimbing Tugas

Akhir yang telah banyak memberikan bimbingan, arahan, dan masukan selama perkuliahan sekaligus selama proses penyelesaian Tugas Akhir ini.

5. Bapak Dr. Abdiansah, S.Kom., M.Sc. dan Ibu Mastura Diana Marieska, M.T., selaku penguji I dan penguji II sidang komprehensif yang telah memberi banyak masukan dan saran terhadap Tugas Akhir ini.
6. Segenap staff pengajar di Fakultas Ilmu Komputer Universitas Sriwijaya yang telah mengajar, membimbing, dan memberikan ilmu kepada penulis.
7. Mbak Wiwin Juliani, S. Si, selaku staff administrasi Teknik Informatika Bilingual yang telah banyak membantu dalam hal urusan akademik dan administrasi selama perkuliahan.
8. Sahabat-sahabat seperjuangan Teknik Informatika Bilingual 2017 yang selalu saling mendukung dalam suka dan duka.
9. Serta pihak-pihak lainnya yang terlibat selama pelaksanaan Tugas Akhir ini yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari dalam penyusunan tugas akhir ini masih banyak kekurangan yang disebabkan kurangnya pengetahuan dan pengalaman. Oleh karena itu kritik dan saran sangat diterima untuk kemajuan penelitian berikutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, Desember 2021



Rizki Ramadandi

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG SKRIPSI	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN.....	xvi

BAB I PENDAHULUAN

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-4
1.7	Sistematika Penulisan	I-4
1.8	Kesimpulan	I-5

BAB II KAJIAN LITERATUR

2.1	Pendahuluan	II-1
2.2	Landasan Teori.....	II-1
2.2.1	Pemrosesan Bahasa Alami	II-1
2.2.2	<i>Text preprocessing</i>	II-2
2.2.3	Pemodelan Topik.....	II-3
2.2.4	<i>Latent Dirichlet Allocation (LDA)</i>	II-4

2.2.5	<i>Gibbs sampling</i>	II-6
2.2.6	<i>Web scraping</i>	II-7
2.2.7	<i>Extreme Programming</i>	II-8
2.2.8	Koherensi Topik	II-10
2.3	Penelitian Lain yang Relevan.....	II-12
2.4	Kesimpulan	II-13

BAB III METODOLOGI PENELITIAN

3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
3.2.1	Jenis Data	III-1
3.2.2	Sumber Data	III-2
3.2.3	Metode Pengumpulan Data	III-2
3.3	Tahapan Penelitian	III-2
3.3.1	Kerangka Kerja.....	III-3
3.3.2	Kriteria Pengujian.....	III-4
3.3.3	Format Data Pengujian	III-5
3.3.4	Alat yang Digunakan dalam Pelaksanaan Penelitian	III-5
3.3.5	Pengujian Penelitian	III-6
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan	III-7
3.4	Metode Pengembangan Perangkat Lunak.....	III-7
3.5	Manajemen Proyek Penelitian	III-8

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1	Pendahuluan	IV-1
4.2	Perencanaan Awal.....	IV-1
4.3	Iterasi Pertama.....	IV-4
4.3.1	<i>Planning</i>	IV-5
4.3.2	<i>Design</i>	IV-21
4.3.2.1	<i>Use Case Diagram</i>	IV-21
4.3.2.2	<i>Activity Diagram</i>	IV-26
4.3.2.3	<i>Sequence Diagram</i>	IV-28
4.3.2.4	<i>Class Diagram</i>	IV-31

4.3.3	<i>Coding</i>	IV-32
4.3.4	<i>Testing</i>	IV-32
4.4	Iterasi Kedua	IV-34
4.4.1	<i>Planning</i>	IV-34
4.4.2	<i>Design</i>	IV-35
4.4.3	<i>Coding</i>	IV-38
4.4.4	<i>Testing</i>	IV-38

BAB V HASIL DAN ANALISIS PENELITIAN

5.1	Pendahuluan	V-1
5.2	Data Hasil Percobaan/Penelitian	V-1
5.2.1	Konfigurasi Percobaan	V-1
5.2.2	Data Hasil Konfigurasi 1	V-2
5.2.3	Data Hasil Konfigurasi 2	V-4
5.3	Analisis Hasil Penelitian	V-7
5.3.1	Konfigurasi 1	V-7
5.3.2	Konfigurasi 2	V-10
5.4	Kesimpulan	V-13

BAB VI KESIMPULAN DAN SARAN

6.1	Kesimpulan	VI-1
6.2	Saran	VI-1

DAFTAR PUSTAKA	xvii
----------------------	------

LAMPIRAN	xx
----------------	----

DAFTAR TABEL

	Halaman
Tabel II-1. Contoh <i>case folding</i>	II-2
Tabel II-2. Contoh <i>tokenizing</i>	II-3
Tabel II-3. Contoh <i>stopword removal</i>	II-3
Tabel III-1. Hasil Pengujian	III-5
Tabel III-2. Daftar Kata Probabilitas Tertinggi K Optimal	III-7
Tabel III-3. Penjadwalan Penelitian Menggunakan <i>Word Breakdown Structure</i> (WBS)	III-8
Tabel IV-1. Kebutuhan Fungsional	IV-2
Tabel IV-2. Kebutuhan Non-fungsional	IV-3
Tabel IV-3 Iterasi dan Tujuannya	IV-4
Tabel IV-4. Contoh Penyeragaman Karakter	IV-6
Tabel IV-5. Contoh Tokenisasi	IV-7
Tabel IV-6. Contoh <i>Stemming</i>	IV-8
Tabel IV-7. Contoh Penghapusan <i>Stopword</i>	IV-9
Tabel IV-8. Contoh Pemberian Topik Setiap Kata Secara Acak (K=2)	IV-10
Tabel IV-9. Contoh Daftar Kata Unik	IV-11
Tabel IV-10. Frekuensi Kemunculan Kata Setiap Topik	IV-12
Tabel IV-11. Contoh Hasil Iterasi <i>Gibbs Sampling</i> (100 Iterasi)	IV-14
Tabel IV-12. Nilai Probabilitas Kata Topik 1	IV-16
Tabel IV-13. Nilai Probabilitas Kata Topik 2	IV-18
Tabel IV-14. Distribusi Topik Pada Dokumen	IV-20
Tabel IV-15. Tabel skor UCI masing-masing K	IV-21
Tabel IV-16. Definisi Aktor	IV-22
Tabel IV-17. Definisi <i>Use Case</i>	IV-23
Tabel IV-18. Skenario <i>Use Case</i> Mengisi Parameter Konfigurasi Penelitian .	IV-23
Tabel IV-19. Skenario <i>Use Case</i> Pemodelan Topik	IV-25
Tabel IV-20. Skenario <i>Use Case</i> Melihat Hasil Akhir Pemodelan Topik	IV-25

Tabel IV-21. Daftar Kelas.....	IV-31
Tabel IV-22. Pengujian Skenario Mengisi Parameter Konfigurasi Penelitian	IV-32
Tabel IV-23. Pengujian Skenario Pemodelan Topik.....	IV-33
Tabel IV-24. Pengujian Skenario Melihat Hasil Akhir Pemodelan Topik	IV-34
Tabel IV-25. Pengujian Skenario Mengisi Parameter Konfigurasi Penelitian	IV-38
Tabel IV-26. Pengujian Skenario Pemodelan Topik.....	IV-38
Tabel IV-27. Pengujian Skenario Melihat Hasil Akhir Pemodelan Topik	IV-39
Tabel V-1. Konfigurasi Parameter 1	V-2
Tabel V-2. Data Hasil Pengujian Konfigurasi 1	V-2
Tabel V-3. Konfigurasi Parameter 2	V-4
Tabel V-4. Data Hasil Pengujian Konfigurasi 2	V-5
Tabel V-5. Daftar Kata Probabilitas Tertinggi (Konfigurasi 1).....	V-7
Tabel V-6. Distribusi Topik Setiap Dokumen (Konfigurasi 1)	V-9
Tabel V-7. Daftar Kata Probabilitas Tertinggi (Konfigurasi 2).....	V-10
Tabel V-8. Distribusi Topik Setiap Dokumen (Konfigurasi 2)	V-12

DAFTAR GAMBAR

	Halaman
Gambar II-1. Representasi Pemodelan Topik	II-4
Gambar II-2. Model Grafis LDA	II-5
Gambar II-3. Ilustrasi Cara Kerja <i>Web Scraper</i>	II-8
Gambar II-4. Metode Extreme Programming (XP)	II-10
Gambar II-5. Pasangan Kata Perhitungan Skor	II-11
Gambar III-1. Tahapan Penelitian.....	III-2
Gambar IV-1. <i>Use Case Diagram</i>	IV-22
Gambar IV-2. <i>Activity Diagram</i> Mengisi Parameter Konfigurasi Penelitian ..	IV-27
Gambar IV-3. <i>Activity Diagram</i> Pemodelan Topik	IV-28
Gambar IV-4. <i>Activity Diagram</i> Melihat Hasil Akhir Pemodelan Topik	IV-28
Gambar IV-5. <i>Sequence Diagram</i> Mengisi Parameter Konfigurasi Penelitian	IV-29
Gambar IV-6. <i>Sequence Diagram</i> Pemodelan Topik	IV-30
Gambar IV-7. <i>Sequence Diagram</i> Melihat Hasil Akhir Pemodelan Topik	IV-30
Gambar IV-8. <i>Class Diagram</i>	IV-32
Gambar IV-9. Tampilan Utama	IV-35
Gambar IV-10. Tampilan Hasil.....	IV-36
Gambar IV-11. Tampilan <i>Error</i>	IV-36
Gambar IV-12. Proses Mengumpulkan Tautan Artikel Berita	IV-37
Gambar IV-13. Proses Mengumpulkan Konten Artikel Berita.....	IV-37
Gambar V-1. Grafik Skor UCI Konfigurasi 1.....	V-4
Gambar V-2. Grafik Skor UCI Konfigurasi 2.....	V-7
Gambar V-3. Grafik Skor Topik (Konfigurasi 1)	V-8
Gambar V-4. Distribusi Topik Dokumen (Konfigurasi 1).....	V-10
Gambar V-5. Grafik Skor Topik (Konfigurasi 2)	V-11
Gambar V-6. Distribusi Topik Dokumen (Konfigurasi 2).....	V-13

DAFTAR LAMPIRAN

1. Contoh Perhitungan Manual Koherensi Topik Skor UCI
2. Log Skenario Pengujian Parameter Valid
3. Log Skenario Pengujian Parameter Tidak Valid
4. Koding Pemodelan Topik
5. Koding *Web Scraper*

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab pendahuluan akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan, serta kesimpulan.

1.2 Latar Belakang Masalah

Dengan semakin banyaknya teks tidak terstruktur di internet, diperlukan sebuah alat yang dapat membantu untuk menemukan topik tersembunyi yang ada pada sekumpulan teks. Untuk menyelesaikan masalah tersebut, maka para peneliti telah mengembangkan suatu teknik yang dinamakan pemodelan topik (Alghamdi & Alfalqi, 2015). Pemodelan topik merupakan bidang penelitian Pemrosesan Bahasa Alami (PBA) yang digunakan dalam berbagai macam bidang, seperti *medical sciences, software engineering, geography, political science*, dan lain-lain. Dalam pemodelan topik, teks yang digunakan dapat berupa *e-mail*, jurnal, artikel berita, dan segala bentuk teks yang tidak terstruktur lainnya.

Dalam pemodelan topik dikenal beberapa metode yang secara umum digunakan oleh para peneliti, seperti *Latent Semantic Analytic (LSA)* dan *Latent Dirichlet Allocation (LDA)*. LSA merupakan metode yang bertujuan membentuk vektor representasi teks (Alghamdi & Alfalqi, 2015). Sedangkan metode LDA merupakan algoritma untuk mendeteksi topik melalui pemodelan probabilistik dalam sekumpulan data (Blei et al., 2003). LDA juga dapat digunakan untuk

meringkas, melakukan pengelompokan, menghubungkan maupun memproses data (Arianto & Anuraga, 2020).

Latent Dirichlet Allocation (LDA) merupakan salah satu metode populer dikalangan peneliti pemrosesan bahasa alami di bidang pemodelan topik. Meskipun begitu, LDA memiliki distribusi yang kompleks, sehingga menyebabkan perhitungan estimasi dari distribusi posterior untuk model LDA menjadi sangat sulit dilakukan secara manual (Kengken, 2014). Untuk menyelesaikan permasalahan distribusi kompleks tersebut, maka digunakan algoritma *Gibbs Sampling* (Qiu, 2014).

Pada penelitian (Gaur, 2017) dengan judul “*Topic Models As A Novel Approach To Identify Themes In Content Analysis: The Example Of Organizational Research Methods*” mendemonstrasikan LDA *Topic modeling* sebagai metode baru dalam analisis data bentuk teks. Adapun data teks yang dianalisis adalah 421 artikel yang dipublikasikan pada *Organization Research Method* (ORM) dan berhasil mengungkapkan 15 topik, dimana hasil analisis tersebut cukup sesuai dengan hasil kajian oleh manusia.

Pada penelitian yang dilakukan (Arianto & Anuraga, 2020) dengan judul “Pemodelan Topik Pengguna Twitter Mengenai Aplikasi “Ruangguru”” menggunakan LDA dan *Gibbs Sampling* pada aplikasi ruangguru, dimana hasil yang didapatkan berupa kelompok data *twitter* yang terbagi menjadi 28 buah topik.

Pada penelitian yang dilakukan (Qiu, 2014) dengan judul “*Topic words analysis based on LDA model*” menunjukkan bahwa model yang dihasilkan oleh

metode LDA dan *Gibbs Sampling* lebih baik dibandingkan dengan model TF-IDF berdasarkan perbandingan tingkat presisi kedua model tersebut.

Berdasarkan penelitian-penelitian yang telah disebutkan diatas, pada penelitian ini akan diterapkan metode *Latent Dirichlet Allocation* (LDA) dan *Gibbs Sampling* untuk memodelkan topik pada teks.

1.3 Rumusan Masalah

Berdasarkan latar belakang masalah diatas, maka rumusan masalah dari penelitian ini adalah sebagai berikut :

1. Bagaimana mengembangkan sistem pemodelan topik menggunakan metode LDA dan *Gibbs Sampling* ?
2. Bagaimana kinerja metode LDA dan *Gibbs Sampling* dalam memodelkan topik?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Menghasilkan sistem pemodelan topik dengan menggunakan metode LDA dan *Gibbs Sampling*.
2. Mengetahui kinerja metode LDA dan *Gibbs Sampling* dalam pemodelan topik.

1.5 Manfaat Penelitian

Manfaat yang didapat dari penelitian ini adalah sebagai berikut :

1. Membantu pembaca untuk mengetahui topik laten pada teks.
2. Sebagai referensi dalam penelitian di bidang *Natural Language Processing*, khususnya pada bidang pemodelan topik.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut :

1. Bahasa yang digunakan adalah Bahasa Indonesia.
2. Dokumen yang digunakan berupa artikel berita.
3. Dokumen berupa file format *.json*.

1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut :

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan dan kesimpulan.

BAB II. KAJIAN LITERATUR

Pada bab ini membahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi pemrosesan bahasa alami, *teks preprocessing*, pemodelan topik, *Latent Dirichlet Allocation (LDA)*, *Gibbs Sampling*, serta beberapa kajian literatur mengenai penelitian lain yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

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

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dibahas mengenai analisis dan perancangan perangkat lunak yang akan digunakan sebagai alat penelitian.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini, hasil pengujian berdasarkan langkah-langkah yang telah direncanakan disajikan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dari penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran berdasarkan hasil dari penelitian.

1.8 Kesimpulan

Dari pendahuluan ini, telah diuraikan secara umum pokok-pokok pikiran yang melandasi penelitian yang akan dilakukan, meliputi latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta sistematika penulisan.

DAFTAR PUSTAKA

- A. F. Octaviansyah, D. Darwis, dan A. Surahman. 2019. "SISTEM Pencarian Lokasi Bengkel Mobil Resmi Menggunakan Teknik Pengolahan Suara dan Pemrosesan Bahasa Alami." *Jurnal TEKNOINFO*, Vol. 13, No. 2 71-77.
- A. J. B. Chaney, dan D. M. Blei. 2012. "Visualizing Topic Models."
- A. Piepenbrink, A. S. Gaur. 2017. "Topic models as a novel approach to identify themes in content analysis: The example of organization research methods."
- B. W. Arianto, G. Anuraga. 2020. "Pemodelan Topik Pengguna Twitter Mengenai Aplikasi “Ruangguru”." *Jurnal ILMU DASAR*, Vol. 21 No. 2 149-154.
- D. M. Blei, A. Y. Ng, M. I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3 993-1022.
- D. Mimno, H. M. Wallach, E. Talley, M. Leenders, A. McCallum. 2011. "Optimizing Semantic Coherence in Topic Models." *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* 262-272.
- D. Newman, J.H. Lau, K. Grieser, T. Baldwin. 2010. "Automatic Evaluation of Topic Coherence." *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL* 100-108.
- D. Newman, S. Karimi, L. Cavedon. 2009. "External Evaluation of Topic Models."
- Darling, W.M. 2011. "A Theoretical and Practical Implementation Tutorial on Topic Modeling and Gibbs Sampling."
- H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, L. Zhao. 2018. "Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey."
- H. M. Wallach, I. Murray, R. Salakhutdinov, D. Mimno. 2009. "Evaluation Methods for Topic Models."
- K. B. Putra, R. P. Kusumawardani. 2017. "ANALISIS TOPIK INFORMASI Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA)." *Jurnal Teknik ITS Vol 6 No. 2*.
- K. Beck, M. Hendrickson, M. Fowler. 2001. *Planning Extreme Programming*. Boston: Pearson Education, Inc. Right and Contracts Department.

- K. Stevens, P. Kegelmeyer, D. Andrzejewski, D. Buttler. 2012. "Exploring Topic Coherence over many models and many topics." *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural* 952-961.
- Kengken, R. I. 2014. "PEMODELAN TOPIK UNTUK MEDIA SOSIAL MENGGUNAKAN LATENT DIRICHLET ALLOCATION."
- L. A. Wirasakti, R. Permadi, A. D. Hartanto, Hartatik. 2020. "Pembuatan Kata Kunci Otomatis Dalam Artikel Dengan Pemodelan Topik." *JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 4, Nomor 1* 27-31.
- M. Cendana, S. D. H. Permana. 2019. "PRA-PEMROSESAN TEKS PADA GRUP WHATSAPP UNTUK PEMODELAN TOPIK." *Jurnal Mantik Penusa Vol. 3 No.3* 107-116.
- Qiu, X. 2014. "Topic words analysis based on LDA model."
- R. Alghamdi, K. Alfalqi. 2015. "A Survey of Topic Modeling in Text Mining." (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol 6 No.1.
- R. Y. K. Lau, Y. Xia, Y. Ye. 2014. "A Probabilistic Generative Model for Mining Cybercriminal Networks from Online Social Media." *IEEE Computational intelligence magazine*.
- S. Bergamaschi, L. Po. 2011. "Comparing LDA and LSA Topic Models for Content-Based Movie Recommendation Systems."
- S. German, D. German. 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. PAMI-6, NO. 6.
- S. J. Lee, E. Brennan, L. A. Gibson, A. S. L. Tan, A. Kybert-Momjian, J. Liu, R. Hornik. 2016. "Predictive Validity of an Empirical Approach for Selecting Promising Message Topics: A Randomized-Controlled Study." *Journal of Communication*, Volume 66, Issue 3 433-453.
- Supriyatna, A. 2018. "Metode Extreme Programming Pada Pembangunan Web Aplikasi Seleksi Peserta Pelatihan Kerja." *Jurnal Teknik Informatika Vol 11 No.1*.
- T. Williams, J. Betak. 2018. "A Comparison of LSA and LDA for the Analysis of Railroad Accident Text." *Procedia Computer Science* 130 98-102.
- Turland, M. 2010. *Php-Architect's Guide to Web Scraping*.
- Y. Hu, J. Boyd-Graber, B. Satinoff, A. Smith. 2013. "Interactive topic modeling." *Mach Learn* 423-469.

Y.U. Al-khairi, Y. Wibisono, B.L. Putro. 2019. "DETEKSI TOPIK FASHION PADA TWITTER DENGAN LATENT DIRICHLET ALLOCATION."

Zulhanif, Sudartianto, B. Tantular, I. G. N. M. Jaya. 2017. "APLIKASI LATENT DIRICHLET ALLOCATION (LDA) PADA CLUSTERING DATA TEKS." *Jurnal "LOG!K@"*, Jilid 7, No. 1 46-51.