

**PEMODELAN TOPIK BERITA DI SITUS WEB UKDW
DENGAN METODE LATENT DIRICHLET ALLOCATION**

Skripsi



oleh :

JOSHUA ALBERT WIJAYA

71170197

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS KRISTEN DUTA WACANA

2023

**PEMODELAN TOPIK BERITA DI SITUS WEB UKDW
DENGAN METODE LATENT DIRICHLET ALLOCATION**

Skripsi



Diajukan kepada Fakultas Teknologi Informasi Program Studi Informatika

Universitas Kristen Duta Wacana

Sebagai salah satu syarat dalam Memperoleh gelar

Sarjana Komputer

Disusun oleh :

JOSHUA ALBERT WIJAYA

71170197

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS KRISTEN DUTA WACANA

2023

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
SKRIPSI UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Kristen Duta Wacana, saya yang bertanda tangan di bawah ini:

Nama : Joshua Albert Wijaya
NIM : 71170197
Program studi : Informatika
Fakultas : Teknologi Informasi
Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Kristen Duta Wacana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

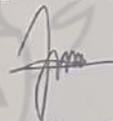
**“PEMODELAN TOPIK BERITA DI SITUS WEB UKDW DENGAN
METODE LATENT DIRICHLET ALLOCATION”**

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Kristen Duta Wacana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta
Pada Tanggal : 20 Juli 2023

Yang menyatakan



(Joshua Albert Wijaya)

NIM.71170197

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul :

PEMODELAN TOPIK BERITA DI SITUS WEB UKDW DENGAN METODE LATENT DIRICHLET ALLOCATION

Yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan sarjana Program Studi Informatika, Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika di kemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 20 Juli 2023



JOSHUA ALBERT WIJAYA
71170197

DUTA WACANA

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
SKRIPSI UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Kristen Duta Wacana, saya yang bertanda tangan di bawah ini:

Nama : Joshua Albert Wijaya
NIM : 71170197
Program studi : Informatika
Fakultas : Teknologi Informasi
Jenis Karya : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Kristen Duta Wacana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**“PEMODELAN TOPIK BERITA DI SITUS WEB UKDW DENGAN
METODE LATENT DIRICHLET ALLOCATION”**

berserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Kristen Duta Wacana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta
Pada Tanggal : 20 Juli 2023

Yang menyatakan



(Joshua Albert Wijaya)
NIM.71170197

DUTA WACANA

HALAMAN PERSETUJUAN

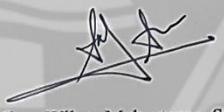
Judul : PEMODELAN TOPIK BERITA DI SITUS WEB
UKDW DENGAN METODE LATENT DIRICHLET
ALLOCATION
Nama : Joshua Albert Wijaya
NIM : 71170197
Mata Kuliah : Skripsi (Tugas Akhir)
Kode : TI0366
Semester : Genap
Tahun Akademik : 2022/2023

Telah diperiksa dan disetujui
Di Yogyakarta,
Pada Tanggal 20 Juli 2023

Dosen Pembimbing I

Dosen Pembimbing II


Dr. Phil. Lucia Dwi Krisnawati, S.S., M.A.


Aditya Wikan Mahastama, S.Kom.,
M.Cs.

DOTA WACANA

HALAMAN KELENGKAPAN PENDADARAN

**PEMODELAN TOPIK BERITA DI SITUS WEB UKDW DENGAN
METODE LATENT DIRICHLET ALLOCATION**

Mata Kuliah Pendukung : 1. Pemrosesan Bahasa Natural A

Mata Kuliah Wajib Profil : 1. Supervised Learning B+

2. Unsupervised Learning A

Profil Profesional Mandiri : ISD

NIM : 71170197

Nama Lengkap : Joshua Albert Wijaya

IPK : 3.47

No. HP : 087738284424

Email : joshua.albert@ti.ukdw.ac.id

Halaman ini digunakan sebagai salah satu persyaratan kelengkapan dan informasi
Ujian Pendadaran Skripsi

Semester Genap 2022/2023

2023

HALAMAN PENGESAHAN

PEMODELAN TOPIK BERITA DI SITUS WEB UKDW DENGAN
METODE LATENT DIRICHLET ALLOCATION

Oleh: JOSHUA ALBERT WIJAYA / 71170197

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana – Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 20 Juni 2023

Yogyakarta, 20 Juli 2023

Mengesahkan,

Dewan Penguji:

1. Lucia Dwi Krisnawati, Dr. Phil.
2. Aditya Wikan Mahastama, S.Kom., M.Cs.
3. Antonius Rachmat C., S.Kom., M.Cs.
4. R. Gunawan Santosa, Drs. M.Si.

Dekan

Ketua Program Studi


(Restyandito, S.Kom., MSIS., Ph.D.)


(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Dengan segala kerendahan hati, saya, Joshua Albert Wijaya ingin mengucapkan terima kasih yang sebesar-besarnya kepada Tuhan Yang Maha Esa atas karunia-Nya yang tak terhingga. Ia memberikan petunjuk, kekuatan, dan berkat-Nya selama saya menjalani proses penulisan skripsi ini. Kepercayaan diri saya diperkuat oleh kehadiran-Nya dan saya bersyukur atas setiap langkah yang telah diberikan-Nya dalam menyelesaikan skripsi ini.

Pertama-tama, kepada para pembimbing saya, Ibu Dr. Phil. Lucia Dwi Krisnawati, S.S., M.A. dan Bapak Aditya Wikan Mahastama, S.Kom., M.Cs., saya tak dapat memadai mengungkapkan rasa terima kasih saya atas dedikasi, kesabaran, dan pengetahuan yang telah diberikan. Bimbingan yang cermat, arahan yang terstruktur, serta saran-saran yang berharga telah memberikan pijakan yang kokoh dalam perjalanan skripsi ini. Terima kasih atas kesediaan mereka untuk meluangkan waktu, memberikan masukan konstruktif, dan membimbing saya menuju pemahaman yang lebih mendalam dalam bidang ini.

Saya juga ingin menyampaikan rasa terima kasih kepada keluarga, teman-teman, dan rekan-rekan seperjuangan yang memberikan dukungan moral dan semangat sepanjang perjalanan ini. Doa, dorongan positif, dan kehadiran kalian memberikan inspirasi dan motivasi untuk menyelesaikan skripsi saya.



INTISARI

Pemodelan Topik Berita di Situs Web UKDW dengan Metode *Latent Dirichlet Allocation*

Situs *website* universitas menjadi sebuah media di mana civitas akademika mengakses informasi berkaitan dengan instansi tersebut. Universitas Kristen Duta Wacana (UKDW) menghadirkan situs yang beralamat ukdw.ac.id sebagai media informasi bagi khalayak umum untuk mendapatkan informasi tetap maupun berita terkini berkaitan dengan UKDW. Ratusan berita telah terunggah semenjak didirikannya situs ini. Berita-berita tersebut rata-rata memiliki 3 – 8 paragraf. Penelitian ini melakukan analisis pemodelan topik untuk berita-berita di situs web UKDW dari tahun dari 1 Maret 2017 hingga 21 Februari 2023. Sebanyak 812 berita telah dikumpulkan menggunakan langkah *webscraping*. Penelitian ini melakukan analisis pemodelan topik dengan menerapkan metode *Latent Dirichlet Allocation* (LDA). Hasil analisis tersebut kemudian dievaluasi menggunakan *log-likelihood* dan *perplexity*. Pemodelan topik yang telah dilaksanakan menghasilkan nilai evaluasi terbaik dengan nilai sebesar -72,978 dan 19,765 pada jumlah topik sebanyak 4 dengan menggunakan parameter *alpha* 0.1 dan *beta* 0.9.

Kata kunci: situs web UKDW, *Latent Dirichlet Allocation* (LDA), berita

DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS SECARA ONLINE UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA	iv
HALAMAN PERSETUJUAN	v
HALAMAN KELENGKAPAN PENDADARAN	vi
HALAMAN PENGESAHAN	vii
UCAPAN TERIMA KASIH	viii
INTISARI	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Perumusan Masalah	2
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	3
1.5. Langkah-langkah Penelitian	3
1.6. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	6
2.1. Tinjauan Pustaka	6
2.2. Landasan Teori	9
2.2.1. <i>Webscraping</i>	9
2.2.2. <i>Pemodelan Topik</i>	9
2.2.3. <i>Pre-processing</i>	10
2.2.4. <i>Term Frequncy Inverse Document Frequency</i>	11
2.2.5. <i>Latent Dirichlet Allocation</i>	12
2.2.6. <i>Log-likelihood</i>	15
2.2.7. <i>Perplexity</i>	17
	x

BAB III METODOLOGI PENELITIAN	19
3.1. Analisis Kebutuhan Sistem	19
3.1.1. Non Fungsional	19
3.1.2. Fungsional	20
3.2. Perancangan Alur Kerja Sistem	22
3.3. Rancangan Pengujian Sistem	34
BAB IV IMPLEMENTASI DAN HASIL	36
4.1. Persiapan Data	36
4.2. Hasil Eksperimen dan Analisis	50
BAB V KESIMPULAN DAN SARAN	60
5.1. Kesimpulan	60
5.2. Saran	60
DAFTAR PUSTAKA	61
LAMPIRAN	L-1



DAFTAR TABEL

Tabel 2. 1 Tabel Contoh Dokumen TF IDF	11
Tabel 2. 2 Distribusi Probabilitas topik pada suatu dokumen.....	14
Tabel 2. 3 Probabilitas kata pada suatu topik.....	15
Tabel 2. 4 Contoh Hasil Akhir LDA	15
Tabel 2. 5 Tabel Contoh Dokumen <i>Perplexity</i>	17
Tabel 3. 1 Contoh Kombinasi Pengujian	34
Tabel 4. 1 Hasil Kombinasi Parameter.....	50



DAFTAR GAMBAR

Gambar 2. 1 Contoh Hasil <i>Topic Modelling</i>	10
Gambar 3. 1 <i>Use Case Diagram</i> Pemodelan Topik LDA Studi Kasus Berita UKDW	22
Gambar 3. 2 Arsitektur Sistem Pemodelan Topik LDA Studi Kasus Berita UKDW	24
Gambar 3. 3 Sistem Pemodelan Topik LDA Studi Kasus Berita UKDW	26
Gambar 3. 4 <i>Mock-up</i> Halaman Utama	30
Gambar 3. 5 <i>Mock-up</i> Halaman Visualisasi Utama	31
Gambar 3. 6 <i>Mock-up</i> Halaman Evaluasi	32
Gambar 3. 7 <i>Mock-up</i> Halaman Evaluasi 2	33
Gambar 3. 8 <i>Mock-up</i> Halaman Visualisasi Evaluasi	33
Gambar 4. 1 Situs Web Berita UKDW	36
Gambar 4. 2 Berita Pertama dari hasil <i>webscraping</i>	37
Gambar 4. 3 Berita Terakhir dari hasil <i>webscraping</i>	38
Gambar 4. 4 Berita Pertama tanpa tanggal dari hasil <i>webscraping</i>	39
Gambar 4. 5 Berita Terakhir tanpa tanggal dari hasil <i>webscraping</i>	39
Gambar 4. 6 Folder berisi 812 Berita tanpa tanggal	40
Gambar 4. 7 Berita Pertama setelah diproses	41
Gambar 4. 8 Matriks TF-IDF	42
Gambar 4. 9 Hasil LDA <i>gensim</i> dan TF-IDF	43
Gambar 4. 10 Hasil LDA <i>gensim</i> dan TF-IDF dengan parameter lain	44
Gambar 4. 11 Hasil LDA <i>gensim</i> dan <i>bag-of-words</i>	44
Gambar 4. 12 Hasil LDA <i>gensim</i> dan <i>bag-of-words</i> dengan parameter lain	45
Gambar 4. 13 Hasil LDA <i>sklearn</i> dan TF-IDF	45
Gambar 4. 14 Hasil LDA <i>sklearn</i> dan TF-IDF dengan parameter lain	46
Gambar 4. 15 Halaman Utama.....	47
Gambar 4. 16 Halaman Utama dengan isi dokumen	47
Gambar 4. 17 Halaman Visualisasi Topik.....	49
Gambar 4. 18 Halaman Evaluasi.....	50

Gambar 4. 19 Grafik Bandingkan Jumlah Topik 1-30.....	53
Gambar 4. 20 Grafik Bandingkan <i>Alpha</i> 0.1 - 1.....	54
Gambar 4. 21 Grafik Bandingkan <i>Beta</i> 0.1 – 1.....	55
Gambar 4. 22 Hasil Topik <i>Top</i> 1 Evaluasi.....	56
Gambar 4. 23 Hasil Topik <i>Top</i> 2 Evaluasi.....	57
Gambar 4. 24 Hasil Topik <i>Top</i> 3 Evaluasi.....	57
Gambar 4. 25 Hasil Topik yang baik.....	58



BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi informasi dan komunikasi yang cepat membuat banyak kemudahan dalam keberlangsungan hidup manusia. Salah satunya ialah mendapatkan suatu informasi. Dengan mengetahui informasi secara terkini, masyarakat dapat terbantu untuk menentukan sebuah keputusan. Media populer yang digunakan untuk menyebarkan informasi salah satunya adalah portal berita daring. Portal berita daring merupakan salah satu jenis media untuk menyebarkan informasi secara daring dan memiliki isi konten yang sangat beragam seperti olahraga, politik, kesehatan, edukasi. Dengan kemampuan teknologi saat ini banyaknya konten dan waktu untuk mendapatkan berita sangat cepat dalam hitungan menit bahkan detik sudah mampu kita dapatkan. Kemunculan berita di media massa yang sangat agresif menyebabkan masyarakat berpotensi mengalami *information overload* atau ledakan informasi. Ledakan informasi mengakibatkan para pencari informasi mengalami kesulitan untuk menaksir kredibilitas informasi (Prasteyawan, 2018).

Salah satu penyedia berita daring adalah Universitas Kristen Duta Wacana (UKDW). UKDW menyebarkan berita melalui situs web dan dapat diakses secara daring oleh seluruh civitas akademika. Saat ini berita yang ditampilkan masih secara acak dan belum ada pengkategorian yang memilah isi kontennya. Sehingga dibutuhkan waktu yang lama untuk membaca berita dengan isi konten yang diinginkan. Proses ini tentunya sangat tidak efisien, sehingga manusia membutuhkan pemanfaatan komputer untuk memecahkan masalah ini, yaitu memunculkan topik dari setiap dokumen. *Topic Modelling* atau pemodelan topik dapat menjadi solusi dari permasalahan ini. Pemodelan topik adalah salah satu metode statistik yang berdasarkan kumpulan dokumen yang digunakan dalam tugas abstraksi otomatis, ekstraksi informasi dan klasifikasi (Kirill dkk, 2020). Dengan penerapan pemodelan topik, kumpulan dokumen berita dapat diklasifikasikan

melalui topik-topiknya. Terdapat dua algoritma yang cukup populer pada pemodelan topik yaitu *Latent Dirichlet Allocation (LDA)* dan *Latent Semantic Analysis (LSA)*. LSA berfokus mempelajari topik laten dengan melakukan *Matrix Decomposition* sedangkan LDA adalah model yang didasari oleh probabilitas. Mulya (2020) menyatakan LDA terbukti bekerja lebih baik dibanding LSA.

Berdasarkan latar belakang, maka penelitian ini berfokus pada penerapan pemodelan topik pada berita di situs web UKDW. Penerapan pemodelan topik dengan menggunakan algoritma *Latent Dirichlet Allocation*. Proses yang dilakukan meliputi pengambilan data berita dari situs web UKDW, kemudian dilakukan normalisasi agar data dapat digunakan dan diproses dengan baik dalam proses berikutnya. Proses selanjutnya yaitu pembentukan transformasi data teks menjadi sekumpulan angka atau biasa disebut dengan *Document Term Matrix*. Setelah itu pemodelan topik dengan metode LDA dapat dilakukan. Karena pemodelan topik merupakan salah satu teknik dari *Unsupervised Learning* maka evaluasi dilakukan secara subjektif dengan melakukan beberapa pengulangan untuk menentukan berapa jumlah topik yang tepat untuk digunakan. Diharapkan dari pembangunan sistem ini dapat membantu menyelesaikan masalah terhadap banyaknya berita di situs web UKDW yang masih belum diklasifikasikan.

1.2. Perumusan Masalah

Berdasarkan latar belakang yang ada, penulis merumuskan masalah sebagai berikut :

1. Bagaimana hasil topik yang muncul setelah pengaplikasian pemodelan topik pada berita UKDW?
2. Bagaimana hasil evaluasi topik yang dihasilkan menggunakan metode *Latent Dirichlet Allocation* dengan *log-likelihood* dan *perplexity*?

1.3. Batasan Masalah

Pada penelitian ini terdapat kemungkinan untuk melebar dari tujuan awal penelitian. Oleh karena itu penulis mempunyai beberapa batasan, antara lain :

1. Dokumen yang digunakan adalah berita UKDW yang diambil dari situs web UKDW berjumlah 812 berita dengan rentang waktu dari 1 Maret 2017 hingga 21 Februari 2023.
2. Sistem yang dibangun menggunakan bahasa pemrograman python versi 3.8.3.
3. Proses Webscraping dan normalisasi berita UKDW dilakukan di luar aplikasi.
4. Sistem berbasis aplikasi desktop.
5. Aplikasi hanya dapat memproses dokumen dengan ekstensi file .txt.
6. Hasil output sistem adalah kumpulan kata yang mewakili dari setiap topik.

1.4. Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengetahui seberapa baik kinerja Latent Dirichlet Allocation untuk melakukan pemodelan topik dengan evaluasi *log-likelihood* dan *perplexity* serta mengetahui hasil pengelompokan topik dalam studi kasus berita di situs web UKDW. Selain itu, sebagai syarat menyelesaikan program pendidikan sarjana (S-1).

1.5. Langkah-langkah Penelitian

1. Diskusi

Pada tahap awal, penulis berkomunikasi dengan Ibu Dr. Phil. Lucia Dwi Krisnawati, S.S., M.A sebagai seorang dosen yang mengampu mata kuliah seputar topik terkait, untuk membahas mengenai sistem yang akan dikembangkan, termasuk studi kasus, metode yang akan digunakan, dan aspek-aspek lain yang terkait.

2. Studi Pustaka

Dalam studi pustaka ini, peneliti melakukan pencarian literatur yang relevan untuk mendapatkan pemahaman yang lebih baik tentang topik penelitian yang akan dilakukan. Peneliti mengakses sumber-sumber seperti jurnal ilmiah, buku, artikel, dan laporan terkait topik yang akan diteliti.

3. Pembuatan Sistem

Pembuatan sistem melibatkan tahapan analisis kebutuhan untuk mengidentifikasi dan memahami kebutuhan sistem. Selanjutnya, perancangan sistem dilakukan dengan merancang arsitektur sistem, mengidentifikasi komponen utama, dan mendetailkan desain antarmuka pengguna, desain basis data, serta alur pemodelan topik LDA. Setelah itu dilakukan persiapan data dengan mengambil data dari situs web berita UKDW yang akan digunakan dalam sistem. Kemudian pada tahap akhir juga dilakukan implementasi, pengujian, dan analisis untuk mengetahui hasil yang diperoleh.

4. Konsultasi

Dalam tahap konsultasi, peneliti berinteraksi dengan dosen pembimbing, atau sesama peneliti untuk mendiskusikan masalah-masalah yang ditemukan selama penelitian. Tujuan konsultasi ini adalah untuk mendapatkan pandangan dan masukan yang berharga dari orang-orang yang memiliki pengetahuan dan pengalaman yang relevan.

5. Pengujian

Pada tahap pengujian, dilakukan serangkaian percobaan pemodelan topik dengan variasi parameter, seperti jumlah topik, *alpha*, dan *beta*. Melalui pengujian ini, tujuannya adalah untuk memahami bagaimana perubahan parameter dapat memengaruhi hasil pemodelan topik, seperti kualitas topik yang dihasilkan dan distribusi kata-kata dalam setiap topik.

6. Analisis Hasil Akhir

Melalui analisis hasil akhir, penulis dapat memaparkan hasil data yang sudah diperoleh dari pengujian. Sehingga dapat menarik kesimpulan dan memberikan saran kepada pembaca untuk melakukan penelitian ke depan.

1.6. Sistematika Penulisan

Penelitian ini terdiri dari beberapa bagian yang disusun dengan tujuan dan maksud tertentu, yaitu:

Bab I Pendahuluan dalam penelitian ini mencakup beberapa aspek penting. Pertama, latar belakang masalah akan menjelaskan konteks dan urgensi

permasalahan yang akan diteliti. Kedua, perumusan masalah akan mengidentifikasi permasalahan utama yang akan diselesaikan dalam penelitian. Terakhir, bab ini juga akan memuat tujuan penelitian, langkah-langkah penelitian, dan sistematika penulisan yang akan diikuti.

Bab II Tinjauan Pustaka dan Dasar Teori akan membahas literatur yang relevan dengan penelitian ini. Tinjauan pustaka akan mengumpulkan penelitian terdahulu yang telah dilakukan dalam bidang yang sama. Sementara itu, dasar teori akan memberikan landasan konseptual yang menjadi dasar analisis dalam penelitian ini.

Bab III Metodologi Penelitian akan menjelaskan alur persiapan data yang berupa dokumen berita dari situs UKDW dan pembuatan aplikasi desktop untuk pemodelan topik dengan metode LDA serta metode evaluasinya. Penjelasan akan mencakup langkah-langkah pengambilan berita, proses pembuatan aplikasi, dan penerapan metode LDA pada data berita yang telah diambil.

Bab IV Implementasi dan Hasil menyajikan hasil riset/implementasi, analisis terpadu dari hasil tersebut, dan pengujian sistem untuk membuktikan kebenaran penelitian. Di sini, akan dijelaskan temuan-temuan penting yang ditemukan dan hasil tersebut akan dianalisis secara menyeluruh dengan pendekatan yang relevan. Pembahasan akan menyoroti implikasi hasil riset terhadap bidang studi yang relevan dan kesesuaian dengan tujuan penelitian yang telah ditetapkan.

Bab V Kesimpulan dan Saran memberikan ringkasan hasil analisis riset/implementasi dalam penyusunan skripsi ini. Kesimpulan akan menyampaikan temuan utama dan jawaban terhadap permasalahan penelitian dengan jelas. Selanjutnya, saran-saran untuk penelitian mendatang akan disampaikan, mencakup langkah-langkah atau metode pengembangan yang dapat meningkatkan kinerja sistem yang diteliti pada penelitian berikutnya.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dari semua penelitian yang telah dilakukan oleh penulis, dapat ditarik kesimpulan berikut ini:

1. Pengaplikasian pemodelan topik dengan metode *Latent Dirichlet Allocation* pada berita UKDW menghasilkan banyak topik tentang kegiatan UKDW.
2. Hasil evaluasi topik yang dihasilkan menggunakan metode *Latent Dirichlet Allocation* dengan *log-likelihood* dan *perplexity* bergantung dengan parameter yang digunakan seperti :
 - Semakin banyak jumlah topik maka nilai *log-likelihood* semakin rendah dan nilai *perplexity* semakin tinggi.
 - Semakin tinggi nilai *alpha* maka nilai *log-likelihood* semakin rendah dan nilai *perplexity* semakin tinggi.
 - Semakin tinggi nilai *beta* maka nilai *log-likelihood* semakin tinggi dan nilai *perplexity* semakin rendah.
3. Hasil pengujian menunjukkan bahwa nilai *log-likelihood* dan *perplexity* yang baik tidak selalu menghasilkan topik yang koheren.

5.2. Saran

Dari semua penelitian yang telah dilakukan oleh penulis, ada beberapa saran yang dapat diambil:

1. Penelitian ini dapat dikembangkan kembali dengan menambah pengujian parameter, seperti jumlah iterasi dalam pemodelan topik, jumlah dokumen yang diproses setiap iterasi. Dengan pengujian parameter yang lebih kompleks memungkinkan untuk mendapat model yang lebih baik.
2. Penelitian ini dapat dikembangkan kembali dengan menambah atau mengganti dengan model lain, untuk melihat apakah korpus berita UKDW lebih cocok jika diterapkan menggunakan model lain.

DAFTAR PUSTAKA

- Ahmad, F. Z., Satria Arifandy, M. F., Caesarardhi, M. R., & Rakhmawati, N. A. (2021). Bagaimana Masyarakat Menyikapi Pembelajaran Tatap Muka: Analisis Komentar Masyarakat pada Media Sosial Youtube Menggunakan Algoritma Deep Learning Sekuensial dan LDA. *Jurnal Linguistik Komputasional*, 40-46.
- Blei, D. M. (2012). *Introduction to Probabilistic Topic Models*. Princeton: Princeton University.
- Chilmi, M. L. (2021). *Latent dirichlet allocation (lda) untuk mengetahui topik pembicaraan warganet twitter*. Jakarta: Universitas Islam Negeri Syarif Hidayatullah.
- Jacobi, C., van Atteveldt, W., & Welbers, K. (2015). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Rethinking Research Methods in an Age of Digital Journalism*, 89-106.
- Josi, A., Abdillah, L. A., & Suryayusra. (2014). Penerapan teknik web scraping pada mesin pencari artikel ilmiah.
- Kirill, Y., Mihail, I. G., Sanzhar, M., Rustam, M., Olga, F., & Ravil, M. (2020). Propaganda Identification Using Topic Modelling. *Procedia Computer Science*, 205-212.
- Mulya, U. (2020). *Analisis Komparasi Pemodelan Topik Latent Dirichlet Allocation Dan Latent Semantic Analysis Pada Ulasan Restoran Di Yogyakarta*. Yogyakarta: Universitas Islam Negeri Sunan Kalijaga.
- Najjichah, H., Syukur, A., & Subagyo, H. (2019). Pengaruh Text Preprocessing dan kombinasinya pada peringkasan dokumen otomatis teks berbahasa indonesia. *Jurnal Teknologi Informasi*, 1-11.
- Patmawati, & Yusuf, M. (2021). Analisis Topik Modelling Terhadap Penggunaan Sosial Media Twitter. *Building of Informatics, Technology and Science (BITS)*, 122–129.
- Prasetyawan, Y. Y. (2018). Menyoal Pentingnya Layanan Informasi dan Referensi di Perguruan Tinggi Indonesia. *Anuva: Jurnal Kajian Budaya*, 393-398.
- Robertson, S. (2004). Understanding Inverse Document Frequency. *Journal of Documentation*, 503–520.

Santika, T. (2019). *evaluasi perplexity untuk permodelan topik diskusi agama islam di media sosial twitter indonesia tahun 2006 - 2018 menggunakan latent dirichlet allocation*. Jakarta: UIN Syarif Hidayatullah.

Suparyati, Utami, E., & Fathurahman, A. (2022). Pengamatan Tren Ulasan Hotel Menggunakan Pemodelan Topik. *Journal of Applied Informatics and Computing (JAIC)*, 71~77.

