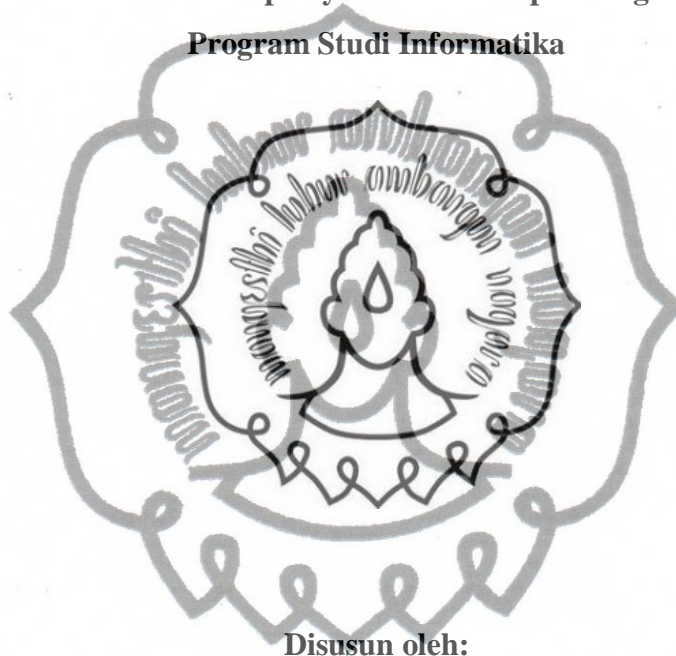


**ANALISIS PEMODELAN TOPIK PADA ARTIKEL BERITA  
MENGUNAKAN METODE LATENT DIRICHLET ALLOCATION**

**SKRIPSI**

**Diajukan untuk memenuhi persyaratan mendapatkan gelar Strata Satu  
Program Studi Informatika**



Disusun oleh:

**HANIF SULTHAN RIZQULLAH**

**M0514019**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS SEBELAS MARET**

**SURAKARTA**

**2019**

**SKRIPSI**  
**ANALISIS PEMODELAN TOPIK PADA ARTIKEL BERITA**  
**MENGGUNAKAN METODE LATENT DIRICHLET ALLOCATION**

**Disusun oleh:**

**HANIF SULTHAN RIZQULLAH**

**NIM. M0514019**

Skripsi ini telah disetujui untuk dipertahankan di hadapan dewan penguji  
pada tanggal 24 Januari 2019

Pembimbing 1,

Pembimbing 2,



**Afrizal Doewes, S.Kom., M.Sc.**

**NIP. 198508312012121004**





**Haryono Setiadi, ST., M.Eng**

**NIP. 198003272005011002**

**SKRIPSI****ANALISIS PEMODELAN TOPIK PADA ARTIKEL BERITA  
MENGUNAKAN METODE LATENT DIRICHLET ALLOCATION****Disusun oleh:****HANIF SULTHAN RIZQULLAH****NIM. M0514019**

telah dipertahankan di hadapan dewan penguji pada tanggal,

29 Januari 2019**Susunan Dewan Penguji**

- |  |              |   |
|--|--------------|---|
| 1. <b><u>Afrizal Doewes, S.Kom., M.Sc.</u></b><br>NIP. 198508312012121004              | (Ketua)      | (  ) |
| 2. <b><u>Haryono Setiadi, ST., M.Eng</u></b><br>NIP. 198003272005011002                | (Sekretaris) | (  ) |
| 3. <b><u>Dr. Techn. Dewi Wisnu Wardani, S.Kom, M.S.</u></b><br>NIP. 197810262005012002 | (Anggota)    | (  ) |
| 4. <b><u>Ristu Saptono, S.Si., M.T.</u></b><br>NIP. 197902102002121001                 | (Anggota)    | (  ) |

**Disahkan oleh****Kepala Program Studi Informatika****Drs. Bambang Harjito, M.App.Sc., Ph.D.****NIP. 196211301991031002**

## MOTTO



## PERSEMBAHAN



*Skripsi ini penulis persembahkan kepada:*

**Ayah, Ibu, adik, dan Keluarga Besar,  
serta teman-teman S1 Informatika UNS khususnya “iFourteen”.**

## KATA PENGANTAR

Segala puji dan syukur Penulis panjatkan kepada Allah *Subhaanahu wa Ta'ala* atas berkat dan rahmat-Nya, Penulis dapat menyelesaikan Tugas Akhir dengan judul “Analisis Pemodelan Topik pada Artikel Berita Menggunakan Metode Latent Dirichlet Allocation” sebagai syarat untuk memperoleh gelar Sarjana Informatika di Universitas Sebelas Maret Surakarta.

Dalam penyusunan Tugas Akhir ini banyak pihak yang telah memberi bantuan, oleh karena itu tidak lupa penulis mengucapkan terima kasih kepada:

1. Bapak Drs. Bambang Harjito, M.App.Sc., Ph.D selaku Kepala Program Studi Informatika Universitas Sebelas Maret Surakarta dan Pembimbing Akademis.
2. Bapak Afrizal Doewes, S.Kom., M.Sc. selaku Dosen Pembimbing I yang telah memberikan bimbingan serta masukan dalam penyusunan Tugas Akhir.
3. Bapak Haryono Setiadi, ST., M.Eng selaku Dosen Pembimbing II yang telah memberikan bimbingan serta masukan dalam penyusunan Tugas Akhir.
4. Kedua Orang tua serta keluarga yang selalu memberikan motivasi dan do'a.
5. Teman-teman khususnya Informatika angkatan 2014 “*iFourteen*” dan KKN Batu UNS 2018 yang telah memberikan bantuan dan motivasi dalam menyelesaikan laporan Tugas Akhir ini.
6. Semua pihak yang tidak dapat Penulis sebutkan satu per satu.

Penulis menyadari masih banyak kekurangan dalam penyusunan laporan Tugas Akhir ini baik dari segi penulisan maupun materi yang dijelaskan. Dengan adanya laporan Tugas Akhir ini, diharapkan mampu memberikan manfaat dan menambah wawasan bagi pembaca maupun Penulis sendiri.

Surakarta, Januari 2019

Penulis



## ANALISIS PEMODELAN TOPIK PADA ARTIKEL BERITA MENGUNAKAN METODE LATENT DIRICHLET ALLOCATION

HANIF SULTHAN RIZQULLAH

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Sebelas Maret

### ABSTRAK

Berita merupakan sebuah informasi berupa fakta atau pendapat seseorang yang berasal dari sebuah kejadian/peristiwa yang sifatnya menarik untuk diketahui dan dimuat melalui media massa, termasuk portal online berita. Dalam menemukan topik tersembunyi yang terdapat dalam kumpulan artikel berita salah satu metode yang cepat dan efisien adalah pemodelan topik, dan *Latent Dirichlet Allocation* (LDA) merupakan salah satunya. Tahap-tahap dalam penelitian ini antara lain pengumpulan data, *text preprocessing*, pemodelan topik, dan analisis hasil. Pemodelan topik dengan metode LDA menghasilkan 8 topik sebagai jumlah topik terbaik berdasarkan nilai *perplexity* terendah dengan nilai *perplexity* 3087,927 dan berdasarkan hasil analisa model dari dua eksperimen yang dilakukan yaitu eksperimen terkait dengan jumlah *passes* dan eksperimen terkait dengan jumlah topik.

**Kata kunci:** *pemodelan topik, Latent Dirichlet Allocation*

# TOPIC MODELING ANALYSIS ON NEWS ARTICLE BY USING LATENT DIRICHLET ALLOCATION

**HANIF SULTHAN RIZQULLAH**

Department of Informatics, Faculty of Mathematics and Natural Sciences,  
Universitas Sebelas Maret

## ABSTRACT

News is an information comes from an event based on fact or someone opinion that is interesting to know and being published on mass media including online news portals. Topic modelling is one of the quickest and efficient methods to discover hidden topics in a collection of news article, and Latent Dirichlet Allocation (LDA) is one of them. Methodology in this research consists of data collection, text preprocessing, topic modelling, and result analysis. Topic modelling using LDA yields 8 topics as the best number of topics based on the lowest perplexity with value 3087.927 and based on model analysis from two experiments, experiments related to the number of passes and experiments related to the number of topics.

**Keywords:** *topic modeling, Latent Dirichlet Allocation*



## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
MOTTO.....	iv
PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
ABSTRAK .....	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN .....	1
1.1    Latar Belakang.....	1
1.2    Rumusan Masalah .....	3
1.3    Batasan Masalah.....	3
1.4    Tujuan Penelitian.....	3
1.5    Manfaat Penelitian.....	3
1.6    Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1    Dasar Teori .....	5
2.1.1    Text Mining .....	5
2.1.2    Text Preprocessing.....	5

2.1.3	Pemodelan Topik .....	8
2.1.4	Latent Dirichlet Allocation .....	8
2.1.5	Perplexity .....	11
2.2	Penelitian Terkait.....	12
	Tabel 2.1 Perbandingan Hasil Penelitian Terkait.....	17
<b>BAB III METODOLOGI PENELITIAN.....</b>		<b>20</b>
3.1	Pengumpulan Data.....	20
3.2	Text Preprocessing .....	21
3.2.1	Unitize and Tokenization .....	22
3.2.2	Standardization and Cleaning .....	22
3.2.3	Stop Word Removal.....	22
3.2.4	Lemmatization.....	22
3.2.5	Bag-of-Words.....	22
3.3	Pemodelan Topik.....	23
3.3.1	Inisialisasi.....	23
3.3.2	E-Step.....	24
3.3.3	M-Step.....	25
3.4	Analisis Hasil.....	25
3.4.1	Analisis berdasarkan jumlah <i>passes</i> .....	25
3.4.2	Analisis berdasarkan jumlah topik .....	26
<b>BAB IV PEMBAHASAN.....</b>		<b>28</b>
4.1	Hasil Pengumpulan Data .....	28
4.2	Hasil Text Preprocessing .....	28
4.2.1	Unitize and Tokenization .....	29
4.2.2	Standardization and Cleaning .....	29

4.2.3	Stop Word Removal.....	29
4.2.4	Lemmatization.....	30
4.2.5	Bag-of-Words.....	30
4.3	Pemodelan Topik.....	32
4.3.1	Inisialisasi.....	32
4.3.2	E-Step.....	33
4.3.3	M-Step.....	36
4.4	Analisis Pemodelan Topik.....	37
4.4.1	Analisis pemodelan topik berdasarkan jumlah <i>passes</i> .....	38
4.4.2	Analisis pemodelan topik berdasarkan jumlah topik .....	38
BAB V PENUTUP.....		42
5.1	Kesimpulan.....	42
5.2	Saran.....	42
DAFTAR PUSTAKA .....		43
LAMPIRAN.....		46

**DAFTAR TABEL**

Tabel 2.1 Perbandingan Hasil Penelitian Terkait.....	17
Tabel 4.1 Jumlah data yang didapatkan .....	28
Tabel 4.2 Representasi id token .....	31
Tabel 4.3 Rincian Inisialisasi Parameter.....	32
Tabel 4.4 Pemodelan jumlah topik 10.....	40
Tabel 4.5 Pemodelan jumlah topik 14.....	41



## DAFTAR GAMBAR

Gambar 2.1 Data teks asli .....	6
Gambar 2.2 Hasil unitize dan tokenize .....	6
Gambar 2.3 Hasil standardization dan cleaning .....	6
Gambar 2.4 Hasil stop word removal .....	7
Gambar 2.5 Hasil stemming .....	7
Gambar 2.6 Hasil lemmatization .....	8
Gambar 2.7 Representasi Model LDA (Blei, et al., 2003) .....	9
Gambar 3.1 Alur Metodologi Penelitian .....	20
Gambar 3.2 Tahap Text Preporcessing .....	21
Gambar 3.3 Pola Regex untuk Tokenization .....	22
Gambar 3.4 Tahap pemodelan topik LDA oleh gensim .....	23
Gambar 3.5 Rancangan visualisasi grafik eksperimen berdasarkan jumlah passes .....	26
Gambar 3.6 Rancangan visualisasi grafik eksperimen berdasarkan jumlah topik .....	27
Gambar 4.1 Data asli .....	29
Gambar 4.2 Tahap Unitize and Tokenization .....	29
Gambar 4.3 Tahap Standardtization and Cleaning .....	29
Gambar 4.4 Tahap Stop Word Removal .....	30
Gambar 4.5 Tahap Lemmatization .....	30
Gambar 4.6 Bag-of-words .....	30
Gambar 4.7 Nilai perplexity berdasarkan jumlah passes .....	38
Gambar 4.8 Nilai perplexity berdasarkan jumlah topik .....	39

**DAFTAR LAMPIRAN**

Lampiran 1 Dataset penelitian .....	46
Lampiran 2 Hasil <i>Text Preprocessing</i> .....	46
Lampiran 3 Hasil Inisialisasi, E-step, dan M-step .....	47
Lampiran 4 Hasil Pemodelan Topik LDA .....	60

