

**IDENTIFIKASI TOPIK TWEET TOKOH PUBLIK POLITIK
MENGUNAKAN TOPIC MODELING DENGAN METODE
LATENT DIRICHLET ALLOCATION**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

Aditya Dwi Rangga Pribadi

18.11.2395

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

**IDENTIFIKASI TOPIK TWEET TOKOH PUBLIK POLITIK
MENGUNAKAN TOPIC MODELING DENGAN METODE
LATENT DIRICHLET ALLOCATION**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

Aditya Dwi Rangga Pribadi

18.11.2395

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**IDENTIFIKASI TOPIK TWEET TOKOH PUBLIK POLITIK
MENGUNAKAN TOPIC MODELING DENGAN METODE LATENT
DIRICHLET ALLOCATION**

yang disusun dan diajukan oleh

Aditya Dwi Rangga Pribadi

18.11.2395

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 10 Mei 2023

Dosen Pembimbing,



Aini Yanti, M. Kom.

NIK. 190302255

HALAMAN PENGESAHAN

SKRIPSI

**IDENTIFIKASI TOPIK TWEET TOKOH PUBLIK POLITIK
MENGUNAKAN TOPIC MODELING DENGAN METODE LATENT
DIRICHLET ALLOCATION**

yang disusun dan diajukan oleh

Aditya Dwi Rangga Pribadi

18.11.2395

Telah dipertabankan di depan Dewan Penguji
pada tanggal 19 Mei 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Subkutiningsih, M.Kom
NIK. 190302413

Anggit Ferdita Nugraha, S.T., M.Eng
NIK. 190302480

Ainul Yaqin, M. Kom.
NIK. 190302255



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Mei 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Aditya Dwi Rangga pribadi
NIM : 18.11.2395

Menyatakan bahwa Skripsi dengan judul berikut:

IDENTIFIKASI TOPIK TWEET TOKOH PUBLIK POLITIK MENGGUNAKAN TOPIC MODELING DENGAN METODE LATENT DIRICHLET ALLOCATION

Dosen Pembimbing : Ainul Yaqin, M. Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Mei 2023

Yang Menyatakan,

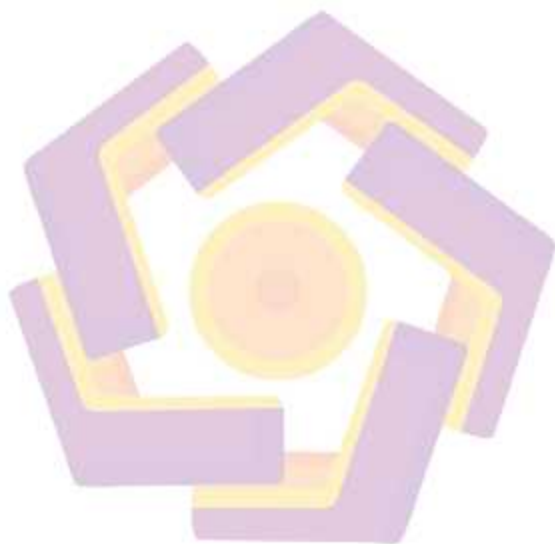


Aditya Dwi Rangga Pribadi

MOTTO

“Sesungguhnya sesudah kesulitan itu ada Kemudahan.”

(Q.S Al-Insyirah: 6)



PERSEMBAHAN

Dengan rasa syukur yang mendalam, dengan telah diselesaikannya skripsi ini Penulis mempersembahkannya kepada :

1. Orang tua saya yang selalu memberikan motivasi dan doa tanpa henti.
2. Bapak Ainul Yaqin, M.Kom. yang telah membimbing saya dalam mengerjakan skripsi.
3. Teman-teman ABB yang selalu menemani saya dalam suka maupun duka, di saat susah maupun bahagia.
4. Teman-teman dari kelas 18-IF09 yang menemani saya dari awal sampai akhir perkuliahan.



KATA PENGANTAR

Puji syukur kehadiran Allah SWT. Yang telah memberikan rahmat dan hidayah-Nya sehingga saya dapat menyelesaikan skripsi ini. Skripsi yang berjudul **“Identifikasi Topik Tweet Tokoh Publik Politik Menggunakan Topik Modeling Dengan Metode Latent Dirichlet Allocation”** ini menjadi salah satu syarat untuk menyelesaikan masa studi program sarjana di Universitas Amikom Yogyakarta.

Dalam penelitian ini, penulis mengucapkan terima kasih atas motivasi, bimbingan saran dan masukan kepada semua pihak baik secara langsung maupun tidak langsung. Pada kesempatan ini penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradnya D., M.Kom. selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
4. Bapak Ainul Yaqin, M.Kom. selaku Dosen Pembimbing yang selalu memberikan bimbingan, saran dan masukan dalam penulisan skripsi.

Penulis menyadari bahwa skripsi ini masih memiliki banyak kekurangan, maka dari itu kritik dan saran membangun sangat dibutuhkan penulis untuk perbaikan karya selanjutnya. Semoga kita semua selalu dalam lindungan Allah SWT. Akhir kata semoga penelitian ini dapat bermanfaat bagi pembaca dan menambah wawasan pengetahuan khususnya dalam bidang informatika.

Yogyakarta, 7 Mei 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
MOTTO	v
PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xii
INTISARI.....	xiii
ABSTRACT	xiv
BAB I.....	1
1.1 Latar Belakang	1
1.2 Rumusan masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	2
1.5 Metode Penelitian.....	3
1.5.1 Metode Pengumpulan Data.....	3
1.6 Sistematika Penulisan.....	3
BAB II.....	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori	8

2.2.1	Text Mining.....	8
2.2.2	<i>Pre-processing</i>	8
2.2.3	Topic Modeling.....	10
2.2.4	Latent Dirichlet Allocation	11
2.2.5	Topic coherence	13
BAB III	15
3.1	Tahapan Penelelitian	15
3.2	Data Penelitian	16
3.3	Langkah-Langkah Penelitian.....	16
3.3.1	Pengambilan Data	16
3.3.2	Pre-processing Data	16
3.3.3	Latent Dirichlet Allocation	21
3.3.4	Analisis Topik	23
3.4	Analisis Kebutuhan	23
3.4.1	Kebutuhan Perangkat Keras.....	24
3.4.2	Kebutuhan Perangkat Lunak	24
BAB IV	25
4.1	Pengumpulan Data	25
4.2	Pre-processing	25
4.2.1	Tokenization.....	25
4.2.2	Stopword Removal.....	26
4.2.3	Stemming	28
4.3	Pemodelan Topik Menggunakan LDA.....	29
4.4	Analisis Topik	67
4.5	Evaluasi LDA.....	70
BAB V	73
5.1	Kesimpulan.....	73
5.2	Saran.....	73
DAFTAR PUSTAKA	74

DAFTAR TABEL

Tabel 2. 1 Perbandingan Penelitian.....	6
Tabel 2. 2 Contoh Tokenization.....	8
Tabel 2. 3 Contoh Stopword Removal.....	9
Tabel 3. 1 Contoh Tokenization.....	17
Tabel 3. 2 Contoh Stopword Removal.....	18
Tabel 3. 3 Contoh Stemming	20
Tabel 3. 4 Hasil 2 Topik 100 Iterasi.....	21
Tabel 3. 5 Hasil Pemetaan Bobot 2 Topik 100 Iterasi	22
Tabel 3. 6 Kebutuhan Perangkat Keras	24
Tabel 3. 7 Kebutuhan Perangkat Lunak	24
Tabel 4. 1 Perbedaan Sebelum Dan Setelah Tokenization	26
Tabel 4. 2 Perbedaan Sebelum Dan Setelah Stopword Removal.....	27
Tabel 4. 3 Perbedaan Sebelum Dan Setelah Stemming	29
Tabel 4. 4 Hasil 2 Topik 100 Iterasi.....	30
Tabel 4. 5 Hasil Pemetaan Bobot 2 Topik 100 Iterasi	30
Tabel 4. 6 Hasil 2 Topik 200 Iterasi.....	32
Tabel 4. 7 Hasil Pemetaan Bobot 2 Topik 200 Iterasi	32
Tabel 4. 8 Hasil 2 Topik 500 Iterasi.....	34
Tabel 4. 9 Hasil Pemetaan Bobot 2 Topik 500 Iterasi	34
Tabel 4. 10 Hasil 2 Topik 1000 Iterasi.....	36
Tabel 4. 11 Hasil Pemetaan Bobot 2 Topik 1000 Iterasi	36
Tabel 4. 12 Hasil 3 Topik 100 Iterasi.....	38
Tabel 4. 13 Hasil Pemetaan Bobot 3 Topik 100 Iterasi	38
Tabel 4. 14 Hasil 3 Topik 200 Iterasi.....	40
Tabel 4. 15 Hasil Pemetaan Bobot 3 Topik 200 Iterasi	40
Tabel 4. 16 Hasil 3 Topik 500 Iterasi.....	42
Tabel 4. 17 Hasil Pemetaan Bobot 3 Topik 500 Iterasi	42
Tabel 4. 18 Hasil 3 Topik 1000 Iterasi.....	44

Tabel 4. 19 Hasil Pemetaan Bobot 3 Topik 1000 Iterasi	44
Tabel 4. 20 Hasil 5 Topik 100 Iterasi.....	46
Tabel 4. 21 Hasil Pemetaan Bobot 5 Topik 100 Iterasi	46
Tabel 4. 22 Hasil 5 Topik 200 Iterasi.....	48
Tabel 4. 23 Hasil Pemetaan Bobot 5 Topik 200 Iterasi	49
Tabel 4. 24 Hasil 5 Topik 500 Iterasi.....	51
Tabel 4. 25 Hasil Pemetaan Bobot 5 Topik 500 Iterasi	51
Tabel 4. 26 Hasil 5 Topik 1000 Iterasi.....	53
Tabel 4. 27 Hasil Pemetaan Bobot 5 Topik 1000 Iterasi	53
Tabel 4. 28 Hasil 7 Topik 100 Iterasi.....	55
Tabel 4. 29 Hasil Pemetaan Bobot 7 Topik 100 Iterasi	56
Tabel 4. 30 Hasil 7 Topik 200 Iterasi.....	58
Tabel 4. 31 Hasil Pemetaan Bobot 7 Topik 200 Iterasi	59
Tabel 4. 32 Hasil 7 Topik 500 Iterasi.....	61
Tabel 4. 33 Hasil Pemetaan Bobot 7 Topik 500 Iterasi	62
Tabel 4. 34 Hasil 7 Topik 1000 Iterasi.....	64
Tabel 4. 35 Hasil Pemetaan Bobot 7 Topik 1000 Iterasi	64
Tabel 4. 36 Rangkuman Visualisasi Pemodelan LDA.....	67
Tabel 4. 37 Coherence Score	71
Tabel 4. 38 Hasil Pemetaan Bobot 3 Topik	71

DAFTAR GAMBAR

Gambar 2. 1 Konsep Topik Modelling Blei [12]	11
Gambar 2. 2 Topik Modelling dengan metode LDA[3]	12
Gambar 3. 1 Tahapan Penelitian	15
Gambar 3. 2 Visualisasi LDA 2 Topik 100 Iterasi.....	23
Gambar 4. 1 Visualisasi LDA 2 Topik 100 Iterasi.....	31
Gambar 4. 2 Visualisasi LDA 2 Topik 200 Iterasi.....	33
Gambar 4. 3 Visualisasi LDA 2 Topik 500 Iterasi.....	35
Gambar 4. 4 Visualisasi LDA 2 Topik 1000 Iterasi	37
Gambar 4. 5 Visualisasi LDA 3 Topik 100 Iterasi.....	39
Gambar 4. 6 Visualisasi LDA 3 Topik 200 Iterasi.....	41
Gambar 4. 7 Visualisasi LDA 3 Topik 500 Iterasi.....	43
Gambar 4. 8 Visualisasi LDA 3 Topik 1000 Iterasi	45
Gambar 4. 9 Visualisasi LDA 5 Topik 100 Iterasi.....	47
Gambar 4. 10 Visualisasi LDA 5 Topik 200 Iterasi.....	50
Gambar 4. 11 Visualisasi LDA 5 Topik 500 Iterasi.....	52
Gambar 4. 12 Visualisasi LDA 5 Topik 1000 Iterasi	54
Gambar 4. 13 Visualisasi LDA 7 Topik 100 Iterasi.....	57
Gambar 4. 14 Visualisasi LDA 7 Topik 200 Iterasi.....	60
Gambar 4. 15 Visualisasi LDA 7 Topik 500 Iterasi.....	63
Gambar 4. 16 Visualisasi LDA 7 Topik 1000 Iterasi	66
Gambar 4. 17 Visualisasi Grafik Coherence Score.....	70
Gambar 4. 18 Visualisasi LDA 3 Topik.....	72

INTISARI

Topic Modeling merupakan teknik yang dapat digunakan untuk menganalisis dokumen yang berukuran besar maupun kecil. *Topic Modeling* dapat memahami topik laten dalam dokumen menggunakan variable yang acak, dan menemukan struktur dokumen tersebut. *Topic Modeling* memiliki berbagai aplikasi seperti mengekstraksi kata kunci, mengkategorikan dokumen, rekomendasi tag dan lainnya. Dalam penelitian ini, dataset yang digunakan berasal dari twitter yang dikumpulkan dari beberapa akun milik tokoh publik politik Indonesia. Tujuan dari penelitian ini adalah untuk mengetahui banyak klaster yang optimal dalam pemodelan topik dari *tweet* tokoh publik politik menggunakan algoritma *Latent Dirichlet Allocation*. Tahapan penelitian ini dimulai dari pengumpulan data menggunakan scraping model *api twitter*, kemudian tahap *preprocessing* seperti menghapus *stopword*, menghapus emoji, *tokenization*, dan *stemming*, lalu tahap berikutnya proses *topic modeling* menggunakan metode *Latent Dirichlet Allocation* dan terakhir hasil penelitian berupa grafik *topic modeling*. Dari percobaan yang dilakukan dengan topik sejumlah 2, 3, 5, dan 7 serta iterasi sejumlah 100, 200, 500, dan 1000. Didapatkan hasil klaster terbaik dengan topik sebanyak 3 dan Algoritma *Latent Dirichlet Allocation* mampu dan cukup baik dalam mengidentifikasi topik pada sebuah *tweet*.

Kata kunci: Tokoh Publik Politik, Topik Modeling, *Latent Dirichlet Allocation*

ABSTRACT

Topic Modeling is a technique that can be used to analyze large and small documents. Topic Modeling can understand the latent topics in a document using random variables, and discover the structure of the document. Topic Modeling has various applications such as extracting keywords, categorizing documents, tag recommendations and others. In this research, the dataset used comes from twitter which is collected from several accounts belonging to Indonesian political public figures. The purpose of this study is to determine the optimal number of clusters in modeling topics from tweets of political public figures using the Latent Dirichlet Allocation algorithm. The stages of this research start from data collection using scraping the twitter api model, then the preprocessing stage such as removing stopwords, removing emojis, tokenization, and stemming, then the next stage of the topic modeling process using the Latent Dirichlet Allocation method and finally the research results in the form of a topic modeling graph. From the experiments conducted with topics of 2, 3, 5, and 7 and iterations of 100, 200, 500, and 1000. The best cluster results were obtained with 3 topics and the Latent Dirichlet Allocation Algorithm was able and good enough to identify topics in a tweet.

Keywords: *political public figures, Topic Modeling, Latent Dirichlet Allocation*