

**RANCANG BANGUN APLIKASI UNTUK MENGELOMPOKKAN
TOPIK TWEET DARI TWITTER DENGAN MENGGUNAKAN
PEMODELAN AUTHOR-TOPIC MODEL (STUDI KASUS: AKUN
TWITTER POLITISI INDONESIA)**

SKRIPSI

Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar Sarjana

Komputer Program Studi Informatika



Oleh:

MUHAMMAD RIFQI NAFIS

NPM. 17081010059

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR**

2021

**RANCANG BANGUN APLIKASI UNTUK MENGELOMPOKKAN
TOPIK TWEET DARI TWITTER DENGAN MENGGUNAKAN
PEMODELAN AUTHOR-TOPIC MODEL (STUDI KASUS: AKUN
TWITTER POLITISI INDONESIA)**

SKRIPSI



Oleh:

MUHAMMAD RIFQI NAFIS

NPM. 17081010059

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2021**

LEMBAR PENGESAHAN SKRIPSI

Judul : RANCANG BANGUN APLIKASI UNTUK MENGELOMPOKKAN TOPIK TWEET DARI TWITTER DENGAN MENGGUNAKAN PEMODELAN AUTHOR-TOPIC MODEL (STUDI KASUS: AKUN TWITTER POLITISI INDONESIA)

Oleh : MUHAMMAD RIFQI NAFIS

NPM : 17081010059

Telah Diseminarkan Dalam Ujian Skripsi Pada :
Hari Jum'at, Tanggal 15 Januari 2021

Mengetahui

Dosen Pembimbing

1.



Rizky Parlika, S.Kom, M.Kom
NPT : 3 8405 07 0219 1

2.



Sugiarto, S.Kom, M.Kom
NPT : 3 8702 13 0343 1

Dosen Penguji

1.



Firza Prima Aditiawan, S.Kom, M.TI
NPT : 3 8605 13 0344 1

2.



Pratama Wirya Atmaja, S.Kom, M.Kom
NIP : 19840106 201803 1 001

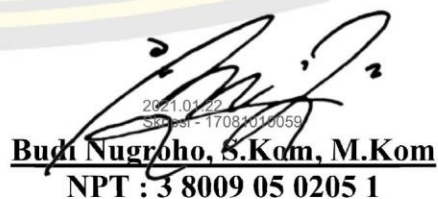
Menyetujui

Dekan
Fakultas Ilmu Komputer



Dr. Ir. Ketut Sari, M.I
NPT : 10650731 1199203 2 001

Koordinator Program Studi
Teknik Informatika



2021.01.22
SKRIPSI - 17081010059

Budi Nugroho, S.Kom, M.Kom
NPT : 3 8009 05 0205 1

SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa Informatika UPN “Veteran” Jawa Timur, yang bertandatangan di bawah ini:

Nama : MUHAMMAD RIFQI NAFIS

NPM : 17081010059

Menyatakan bahwa Judul Skripsi/ Tugas Akhir yang Saya ajukan dan akan dikerjakan, yang berjudul:

**“RANCANG BANGUN APLIKASI UNTUK MENGELOMPOKKAN
TOPIK TWEET DARI TWITTER DENGAN MENGGUNAKAN
PEMODELAN AUTHOR-TOPIC MODEL (STUDI KASUS: AKUN
TWITTER POLITISI INDONESIA)”**

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk dan atau *software* yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN “Veteran” Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka Saya siap menerima segala konsekuensinya.

Surabaya, 15 Januari 2021

Hormat Sava,

MUHAMMAD RIFQI NAFIS
NPM. 17081010059

**RANCANG BANGUN APLIKASI UNTUK MENGELOMPOKKAN TOPIK
TWEET DARI TWITTER DENGAN MENGGUNAKAN PEMODELAN
AUTHOR-TOPIC MODEL (STUDI KASUS: AKUN TWITTER POLITISI
INDONESIA)**

Nama : Muhammad Rifqi Nafis
NPM : 17081010059
Program Studi : Teknik Informatika
Dosen Pembimbing : 1. Rizky Parlika, S.Kom., M.Kom
2. Sugiarto, S.Kom., M.Kom.

ABSTRAK

Twitter merupakan sebuah platform untuk berhubungan satu sama lain agar tetap berkomunikasi dan tetap terhubung dengan mengirim pesan cepat dan daring. Pada Twitter juga dapat membuat *tweet* yang berisi foto, video, tautan, atau teks. Pada aplikasi Twitter pengguna cenderung lebih mengutarakan isi hati yang sebenarnya dibandingkan dengan platform media sosial lainnya. Pada tahun 2020 ini telah terjadi wabah atau pandemi Covid-19 dan pada akhir tahun bulan Desember akan diselenggarakan pilkada serentak seluruh Indonesia. Dari hal tersebut dibutuhkannya sebuah platform yang mampu memberikan informasi visual terhadap aktivitas politisi dalam hal mengutarakan isi hatinya di media sosial Twitter dengan melakukan *topic modelling* terhadap aktivitas politisi tersebut ketika membuat sebuah tulisan atau *tweet* menggunakan metode *Author-Topic Models*. Penelitian ini dikhususkan untuk menganalisa data *tweet* akun Twitter politisi di Indonesia, proses pengambilan data dilakukan dengan cara *crawling* data pada *user timeline* aplikasi Twitter. Setelah data didapatkan kemudian akan dianalisa dengan menggunakan *Author-Topic Models* dan dilakukan visualisasi terhadap topik pembahasan berdasarkan *tweet* yang telah dibuat oleh masing-masing akun Twitter politisi. Melalui proses *topic modelling*, telah didapatkan hasil terbaik berupa 3 topik. Model 3 topik tersebut dikatakan terbaik karena memiliki nilai *perplexity* terendah diantara jumlah topik lain yang diuji. 3 topik yang terbentuk dianalisis dan diterjemahkan ke dalam label kategori, yaitu “Rakyat Tolak RUU Cipta Kerja”, “Ucapan Syukur saat Pandemi”, “Kondisi Ekonomi saat Pandemi” dengan jumlah data *tweet* yang digunakan adalah 24846 data *tweet* dari 36 akun Twitter politisi.

Kata kunci: Twitter, Tweet, Politisi, Topic Modelling, Author-Topic Models

KATA PENGANTAR

Segala puji dan syukur bagi Allah SWT atas ridho dan rahmat-Nya, penulis dapat menyelesaikan laporan skripsi dengan judul “RANCANG BANGUN APLIKASI UNTUK MENGELOMPOKKAN TOPIK TWEET DARI TWITTER DENGAN MENGGUNAKAN PEMODELAN AUTHOR-TOPIC MODELS (STUDI KASUS: AKUN TWITTER POLITISI INDONESIA)” dengan baik. Skripsi ini ditujukan untuk memenuhi salah satu syarat kelulusan program Sarjana di Jurusan Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam pembuatan laporan skripsi ini penulis menyadari bahwa penulisan laporan skripsi ini masih jauh dari kata sempurna, masih banyak kekurangan mengingat keterbatasan pengetahuan dan kemampuan penulis. Penulis juga mengharapkan hasil dari pengerjaan skripsi ini dapat bermanfaat bagi pembaca perihal permasalahan politik yang sedang terjadi di Negara Kesatuan Republik Indonesia tercinta ini.

Dalam penyusunan skripsi ini juga tidak terlepas dukungan dari berbagai pihak. Untuk itu, secara khusus penulis akan menyampaikan ucapan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu. Penulis banyak menerima bimbingan, petunjuk, dan bantuan serta dorongan dari berbagai pihak baik yang bersifat moral maupun material. Pada kesempatan ini penulis menyampaikan rasa terima kasih yang sedalam-dalamnya kepada:

1. Keluarga penulis terutama kedua orang tua, Bapak Suwandi dan Ibu Dra. Siti Nur Hasanah yang selalu senantiasa mendoakan, memberikan motivasi, dan kebutuhan materiil maupun non-materiil serta kakak saya

Faiz Nur Fitrah Insani yang telah membantu dengan memberikan ide dan masukan sehingga penulis mampu untuk menyelesaikan pendidikan S1 ini dengan baik.

2. Prof. Dr. Ir. Ahmad Fauzi, M.MT. selaku Rektor UPN “Veteran” Jawa Timur.
3. Ibu Dr. Ir. Ni Ketut Sari, MT. selaku Dekan Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
4. Bapak Budi Nugroho, S.Kom., M.Kom. selaku Koordinator Program Studi Teknik Informatika Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
5. Bapak Rizky Parlika, S.Kom., M.Kom. selaku Dosen Pembimbing 1 yang telah meluangkan waktu untuk memberikan ilmu dan dorongan serta motivasi kepada penulis dalam menyelesaikan tugas akhir ini dari awal hingga akhir.
6. Bapak Sugiarto, S.Kom., M.Kom. selaku Dosen Pembimbing 2 yang telah meluangkan waktu untuk memberikan ilmu dan dorongan serta motivasi kepada penulis dalam menyelesaikan tugas akhir ini dari awal hingga akhir.
7. Teman-teman seperjuangan saya Teknik Informatika angkatan 2017 terutama Herdi, Harris, Bimo, Taruna, dan Taufiq yang telah memberikan semangat dan motivasinya untuk segera menyelesaikan tugas akhir ini.

Penulis juga ucapkan terima kasih kepada segenap dosen dan karyawan di Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah membimbing dan memberi bekal ilmunya selama ini serta

memberikan kemudahan dalam setiap kegiatan akademik. Terima kasih sekali lagi untuk teman-teman dari Teknik Informatika UPN “Veteran” Jawa Timur angkatan 2017 yang solid dan selalu membantu penulis apabila terdapat kesulitan selama masa perkuliahan. Terima kasih atas bantuannya semoga Allah SWT membalas semua kebaikan dan bantuan yang telah diberikan kepada penulis.

Surabaya, 15 Januari 2021

Muhammad Rifqi Nafis

DAFTAR ISI

LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN ANTI PLAGIAT	iii
ABSTRAK	iv
KATA PENGANTAR	v
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
DAFTAR KODE.....	xv
BAB I	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
BAB II.....	6
TINJAUAN PUSTAKA.....	6
2.1. Hasil Penelitian Terdahulu	6
2.2. Landasan Teori.....	8
2.2.1. Twitter	8
2.2.2. Web Crawler	9
2.2.3. Natural Language Processing.....	10
2.2.4. Topic Modelling.....	10
2.2.5. Latent Dirichlet Allocation	11
2.2.6. Author-Topic Models.....	13
2.2.7. Perplexity	15
2.2.8. Agile Development Model Extreme Programming	16
BAB III	17
METODOLOGI.....	17

3.1.	Metodologi Penelitian	17
3.1.1.	Studi Literatur	18
3.1.2.	Pengumpulan Data	18
3.1.3.	Topic Modelling Menggunakan Author-Topic Models	19
3.1.3.1.	Case Folding	20
3.1.3.2.	Cleansing	21
3.1.3.3.	Stemming	21
3.1.3.4.	Tokenization	22
3.1.3.5.	Filtering (Stopword Removal)	23
3.1.4.	Perancangan dan Pembuatan Visualisasi dengan Website.....	24
3.1.4.1.	Planning atau Perencanaan	25
3.1.4.2.	Design atau Rancangan.....	26
3.1.4.3.	Coding atau Rancang Bangun.....	26
3.1.4.4.	Testing atau Pengujian.....	27
3.1.5.	Penyusunan Laporan Skripsi.....	27
3.2.	Urutan Pelaksanaan Percobaan	27
3.3.1.	Crawling Data	28
3.3.2.	Praproses data.....	29
3.3.3.	Proses Data.....	29
3.3.3.1.	Pembentukan Author2doc.....	30
3.3.3.2.	Pembentukan Dictionary dan Corpus	31
3.3.3.3.	Pemodelan Topik dengan Author-Topic Model	32
3.3.4.	Visualisasi Data.....	32
3.3.4.1.	Arsitektur Aplikasi.....	33
3.3.4.2.	Perancangan Sistem	33
3.3.4.3.	Perancangan Database	47
3.3.4.4.	User Interface Aplikasi	53
BAB IV		57
HASIL DAN PEMBAHASAN		57
4.1.	Bahan dan Peralatan.....	57
4.2.	Data Penelitian	59
4.2.1.	Pengumpulan Username Akun.....	59

4.2.2.	Crawling Data	61
4.2.3.	Menyimpan Data.....	62
4.2.4.	Memuat Data.....	62
4.2.5.	Hasil Crawling Data.....	62
4.3.	Praproses Data.....	63
4.3.1.	Case Folding	63
4.3.2.	Cleansing.....	63
4.3.3.	Stemming	64
4.3.4.	Tokenization.....	64
4.3.5.	Stopword Filtering	64
4.3.6.	Hasil Praproses Data	66
4.4.	Proses Data.....	66
4.4.1.	Pembuatan Dictionary Author2doc.....	67
4.4.2.	Pembuatan Dictionary dan Korpus data tweet.....	67
4.5.	Pemodelan Topik dengan Author-Topic Models.....	68
4.5.1.	Uji Coba Pembuatan Model Topik	68
4.5.2.	Menyimpan dan Memuat Model.....	69
4.5.3.	Validasi Model Topik	69
4.6.	Analisa Topik	71
4.7.	Visualisasi Data.....	74
BAB V.....		82
KESIMPULAN DAN SARAN.....		82
5.1.	Kesimpulan	82
5.2.	Saran.....	83
DAFTAR PUSTAKA		84
LAMPIRAN.....		86
BIODATA PENULIS		92

DAFTAR GAMBAR

Gambar 2.1 Arsitektur Web Crawler.	9
Gambar 2.2 Natural Language Processing.....	10
Gambar 2.2 Konsep Topic Modelling.....	11
Gambar 2.3 Model Representasi Latent Dirichlet Allocation.....	12
Gambar 2.4 Konsep Author-Topic Model.	13
Gambar 2.5 Contoh 4 topik dari <i>dataset</i> NIPS menampilkan 10 kata teratas dan 10 author teratas berdasarkan topik.....	14
Gambar 2.6 Contoh <i>author</i> terpilih dari <i>dataset</i> NIPS menampilkan 4 topik yang memiliki probabilitas teratas pada setiap <i>author</i>	15
Gambar 2.7 Tahapan Extreme Programming.....	16
Gambar 3.1 Metodologi Penelitian Skripsi.....	17
Gambar 3.2 Proses pengumpulan data <i>tweet</i> dari akun Twitter.....	19
Gambar 3.3 Tahapan Topic Modelling dengan Author-Topic Models.....	19
Gambar 3.4 Alur proses <i>crawling</i> data	28
Gambar 3.5 Alur praproses data.....	29
Gambar 3.6 Alur pemodelan topik dengan Author-Topic Model.....	30
Gambar 3.7 Pembentukan Author2doc list.	31
Gambar 3.8 Proses pembentukan dictionary dan corpus.	31
Gambar 3.9 Alur pemodelan topik dengan <i>Author-Topic Model</i>	32
Gambar 3.10 Arsitektur Aplikasi	33
Gambar 3.11 Use Case Diagram Sistem Aplikasi Visualisasi Pemodelan Topik.	33
Gambar 3.12 Activity Diagram untuk Melihat Data Tweet.....	40
Gambar 3.13 Activity Diagram untuk Mencari Data Tweet.....	41
Gambar 3.14 Activity Diagram untuk Melihat Grafik Tweet.....	41
Gambar 3.15 Activity Diagram untuk Melihat Data Author.	42
Gambar 3.16 Activity Diagram untuk Mencari Data Author.	42
Gambar 3.17 Activity Diagram untuk Melihat Grafik Author.	43
Gambar 3.18 Activity Diagram untuk Melihat Diagram Topik.....	43
Gambar 3.19 Activity Diagram untuk Melihat Diagram Wordcloud.	44

Gambar 3.20 Class Diagram Controller.....	44
Gambar 3.21 Class Diagram Model.....	45
Gambar 3.22 Sequence Diagram Melihat dan Mencari Data Tweet.	46
Gambar 3.23 Sequence Diagram Melihat dan Mencari Data Author.	46
Gambar 3.24 Sequence Diagram Melihat Diagram Topik.....	47
Gambar 3.25 ERD Database	51
Gambar 3.26 CDM database	52
Gambar 3.27 PDM database.	52
Gambar 3.28 Mockup halaman utama.	53
Gambar 3.29 Mockup Daftar Data Tweet.....	54
Gambar 3.30 Mockup Daftar Data Author.	54
Gambar 3.31 Mockup Diagram Topik.	55
Gambar 3.32 Mockup Diagram Authortopik.	55
Gambar 3.33 Diagram Wordcloud.	56
Gambar 4.1 Hasil percobaan menggunakan 100 passes dengan 1 hingga 20 jumlah topik.	70
Gambar 4.2 Halaman utama <i>website</i>	75
Gambar 4.3 Daftar <i>tweet</i> beserta fitur pencarian.	75
Gambar 4.4 Jumlah <i>tweet</i> per bulan pada tahun 2020.	76
Gambar 4.5 Jumlah <i>tweet</i> per hari pada tahun 2020.	76
Gambar 4.6 Jumlah <i>tweet</i> per jam pada tahun 2020.	77
Gambar 4.7 Daftar <i>author</i> beserta fitur pencarian.	77
Gambar 4.8 Jumlah <i>author</i> berdasarkan umur.....	78
Gambar 4.9 Grafik Topik 0 beserta kata dan probabilitasnya.	78
Gambar 4.10 Grafik Topik 0 beserta <i>author</i> dan probabilitasnya.	79
Gambar 4.11 Grafik Topik 1 beserta kata dan probabilitasnya.	79
Gambar 4.12 Grafik Topik 1 beserta <i>author</i> dan probabilitasnya.	80
Gambar 4.13 Grafik Topik 2 beserta kata dan probabilitasnya.	80
Gambar 4.14 Grafik Topik 2 beserta <i>author</i> dan probabilitasnya.	81
Gambar 4.15 Interface Word Cloud.....	81

DAFTAR TABEL

Tabel 3.1 Teks sebelum dan sesudah Case Folding.....	20
Tabel 3.2 Teks sebelum dan sesudah Cleansing	21
Tabel 3.3 Teks sebelum dan sesudah Stemming.....	22
Tabel 3.4 Teks sebelum dan sesudah Tokenization	23
Tabel 3.5 Teks sebelum dan sesudah Filtering	24
Tabel 3.9 Use Case Specification Melihat Data Tweet.....	36
Tabel 3.10 Use Case Specification Mencari Data Tweet.....	36
Tabel 3.11 Use Case Specification Melihat Grafik Tweet.....	37
Tabel 3.12 Use Case Specification Melihat Data Author.	37
Tabel 3.13 Use Case Specification Mencari Data Author.	38
Tabel 3.14 Use Case Specification Melihat Grafik Author.	38
Tabel 3.15 Use Case Specification Melihat Diagram Topik.....	39
Tabel 3.16 Use Case Specification Melihat Diagram Wordcloud.	39
Tabel 3.17 Struktur Tabel Akun	47
Tabel 3.18 Struktur Tabel Tweet	48
Tabel 3.19 Struktur Tabel Praproses.....	48
Tabel 3.20 Struktur Tabel Topik.....	49
Tabel 3.21 Struktur Tabel Toptopik.....	49
Tabel 3.22 Struktur Tabel Authortopik.....	49
Tabel 3.23 Struktur Tabel Wordcloud	50
Tabel 4.1 Spesifikasi perangkat keras.....	57
Tabel 4.2 Perangkat lunak yang digunakan	58
Tabel 4.3 Library yang digunakan	58
Tabel 4.4 Data <i>username</i> yang terkumpul.	59
Tabel 4.5 Detail data hasil <i>crawling</i>	62
Tabel 4.6 Kata Depan.....	65
Tabel 4.7 Kata Ganti.	65
Tabel 4.8 Kata Sambung.....	65
Tabel 4.9 Kata Sandang	65

Tabel 4.10 Hasil Praproses Data.	66
Tabel 4.11 Toptopik dari 3 jumlah topik dengan 10 kata tiap topiknya.	71
Tabel 4.12 Pemberian label pada tiap topik.	73
Tabel 4.13 Probabilitas <i>author</i> yang tertinggi pada tiap topik.	73

DAFTAR KODE

Kode 1 Potongan <i>script</i> untuk menyimpan token <i>API</i> ke dalam variabel.	86
Kode 2 Potongan <i>script crawling</i> data berdasarkan <i>username</i>	86
Kode 3 Potongan <i>script</i> untuk menyimpan data ke dalam <i>database</i>	87
Kode 4 Potongan <i>script</i> memuat data dari <i>database</i>	87
Kode 5 keseluruhan <i>script</i> dalam praproses data.	88
Kode 6 Potongan <i>script</i> untuk <i>case folding</i>	88
Kode 7 Potongan <i>script</i> untuk data <i>cleansing</i>	88
Kode 8 Potongan <i>script</i> untuk <i>stemming</i>	88
Kode 9 Potongan <i>script</i> untuk <i>tokenization</i>	88
Kode 10 Potongan <i>script</i> untuk <i>stopword filtering</i>	88
Kode 11 keseluruhan <i>script</i> dalam proses data.	89
Kode 12 Potongan <i>script</i> untuk membuat <i>dictionary author2doc</i>	89
Kode 13 Potongan <i>script</i> untuk membuat <i>dictionary</i> kata.	89
Kode 14 Potongan <i>script</i> untuk membuat korpus.	89
Kode 15 <i>Script</i> untuk uji coba pembentukan model topik.	90
Kode 16 <i>Script</i> untuk dokumentasi <i>logging</i>	90
Kode 17 <i>Script</i> untuk menyimpan dan memuat model topik.	90
Kode 18 <i>Script</i> model topik terbaik dengan topik berjumlah 3.	90
Kode 19 <i>Script</i> untuk menampilkan topik teratas.	90
Kode 20 <i>Script</i> untuk menampilkan topik secara spesifik.	90
Kode 21 <i>Script</i> untuk memberi label pada topik.	90
Kode 22 <i>Script</i> untuk menampilkan <i>author</i> beserta probabilitas tiap topik.	91