

**PEMODELAN TOPIK TERKAIT BANJIR PADA TWITTER DENGAN
MENGUNAKAN *LATENT DIRICHLET ALLOCATION***

TUGAS AKHIR

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana
Program Studi Sistem Informasi



Disusun oleh:

MUHAMMAD SUTAN IRWANSYAH

2017.02779.12.0009

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK & TEKNOLOGI INFORMASI
UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA
2022**

HALAMAN PENGESAHAN

TUGAS AKHIR

**PEMODELAN TOPIK TERKAIT BANJIR PADA TWITTER DENGAN
MENGUNAKAN *LATENT DIRICHLET ALLOCATION***

Diajukan oleh:

MUHAMMAD SUTAN IRWANSYAH

2017.02779.12.0009


Telah dipertahankan di depan Dewan Penguji dan dinyatakan sah sebagai salah satu syarat untuk memperoleh gelar Sarjana di Fakultas Teknik & Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta

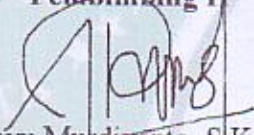
Tanggal: 29 Agustus 2022

Mengesahkan:

Pembimbing I

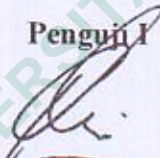
Pembimbing II

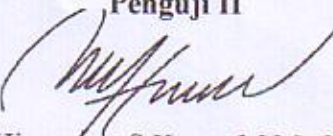

Muhammad Habibi, S.Kom., M.Cs
NIDN: 0525018901


Aris Wahyu Murtiyanto, S.Kom., M.Cs
NIDN: 0502098501

Penguji I

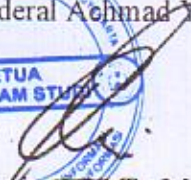
Penguji II


Kharisma, S.T., M.Cs.
NIDN: 0502108201


Arif Himawan, S.Kom., M.M., M.Eng.
NIDN: 0517127402

Ketua Program Studi S-1 Sistem Informasi
Fakultas Teknik & Teknologi Informasi
Universitas Jenderal Achmad Yani Yogyakarta




Kharisma, S.T., M.Cs.
NPP: 2016.13.0082

HALAMAN PERNYATAAN

Saya yang bertanda tangan dibawah ini, adalah mahasiswa Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta.

Nama : Muhammad Sutan Irwansyah

NPM : 2017.02779.12.0009

Program Studi : Sistem Informasi (S1)

Judul TA : Pemodelan Topik Terkait Banjir Pada Twitter Dengan Menggunakan *Latent Dirichlet Allocation*

Menyatakan bahwa Laporan Tugas Akhir dengan judul tersebut diatas merupakan karya saya sendiri dan bukan hasil plagiarisme, bebas dari peniruan terhadap karya orang lain. Semua referensi dan sumber terkait dari karya tulis ini yang merupakan kutipan dari hasil karya orang lain yang telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan karya ilmiah.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa dalam Tugas Akhir ini terkandung ciri-ciri plagiat dan bentuk-bentuk peniruan lain yang dianggap melanggar peraturan, maka saya bersedia menerima sanksi atas ketentuan perbuatan tersebut.

Yogyakarta, 29 Agustus 2022



Muhammad Sutan Irwansyah

KATA PENGANTAR

Assalamua'alaikum Warahmatullahi Wabarokatuh.

Puji Syukur Alhamdulillah penulis panjatkan kepada Allah SWT yang telah memberikan rahmat dan karunia-Nya, sehingga disini penulis dapat menyelesaikan laporan penelitian tugas akhir yang berjudul “Pemodelan Topik Terkait Banjir pada Twitter Dengan Menggunakan Latent Dirichlet Allocation (LDA)”.

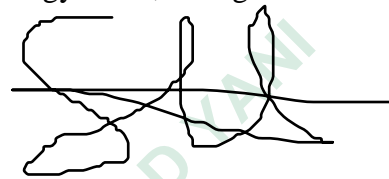
Laporan penelitian ini diajukan sebagai salah satu syarat untuk memenuhi kelulusan jenjang S-1 di Universitas Jenderal Achmad Yani Yogyakarta Prodi FTTI Jurusan Sistem Informasi.

Penulis disini ingin menyampaikan terima kasih kepada Orang Tua serta Keluarga besar saya yang selalu memberi semangat kepada saya. Pada kesempatan ini pula, penulis ingin menyampaikan kata terima kasih yang banyak banyak atas bantuannya kepada penulis selama ini, antara lain kepada yang terhormat :

1. Bapak Dr. Drs. Djoko Susilo, S.T., M.T. selaku Rektor Universitas Jenderal Achmad Yani Yogyakarta.
2. Bapak Aris Wahyu Murdiyanto, S.Kom., M.Cs. selaku Dekan Fakultas Teknik dan Teknologi Informasi (FTTI).
3. Bapak Kharisma, S.T., M.Cs. selaku Kepala Prodi Sistem Informasi S1.
4. Bapak Muhammad Habibi, S.Kom., M.Cs. selaku pembimbing tugas akhir.
5. Seluruh Dosen serta Staff Administrasi yang telah memberikan ilmu dan kemudahan dalam penyelesaian tugas akhir.
6. Teman-teman yang telah mendukung, membantu, dan memberikan motivasi untuk menyelesaikan naskah tugas akhir.

Peneliti disini menyadari bahwa masih terdapat kekurangan dalam penulisan atau penyusunan aplikasi program ini. Oleh sebab itu, dibutuhkan kritik serta saran yang dapat membangun dari semua pihak.

Yogyakarta, 29 Agustus 2022

A handwritten signature in black ink, consisting of several overlapping loops and a horizontal line, positioned above the name.

Muhammad Sutan Irwansyah

UNIVERSITAS JENDERAL ACHMAD YAN
PEPUSTAKAAN
YOGYAKARTA

DAFTAR ISI

Halaman Judul	i
Halaman Pengesahan.....	ii
Halaman Pernyataan	iii
Kata Pengantar	iv
Daftar Isi	vi
Daftar Tabel.....	viii
Daftar Gambar	ix
Daftar Lampiran	xi
Daftar Singkatan	xii
Intisari	xiii
Abstract.....	xiv
Bab 1 Pendahuluan	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Manfaat Hasil Penelitian	3
1.4 Tujuan Penelitian	3
1.5 Pertanyaan Penelitian	3
Bab 2 Tinjauan Pustaka dan Landasan Teori.....	4
2.1 Tinjauan Pustaka	4
2.2 Landasan Teori.....	5
2.2.1 Data Mining.....	5
2.2.2 Tahapan Data Mining	6
2.2.3 Text Mining.....	7
2.2.4 Preprocessing	7
2.2.5 Feature Extraction	8
2.2.6 Latent Dirichlet Allocation.....	8
2.2.7 Topic Coherence	9
2.2.8 Topic Analysis.....	10
Bab 3 Metode Penelitian.....	11

3.1	Bahan Penelitian.....	11
3.2	Alat Penelitian.....	11
3.3	Jalan Penelitian.....	13
3.3.1	Web Data Extraction	14
3.3.2	<i>Preprocessing</i>	15
3.3.3	Feature Extraction menggunakan TF-IDF	19
3.3.4	<i>Topic Modelling</i> Menggunakan LDA	20
3.3.5	Visualisasi	21
3.3.6	<i>Topic Analysis</i>	23
Bab 4	Hasil Penelitian.....	25
4.1	Ringkasan Hasil Penelitian	25
4.2	Pembahasan Hasil Topic Coherence.....	25
4.3	Hasil Wordcloud Pertopik.....	27
4.4	Hasil Analisis Pertopik.....	42
4.5	Hasil Wordcloud Keseluruhan Topic.....	45
Bab 5	Kesimpulan dan Saran	48
5.1	Kesimpulan	48
5.2	Saran.....	48
	Daftar Pustaka.....	49
	Lampiran	51

DAFTAR TABEL

Tabel 3.1 Daftar kata stopwords.....	17
Tabel 4. 1 Nilai Coherence	26
Tabel 4. 2 Contoh <i>tweet</i> terkait Topik 1	27
Tabel 4. 3 Contoh <i>tweet</i> terkait Topik 2	29
Tabel 4. 4 Contoh <i>tweet</i> terkait Topik 3	30
Tabel 4. 5 Contoh <i>tweet</i> terkait Topik 4	32
Tabel 4. 6 Contoh <i>tweet</i> terkait Topik 5	34
Tabel 4. 7 Contoh <i>tweet</i> terkait Topik 6	35
Tabel 4. 8 Contoh <i>tweet</i> terkait Topik 7	37
Tabel 4. 9 Contoh <i>tweet</i> terkait Topik 8	38
Tabel 4. 10 Contoh <i>tweet</i> terkait Topik 9	40
Tabel 4. 11 Contoh <i>tweet</i> terkait Topik 10	41
Tabel 4. 12 Kata-kata Pertopik	42
Tabel 4. 13 Contoh <i>tweet</i> terkait kata yang paling banyak muncul.....	46

DAFTAR GAMBAR

Gambar 2.1 Tahapan data mining	6
Gambar 2.2 Visualisasi <i>Topic Modelling</i> memakai metode LDA.....	9
Gambar 3.1 Alur Penelitian	14
Gambar 3.2 Library proses pengolah data.....	15
Gambar 3.3 Perintah import data.....	15
Gambar 3.4 Library preprocessing	16
Gambar 3.5 Tahapan Cleaning	16
Gambar 3.6 Tahapan Casefolding serta Stopwords.....	18
Gambar 3.7 Library Tokenizing	18
Gambar 3.8 Fungsi dari token Tokenizing	19
Gambar 3.9 Perulangan data dari range variabel df	19
Gambar 3.10 Perulangan dan menambahkan data tokens	19
Gambar 3.11 Membuat kamus dari kumpulan beberapa tulisan	20
Gambar 3.12 Perhitungan TF-IDF.....	20
Gambar 3.13 Library gensim dan menjalankan LDA	20
Gambar 3.14 Memasukkan kata-kata pertopik kedalam list	21
Gambar 3.15 Diagram LDA	21
Gambar 3.16 Library untuk modul wordcloud.....	22
Gambar 3.17 Wordcloud keseluruhan topik.....	22
Gambar 3.18 Wordcloud pertopik	22
Gambar 3.19 Memanggil topik LDA yang sudah ada.....	23
Gambar 3.20 Memasukkan library gensim dan fungsi dari <code>evaluate_graph</code>	23
Gambar 3.21 Menampilkan diagram Topic Coherence.....	24
Gambar 4. 1 Diagram grafik nilai coherence	25
Gambar 4. 2 Wordcloud Topik ke 1	27
Gambar 4. 3 <i>Wordcloud</i> Topik ke 2	28
Gambar 4. 4 <i>Wordcloud</i> Topik ke 3	30
Gambar 4. 5 <i>Wordcloud</i> Topik ke 4	32
Gambar 4. 6 <i>Wordcloud</i> Topik ke 5	33

Gambar 4. 7 <i>Wordcloud</i> Topik ke 6	35
Gambar 4. 8 <i>Wordcloud</i> Topik ke 7	36
Gambar 4. 9 <i>Wordcloud</i> Topik ke 8	38
Gambar 4. 10 <i>Wordcloud</i> Topik ke 9	39
Gambar 4. 11 <i>Wordcloud</i> Topik ke 10	41
Gambar 4. 12 <i>Wordcloud</i> keseluruhan topik	45

PEPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA

DAFTAR LAMPIRAN

LAMPIRAN 1 JADWAL PENELITIAN.....	51
LAMPIRAN 2 KARTU BIMBINGAN TUGAS AKHIR.....	52
LAMPIRAN 3 HASIL CEK PLAGIARISME	54

UNIVERSITAS JENDERAL ACHMAD YANI
PEPUSTAKAAN
YOGYAKARTA

DAFTAR SINGKATAN

LDA	Latent Dirichlet Allocation
TF-IDF	Term Frequency-Inverse Document Frequency

PEPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA