

**ANALISIS SENTIMEN BERDASARKAN TOPIK TERKAIT WABAH
COVID-19 DI TWITTER MENGGUNAKAN LATENT DIRICHLET
ALLOCATION (LDA) DAN NAIVE BAYES CLASSIFIER (NBC)**

TUGAS AKHIR

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana
Program Studi Informatika



Disusun oleh:

PANGKY PUTRA AZIZTIYA

182102028

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

HALAMAN PENGESAHAN

TUGAS AKHIR

**ANALISIS SENTIMEN BERDASARKAN TOPIK TERKAIT WABAH
COVID-19 DI TWITTER MENGGUNAKAN LATENT DIRICHLET
ALLOCATION (LDA) DAN NAIVE BAYES CLASSIFIER (NBC)**

Diajukan oleh:

PANGKY PUTRA AZIZTIYA


182102001

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: 25 Agustus 2022

Mengesahkan:


Pembimbing I


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

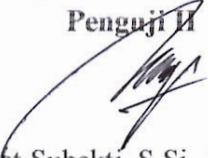
Pembimbing II


Puji Winar Cahyo, S.Kom., M.Cs.
NIDN: 0519119003


Penguji I


Choerun Asnawi, S.Kom., M.Kom.
NIDN: 0011077702

Penguji II


Dayat Subekti, S.Si., M.Kom.
NIDN: 0507037401

Ketua Program Studi Informatika
Fakultas Teknik & Teknologi Informasi
Universitas Jenderal Achmad Yani Yogyakarta


Puji Winar Cahyo, S.Kom., M.Cs.
NPP: 2018.13.0109

PERNYATAAN

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

Nama : Pangky Putra Aziztiya
NPM : 182102028
Program Studi : Informatika (S-1)
Judul Tugas Akhir : Analisis sentimen berdasarkan topik terkait wabah COVID-19 di Twitter menggunakan Latent Dirichlet Allocation (LDA) dan Naïve Bayes Classifier (NBC)

Menyatakan bahwa hasil penelitian dengan judul tersebut di atas adalah asli karya saya sendiri dan bukan hasil plagiarisme. Semua referensi dan sumber terkait yang dikutip dalam karya ilmiah ini telah ditulis sesuai kaidah penulisan ilmiah yang berlaku. Dengan ini, saya menyatakan untuk menyerahkan hak cipta penelitian kepada Universitas Jenderal Achmad Yani Yogyakarta guna kepentingan ilmiah.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya tanpa ada paksaan dari pihak mana pun. Apabila terdapat kekeliruan atau ditemukan adanya pelanggaran akademik di kemudian hari, maka saya bersedia menerima konsekuensi yang berlaku sesuai ketentuan akademik.

Yogyakarta, 25 Agustus 2022



Pangky Putra Aziztiya

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul: “Analisis sentiment berdasarkan topik terkait wabah COVID-19 di Twitter menggunakan menggunakan Latent Dirichlet Allocation (LDA) dan Naïve Bayes Classifier (NBC)”. Penyusunan laporan ini merupakan salah satu persyaratan untuk menyelesaikan studi di Program Studi Informatika (S-1) Fakultas Teknik & Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta. Laporan ini dapat diselesaikan atas bimbingan, arahan, dan bantuan dari berbagai pihak. Pada kesempatan ini penulis dengan rendah hati mengucapkan terima kasih dengan setulus-tulusnya kepada:

1. Bapak Aris Wahyu Murdiyanto, S.Kom., M.Cs. selaku Dekan Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
2. Bapak Puji Winar Cahyo, S.Kom., M.Cs. selaku Ketua Program Studi Informatika (S-1) Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
3. Bapak Muhammad Habibi, S.Kom., M.Cs. selaku Dosen Pembimbing Tugas Akhir;
4. Para dosen yang telah memberikan banyak bekal ilmu pengetahuan kepada penulis selama menjadi mahasiswa di Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
5. Ayah, ibu, dan adik saya, Clara Astrid Mida Aprodid Octavia, yang telah memberikan dukungan semangat serta doa restu kepada saya, sehingga dapat menyelesaikan studi saya;
6. Sahabatku Lintang Permadi, Ayudya, Gilang berilian, dan Rizky Fauzi yang telah memberikan semangat dan doa kepada saya;
7. Rekan-rekan mahasiswa Informatika (S-1) di Universitas Jenderal Achmad Yani Yogyakarta yang sudah memberi dukungan dan kerja sama selama pembuatan tugas akhir.

Penulis menyadari bahwa laporan tugas akhir ini masih jauh dari kata sempurna. Maka dari itu dengan segala kerendahan hati penulis sangat menghargai adanya kritik dan saran yang membangun dari semua pihak yang bersedia meluangkan waktu untuk membaca laporan tugas akhir ini.

Yogyakarta, 31 Agustus 2022



Pangky Putra Aziztiya

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA

DAFTAR ISI

Halaman Judul	i
Halaman Pengesahan.....	ii
Halaman Pernyataan	iii
Kata Pengantar	iv
Daftar Isi	vi
Daftar Tabel.....	ix
Daftar Gambar	xi
Daftar Lampiran	xii
Daftar Singkatan	xiii
Intisari	xiv
Abstract	xv
Bab 1 Pendahuluan	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Pertanyaan Penelitian	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Hasil Penelitian	4
Bab 2 Tinjauan Pustaka Dan Landasan Teori.....	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori.....	7
2.2.1 Twitter	7
2.2.2 Python.....	8
2.2.3 Flask	8
2.2.4 Data Mining.....	8
2.2.5 Text Mining.....	9
2.2.6 Analisis Sentimen.....	9

2.2.7 Naive Bayes Classifier	9
2.2.8 Feature Extraction	10
2.2.9 Topic Modelling	10
2.2.10 Latent Dirichlet Allocation.....	11
2.2.11 Evaluasi Sistem	12
2.2.12 Topic Coherence	13
Bab 3 Metode Penelitian.....	14
3.1 Bahan Penelitian.....	14
3.2 Alat Penelitian.....	15
3.3 Jalan Penelitian.....	15
3.3.1 Web Data Extraction	16
3.3.2 Data Collection.....	16
3.3.3 Preprocessing.....	18
3.3.4 Data Clean	23
3.3.5 Topic Modelling Menggunakan LDA.....	23
3.3.6 Naïve Bayes Classifier	27
3.3.7 Pelabelan Manual	27
3.3.8 Data Training.....	29
3.3.9 Testing.....	38
3.3.10 Klasifikasi.....	39
Bab 4 Hasil Penelitian.....	40
4.1 Ringkasan Hasil Penelitian	40
4.2 Pembahasan Hasil Topik Coherence.....	41
4.3 Hasil Topik Modelling	42
4.4 Hasil Wordcloud Pertopik.....	43
4.5 Hasil Analisis Pertopik.....	49
4.6 Hasil Wordcloud keseluruhan Topik	50
4.7 Evaluasi Serta Hasil Klasifikasi	51
4.8 Hasil Klasifikasi Data Keseluruhan	54
4.9 Web Dashbord.....	63
Bab 5 Kesimpulan Dan Saran.....	70

5.1 kesimpulan	70
5.2 saran	70
Daftar Pustaka.....	71
Lampiran	73

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA

DAFTAR TABEL

Tabel 3.1 Data Collection.....	17
Tabel 3.2 Data Cleaning	19
Tabel 3.3 Daftar Kata Stopwords	20
Tabel 3.4 Tabel Kata Normalisasi	22
Tabel 3.5 Daftar Data Clean	23
Tabel 3.6 Pelabelan Manual	28
Tabel 3.7 Dokumen TF-IDF.....	29
Tabel 3.8 Perhitungan pada TF	30
Tabel 3.9 Perhitungan Pada IDF	31
Tabel 3.10 Perhitungan Pada TF-IDF	32
Tabel 4.1 Nilai Coherence	41
Tabel 4.2 Contoh Tweet Yang Terkait Dengan Topik 1	44
Tabel 4.3 Tweet Yang Terkait Topik 2	45
Tabel 4.4 Tweet Yang Terkait Topik 3	46
Tabel 4.5 Tweet Yang Terkait Topik 4	47
Tabel 4.6 Tweet Yang Terkait Topik 5	48
Tabel 4.7 Kata Dalam Pertopik	49
Tabel 4.8 Tweet Dari Keseluruhan Kata Yang Sering Muncul	51
Tabel 4.9 K-fold Accuracy Dan F-measure	51
Tabel 4.10 Confusion Matrix Pada Cross Validatiion	52
Tabel 4.11 Hasil Tata Training.....	53
Tabel 4.12 Confusion Matrix Pada Data Testing	53
Tabel 4.13 Akurasi Pada Data Testing	54
Tabel 4.14 Hasil Klasifikasi Keseluruhan	55
Tabel 4.15 Hasil Klasifikasi Keseluruhan	55
Tabel 4.16 Hasil Tweet Positif Pada Topik 1	57
Tabel 4.17 Hasil Tweet Negatif Pada Topik 1	57
Tabel 4.18 Hasil Tweet Positif Pada Topik 2.....	58
Tabel 4.19 Hasil Tweet Negatif Pada Topik 2	59

Tabel 4.20 Hasil Tweet Positif Pada Topik 3.....	59
Tabel 4.21 Hasil Tweet Negatif Pada Topik 3	60
Tabel 4.22 Hasil Tweet Positif Pada Topik 4.....	61
Tabel 4.23 Hasil Tweet Negatif Pada Topik 4	61
Tabel 4.24 Hasil Tweet Positif Pada Topik 5.....	62
Tabel 4.25 Hasil Tweet Negatif Pada Topik 5	62

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA

DAFTAR GAMBAR

Gambar 2.1 Visualisasi Topic Modelling.....	11
Gambar 2.2 Visualisasi Pemodelan Topic Dengan Metode LDA	12
Gambar 3.1 Alur Penelitian	16
Gambar 3.2 Data Pelabelan Manual	28
Gambar 3.3 Hasil Perhitungan Pertama	34
Gambar 3.4 Hasil Perhitungan Kedua	35
Gambar 3.5 Penggabungan Data Testing	38
Gambar 4.1 Grafik Diagram Pada Topic Coherence.....	41
Gambar 4.2 Topic Modelling Berdasarkan 5 Topik.....	43
Gambar 4.3 Wordcloud Topik Ke 1	44
Gambar 4.4 Wordcloud Topik Ke 2	45
Gambar 4.5 Wordcloud Topik Ke 3	46
Gambar 4.6 Wordcloud Topik Ke 4	47
Gambar 4.7 Wordcloud Topik Ke 5	48
Gambar 4.8 Wordcloud Keseluruhan Topik.....	50
Gambar 4.9 Hasil Diagram Dari Keseluruhan Data	56
Gambar 4.10 Tampilan Awal Dashbord.....	63
Gambar 4.11 Bagian Dataset	64
Gambar 4.12 Table Tokenizing	64
Gambar 4.13 Table Data Stopwords.....	65
Gambar 4.14 Tabel Data Stemming	65
Gambar 4.15 Tabel Data Clean	66
Gambar 4.16 Tabel Data Training	66
Gambar 4.17 Pemodelan Pickle.....	67
Gambar 4.18 Hasil Testing	67
Gambar 4.19 Tabel Result Topik	68
Gambar 4.20 Wordcloud Pertopik.....	68
Gambar 4.21 Hasil Keseluruhan Data	69

DAFTAR LAMPIRAN

Lampiran 1 Jadwal Penelitian	73
Lampiran 2 Kartu Bimbingan Tugas Akhir	74
Lampiran 3 Hasil Cek Plagiarisme.....	75

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA

DAFTAR SINGKATAN

LDA	Latent Dirichlet Allocation
NBC	Naïve Bayes Classifier
CSV	Comma Separated Values
TF-IDF	Term Frequency-Inverse Document Frequency
COVID-19	Kode <i>Corona Virus Disease</i> yang terjadi pada akhir tahun 2019

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA