PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN BOW PADA ANALISIS SENTIMEN TERKAIT WABAH CAMPAK MENGGUNAKAN LOGISTIC REGRESSION

SKRIPSI

Oleh:

Tri Ayuningtyas
220441100042

Dosen Pembimbing:

1. **Dr. Fitri Damayanti, S.Kom., M.Kom 197508272003122001**
2. **Dr. Imamah, S.Kom., M.Kom. 198507212014042001**

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS TRUNODJOYO MADURA
BANGKALAN

2026

HALAMAN JUDUL

PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN BOW PADA ANALISIS SENTIMEN TERKAIT WABAH CAMPAK MENGGUNAKAN LOGISTIC REGRESSION

SKRIPSI

Diajukan Untuk Memenuhi Persyaratan Penyelesaian Studi Strata Satu (S1)
Dan Memperoleh Gelar Sarjana Komputer (S.Kom) Di Program Studi
Sistem Informasi
Jurusan Teknik Informatika
Universitas Trunojoyo Madura

Tri Ayuningtyas
220441100042

PROGRAM STUDI SISTEM INFORMASI
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS TRUNODJOYO MADURA
2026

HALAMAN PENGESAHAN

PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN BOW PADA ANALISIS SENTIMEN TERKAIT WABAH CAMPAK MENGGUNAKAN LOGISTIC REGRESSION

Skripsi diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) di Universitas Trunodjoyo Madura

Oleh:

Nama : Tri Ayuningtyas

NIM : 220441100042

Tanggal Sidang: 30 September 2026
Disetujui oleh Tim Penguji Skripsi :

Dr. Fitri Damayanti, S.Kom., M.Kom.		(Pembimbing I)
NIP. 197508272003122001		
Dr. Imamah, S.Kom., M.Kom.		(Pembimbing II)
NIP. 198507212014042001		
Dr. Fitri Damayanti, S.Kom., M.Kom.		(Ketua Penguji)
NIP. 197508272003122001		
Dr. Yeni Kustiyahningsih, S.Kom., M.Kom.		(Penguji Utama)
NIP. 197709212008122002		
Doni Abdul Fatah, S.Kom., M.Kom.		(Anggota Penguji)
NIP. 198705202019031013		

Bangkalan, 07 Oktober 2026
Mengetahui,
Ketua Jurusan Teknik Informatika

Husni ., S.Kom., MT
NIP. 197707222003121006

HALAMAN PERNYATAAN ORISINALITAS

Saya yang bertanda tangan dibawah ini, menyatakan bahwa skripsi saya dengan judul :

PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN BOW PADA ANALISIS SENTIMEN TERKAIT WABAH CAMPAK MENGGUNAKAN LOGISTIC REGRESSION

1. Adalah asli, bukan merupakan karya pihak lain serta belum pernah diajukan untuk mendapatkan gelar akademik Sarjana baik di Universitas Trunojoyo Madura maupun di Perguruan Tinggi yang lain di Indonesia.
2. Tidak terdapat karya atau pendapat pihak lain yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis telah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila dikemudian hari terbukti skripsi ini sebagian atau seluruhnya merupakan hasil plagiasi atau terdapat hal-hal yang tidak sesuai dengan pernyataan diatas, maka saya sanggup menerima sanksi akademis yang berlaku dengan segala akibat hukumannya sesuai peraturan Universitas Trunojoyo Madura dan atau peraturan perundang-undangan yang berlaku.

Bangkalan, 07 Oktober 2026

Yang menyatakan,

Tri Ayuningtyas

220441100042

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, karunia, dan pertolongan-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul **"Perbandingan Ekstraksi Fitur TF-IDF dan BoW pada Analisis Sentimen Terkait Wabah Campak Menggunakan Logistic Regression"** dengan baik. Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Teknik.

Proses penyusunan Tugas Akhir ini bukanlah perjalanan yang selalu mudah. Banyak waktu, tenaga, pikiran, serta kesabaran yang penulis curahkan hingga akhirnya penelitian ini dapat diselesaikan. Dalam proses tersebut, penulis menyadari bahwa penyelesaian Tugas Akhir ini tidak terlepas dari doa, dukungan, bimbingan, bantuan, dan semangat dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati dan rasa hormat, penulis menyampaikan terima kasih kepada:

1. Allah SWT, yang telah memberikan kesehatan, kekuatan, kemudahan, kesabaran, serta pertolongan kepada penulis dalam setiap proses penyusunan Tugas Akhir ini. Tanpa kehendak dan izin-Nya, penulis tidak akan mampu melewati setiap proses hingga sampai pada tahap ini.
2. Alm. Sugaib, cinta pertama penulis yang telah lebih dahulu berpulang dan beristirahat dalam keabadian. Meskipun Bapak tidak sempat menemani perjalanan penulis hingga sampai pada tahap ini, kasih sayang, perjuangan, dan segala pesan yang pernah diberikan akan selalu hidup dalam ingatan dan hati penulis. Terima kasih karena telah menjadi bagian dari alasan penulis untuk terus berjuang. Salah satu harapan terakhir Bapak yang masih penulis ingat telah menjadi salah satu alasan penulis untuk tetap melangkah dan menyelesaikan apa yang telah dimulai. Semoga pencapaian ini dapat menjadi persembahan kecil untuk Bapak, dan semoga Bapak mendapatkan tempat terbaik di sisi Allah SWT.
3. Ibu Sulastri, perempuan hebat yang telah melahirkan, membesarkan, dan menjadi alasan penulis untuk terus bertahan dan melanjutkan perjalanan. Terima kasih atas doa yang tidak pernah putus, kasih sayang yang tidak pernah berkurang, serta kesabaran dan pengorbanan yang mungkin tidak

akan pernah mampu penulis balas sepenuhnya. Terima kasih karena selalu mendampingi, memberikan semangat ketika penulis merasa lelah, menguatkan ketika penulis merasa ingin menyerah, dan selalu percaya bahwa penulis mampu menyelesaikan perjalanan ini. Tugas Akhir ini merupakan salah satu bentuk kecil dari rasa terima kasih dan cinta penulis kepada Ibu.

4. Kedua kakak saya, yang selalu memberikan dukungan, perhatian, semangat, dan bantuan kepada penulis dalam berbagai keadaan. Terima kasih telah menjadi sosok yang dapat diandalkan, menemani penulis dalam proses perjalanan ini, serta memberikan kekuatan ketika penulis menghadapi kesulitan. Kehadiran dan dukungan kalian menjadi salah satu bagian penting yang membuat penulis mampu bertahan dan menyelesaikan Tugas Akhir ini. Semoga keberhasilan sederhana ini juga dapat menjadi kebahagiaan dan kebanggaan untuk kalian.
5. Dosen pembimbing, yang telah meluangkan waktu untuk memberikan bimbingan, arahan, masukan, koreksi, serta motivasi selama proses penyusunan Tugas Akhir. Terima kasih atas kesabaran dan perhatian yang diberikan sehingga penulis dapat menyelesaikan penelitian ini dengan lebih baik.
6. Dosen penguji, yang telah memberikan kritik, saran, dan masukan yang membangun dalam proses penyempurnaan Tugas Akhir ini. Setiap masukan yang diberikan menjadi bagian penting dalam memperbaiki dan meningkatkan kualitas penelitian yang dilakukan.
7. Teman-teman dan orang-orang terdekat penulis, yang telah menjadi tempat berbagi cerita, memberikan semangat, menemani proses yang panjang, serta memberikan dukungan ketika penulis menghadapi berbagai kesulitan selama menyelesaikan Tugas Akhir. Terima kasih telah menjadi bagian dari perjalanan ini.
8. Terakhir, penulis ingin mengucapkan terima kasih kepada satu sosok yang selama ini diam-diam berjuang tanpa henti, seorang perempuan sederhana dengan impian yang tinggi, yang terkadang sulit memahami isi pikiran dan hatinya sendiri. Terima kasih kepada penulis skripsi ini, yaitu diriku sendiri,

Tri Ayuningtyas. Seorang perempuan berusia 22 tahun yang dikenal keras kepala, namun terkadang masih memiliki sisi kekanak-kanakan seperti perempuan pada umumnya. Terima kasih telah hadir di dunia ini, telah bertahan sejauh ini, dan terus berjalan melewati segala tantangan yang semesta hadirkan. Terima kasih karena tetap berani menjadi dirimu sendiri. Aku bangga atas setiap langkah kecil yang telah kau ambil, atas segala pencapaian yang mungkin tidak selalu dirayakan oleh orang lain. Walaupun terkadang harapanmu tidak sesuai dengan apa yang semesta berikan, tetaplah belajar menerima dan mensyukuri apa pun yang kamu dapatkan. Jangan pernah lelah untuk terus berusaha dan berbahagialah di mana pun kamu berada. Rayakan setiap proses dalam dirimu dan jadikan dirimu tetap bersinar di mana pun tempatmu berpijak. Aku berdoa, semoga setiap langkah kakimu selalu dikuatkan, dikelilingi oleh orang-orang baik dan hebat, serta semoga satu per satu impian yang kamu perjuangkan dapat terwujud.

Akhir kata, penulis mengucapkan terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah hadir, membantu, mendoakan, dan memberikan dukungan selama proses penyelesaian Tugas Akhir ini. Setiap doa, bantuan, dan kebaikan yang diberikan tidak akan penulis lupakan. Semoga Allah SWT membalas segala kebaikan tersebut dengan keberkahan, kesehatan, dan kebahagiaan. Semoga Tugas Akhir ini dapat menjadi langkah awal bagi penulis untuk terus belajar, berkembang, dan memberikan manfaat bagi orang lain.

Bangkalan, 07 September 2026

Penyusun



Tri Ayuningtyas

220441100042

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PENGESAHAN ii
HALAMAN PERNYATAAN ORISINALITAS iii
LEMBAR PERSEMBAHAN iv
MOTTO v
ABSTRAK ix
ABSTRACT x
KATA PENGANTAR vi
DAFTAR ISI xi
DAFTAR GAMBAR xiv
DAFTAR TABEL xv
DAFTAR LAMPIRAN xvi
DAFTAR KODE PROGRAM xvii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Perumusan Masalah 4
1.2.1 Permasalahan 4
1.2.2 Metode Usulan 5
1.2.3 Pertanyaan Penelitian 5
1.3 Tujuan dan Manfaat 5
1.3.1 Tujuan Penelitian 5
1.3.2 Manfaat Penelitian 6
1.4 Batasan Masalah 6
1.5 Sistematika Penulisan 7
BAB II TINJAUAN PUSTAKA 9
2.1 Dasar Teori 9
2.1.1 Analisis Sentimen 9
2.1.2 Aplikasi X 10
2.1.3 Penyakit Campak 11
2.1.4 *Crawling* 12
2.1.5 *Preprocessing* 12
2.1.6 Pelabelan Data 13
2.1.7 Ekstrasi Fitur 14

2.1.8 *Logistic Regression* 16
2.1.9 *Confusion Matrix* 17
2.2 Penelitian Terkait 18
BAB III METODOLOGI PENELITIAN 27
3.1 Metodologi Penelitian 27
3.2 Pengumpulan Data 29
3.3 Text Preprocessing 30
3.3.1 Case Folding 31
3.3.2 Cleaning 33
3.3.3 Tokenizing 35
3.3.4 Filtering/Stopword Removal 37
3.3.5 Stemming 39
3.4 Pelabelan Data 41
3.5 Split Data 42
3.6 Pembobotan TF-IDF 44
3.7 Pembobotan BoW 48
3.8 Klasifikasi Logistic Regression 51
3.9 Evaluasi 53
3.10 Deployment 54
3.11 Skenario Uji Coba 54
BAB IV HASIL DAN PEMBAHASAN 56
4.1 Business Understanding 56
4.1.1 Identifikasi Masalah 56
4.1.2 Tujuan Bisnis 56
4.2 Data Understanding 57
4.2.1 Pengumpulan Data 57
4.2.2 Eksplorasi Data 58
4.3 Data Preparation 59
4.3.1 Text Preprocessing 59
4.3.2 Pelabelan Data 65
4.3.3 *Visualisasi* Data 66
4.3.4 Split Data 68
4.3.5 Ekstrasi Fitur TF-IDF 69
4.3.6 Ekstrasi Fitur BoW 70
4.4 Klasifikasi menggunakan *Logistic Regression* 71
4.5 Evaluasi Model 72

4.5.1 Skenario 1 TF-IDF dan BoW terhadap Logistic Regression tanpa menggunakan SMOTE 72
4.5.2 Skenario 2 TF-IDF dan BoW terhadap Logistic Regression menggunakan SMOTE 79
4.5.3 Analisis Hasil Perbandingan Keseluruhan....85
4.6 Deployment 87
4.6.1 Halaman Home 87
4.6.2 Analisis Batch (Upload Dataset)....88
4.6.3 Prediksi Sentimen/Analisis Teks 91
4.6.4 Perbandingan Model....92
BAB V PENUTUP....94
5.1 Kesimpulan....94
5.2 Saran....94
REFERENSI 95
Lampiran 1. Surat Pernyataan Validator....98
Lampiran 2. Sampel Data Validator....99
DAFTAR RIWAYAT HIDUP....106

DAFTAR GAMBAR

Gambar 3. 1 Alur Metodologi Penelitian 27
Gambar 3. 2 Alur Pengumpulan Data 29
Gambar 3. 3 Alur Text Preprocessing 31
Gambar 3. 4 Alur Case Folding 32
Gambar 3. 5 Alur Cleaning 34
Gambar 3. 6 Alur Tokenizing 36
Gambar 3. 7 Alur Stopword 38
Gambar 3. 8 Alur Stemming 40
Gambar 3. 9 Alur Split Data 43
Gambar 3. 10 Alur TF-IDF 44
Gambar 3. 11 Alur BoW 49

Gambar 4. 1 Wordcloud Sentimen Positif 67
Gambar 4. 2 Wordcloud Sentimen Negatif 67
Gambar 4. 3 Nilai Rata-Rata TF-IDF 70
Gambar 4. 4 Fitur BoW 71
Gambar 4. 5 Confusion Matrix TF-IDF Tanpa Smote 75
Gambar 4. 6 Confusion Matriks BoW Tanpa Smote 78
Gambar 4. 7 Hasil Confusion Matriks TF-IDF dengan Smote 81
Gambar 4. 8 Confusion Matriks BoW dengan Smote 84
Gambar 4. 9 Hasil Streamlit Home 88
Gambar 4. 10 Hasil Streamlit Upload CSV 88
Gambar 4. 11 Hasil Streamlit Proses Analisis 89
Gambar 4. 12 Hasil Streamlit Klasifikasi Logistic Regression 90
Gambar 4. 13 Hasil Streamlit Prediksi Akhir 90
Gambar 4. 14 Hasil Streamlit Analisis Teks 91
Gambar 4. 15 Hasil Streamlit Analisis Preprocessing 91
Gambar 4. 16 Hasil Streamit Analisis Hasil Prediksi 92
Gambar 4. 17 Hasil Streamlit Perbandingan Model 92
Gambar 4. 18 Hasil Streamlit Visualisasi Perbandingan Model 93
Gambar 4. 19 Hasil Streamlit Visualisasi CR dan Rangking Model 93

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait 20

Tabel 3. 1 Hasil Ulasan 30
Tabel 3. 2 Hasil Preprocesing Case Folding 33
Tabel 3. 3 Hasil Preprocesing Cleaning 35
Tabel 3. 4 Hasil Preprocesing Tokenizing 37
Tabel 3. 5 Hasil Preprocesing Stopword 39
Tabel 3. 6 Hasil Preprocesing Stemming 41
Tabel 3. 7 Hasil Pelabelan 42
Tabel 3. 8 Tabel Data 45
Tabel 3. 9 Perhitungan TF 45
Tabel 3. 10 Perhitungan IDF 46
Tabel 3. 11 Hasil Perhitungan TF-IDF 47
Tabel 3. 12 Perhitungan BoW 50
Tabel 3. 13 Hasil Representasi BoW 51
Tabel 3. 14 Perhitungan Klasifikasi menggunakan TF-IDF 52
Tabel 3. 15 Skenario Uji 54

Tabel 4. 1 Hasil Case Folding 60
Tabel 4. 2 Hasil Cleansing 61
Tabel 4. 3 Hasil Normalisasi 62
Tabel 4. 4 Hasil Tokenizing 63
Tabel 4. 5 Stopword 64
Tabel 4. 6 Hasil Stemming 65
Tabel 4. 7 Evaluasi TF-IDF Tanpa Smote 74
Tabel 4. 8 CR TFF-IDF Tanpa Smote 74
Tabel 4. 9 Hasil Evaluasi BoW tanpa Smote 76
Tabel 4. 10 Hasil CF BoW Tanpa Smote 77
Tabel 4. 11 Hasil Perbandingan TF-IDF dan BoW Tanpa Smote 78
Tabel 4. 12 Hasil Evaluasi TF-IDF dengan BoW 80
Tabel 4. 13 Hasil CR TF-IDF dengan Smote 80
Tabel 4. 14 Hasil Evaluasi Bow dengan Smote 83
Tabel 4. 15 Hasil CR BoW dengan Smote 83
Tabel 4. 16 Perbandingan TF-IDF dan BoW menggunakan Smote 84

DAFTAR LAMPIRAN

Lampiran 1. Surat Pernyataan Validator.....98
Lampiran 2. Sampel Data Validator.....105

DAFTAR KODE PROGRAM

Kode Program 4.1 Kode Program Crawling Data 57
Kode Program 4.2 Case Folding 59
Kode Program 4.3 Clean 60
Kode Program 4.4 Normalisasi 61
Kode Program 4.5 Tokenizing 62
Kode Program 4.6 Stopword 63
Kode Program 4.7 Stemming 64
Kode Program 4.8 Visualisasi Data 66
Kode Program 4.9 Split Data 68
Kode Program 4.10 Fitur TF - IDF 69
Kode Program 4.11 Fitur BoW 70
Kode Program 4.12 Klasifikasi Logistic Regression 72
Kode Program 4.13 TF-IDF Tanpa Smote 73
Kode Program 4.14 BoW tanpa Smote 76
Kode Program 4.15 TF-IDF dengan Smote 79
Kode Program 4.16 Bow dengan Smote 82