

TUGAS AKHIR

PENERAPAN METODE *MULTINOMIAL NAIVE BAYES* DENGAN OPTIMASI TF-IDF DAN SELEKSI FITUR *CHI-SQUARE* UNTUK ANALISIS SENTIMEN PADA ULASAN APLIKASI *GETCONTACT* DI *GOOGLE PLAY STORE*



**MOH. REZA ALFI ALI
2210651053**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2026**

TUGAS AKHIR

PENERAPAN METODE *MULTINOMIAL NAIVE BAYES* DENGAN OPTIMASI TF-IDF DAN SELEKSI FITUR *CHI-SQUARE* UNTUK ANALISIS SENTIMEN PADA ULASAN APLIKASI *GETCONTACT* DI *GOOGLE PLAY STORE*

Diajukan untuk memenuhi persyaratan guna meraih gelar Sarjana Komputer pada Program Studi Teknik Informatika Universitas Muhammadiyah Jember



MOH. REZA ALFI ALI
2210651053

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2026**

HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR

PENERAPAN METODE *MULTINOMIAL NAIVE BAYES* DENGAN OPTIMASI TF-IDF DAN SELEKSI FITUR *CHI-SQUARE* UNTUK ANALISIS SENTIMEN PADA ULASAN APLIKASI *GETCONTACT* DI *GOOGLE PLAY STORE*

Oleh:

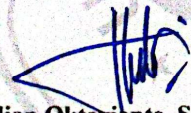
**MOH. REZA ALFI ALI
2210651053**

Telah disetujui bahwa Laporan Tugas Akhir ini diajukan dalam Sidang Tugas Akhir sebagai salah satu syarat kelulusan untuk memperoleh gelar Sarjana Komputer (S.Kom.) di Universitas Muhammadiyah Jember

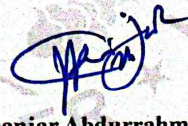
Disetujui oleh:

Pembimbing I

Pembimbing II



Hardian Oktavianto, S.Si., M.Kom
NIDN. 0722108105



Ginanjar Abdurrahman, S.Si., M.Pd.
NIDN. 0714078704

HALAMAN PENGESAHAN

**PENERAPAN METODE *MULTINOMIAL NAIVE*
BAYES DENGAN OPTIMASI TF-IDF DAN SELEKSI FITUR
CHI-SQUARE UNTUK ANALISIS SENTIMEN PADA ULASAN
APLIKASI *GETCONTACT* DI *GOOGLE PLAY STORE***

Oleh:

MOH. REZA ALFI ALI
2210651053

Telah mempertanggungjawabkan Laporan Tugas Akhirnya pada Sidang Tugas Akhir yang dilaksanakan pada tanggal 8 Mei 2026, sebagai salah satu syarat kelulusan untuk memperoleh gelar Sarjana Komputer (S.Kom.)

Disetujui oleh:

Penguji I

Nur Oodariyah Fitriyah, S.T., M.Kom
NIDN. 0727097501

Pembimbing I

Hardian Oktavianto, S.Si., M.Kom
NIDN. 0722108105

Penguji II

Luluk Handayani, S.Si., M.Si.
NIDN. 0725108003

Pembimbing II

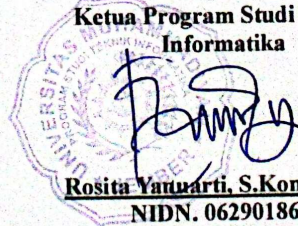
Ginanjar Abdurrahman, S.Si., M.Pd.
NIDN. 0714078704

Mengesahkan,
Dekan Fakultas Teknik



Prof. Dr. Ir. Muhtar, ST., MT., IPM.
NIDN. 0010067301

Mengetahui,
Ketua Program Studi Teknik
Informatika



Rosita Yauwanti, S.Kom., M. Cs
NIDN. 0629018601

HALAMAN PERNYATAAN ORISINALITAS

Saya yang bertanda tangan di bawah ini:

Nama : MOH. REZA ALFI ALI
NIM : 2210651053
Program Studi : Teknik Informatika
Judul Tugas Akhir : Penerapan Metode *Multinomial Naive Bayes* dengan Optimasi TF-IDF dan Seleksi Fitur *Chi-Square* untuk Analisis Sentimen pada Ulasan Aplikasi *Getcontact* di *Google Play Store*

dengan ini menyatakan bahwa karya ilmiah yang saya susun adalah murni hasil pemikiran, penelitian, dan penyusunan saya sendiri. Segala sumber informasi yang berasal atau dikutip dari karya orang lain, baik berupa publikasi, artikel, buku, maupun sumber lainnya, telah disebutkan dengan jelas di dalam daftar pustaka sesuai dengan kaidah penulisan ilmiah.

Saya menyadari bahwa dalam proses penyusunan Tugas Akhir ini, saya memanfaatkan bantuan teknologi kecerdasan buatan (AI) seperti *ChatGPT/Grammarly/Alat AI* lain hanya sebatas untuk:

1. Membantu dalam penyusunan tata bahasa, parafrasa, atau perbaikan redaksi.
2. Memberikan inspirasi ide awal atau kerangka berpikir, yang selanjutnya saya kembangkan secara mandiri.
3. Membantu dalam pengecekan konsistensi format, ejaan, dan tata tulis.

Saya menegaskan bahwa tidak ada bagian dari karya ilmiah ini yang seluruhnya dibuat oleh AI tanpa keterlibatan pemikiran kritis saya sendiri. Tanggung jawab penuh atas isi, keaslian, dan kebenaran karya ilmiah ini ada pada diri saya.

Apabila di kemudian hari terbukti bahwa pernyataan ini tidak benar, maka saya bersedia menerima segala konsekuensi sesuai dengan peraturan yang berlaku di Universitas Muhammadiyah Jember.

Jember, 8 Mei 2026

Yang membuat pernyataan,



MOH. REZA ALFI ALI
NIM 2210651053

KATA PENGANTAR

Segala puji bagi Allah SWT atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Penerapan Metode *Multinomial Naive Bayes* dengan Optimasi TF-IDF dan Seleksi Fitur *Chi-Square* untuk Analisis Sentimen pada Ulasan Aplikasi *Getcontact* di *Google Play Store*”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember.

Penyusunan skripsi ini tidak terlepas dari dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Rektor Universitas Muhammadiyah Jember, Dr. Hanafi, M.Pd.
2. Dekan Fakultas Teknik, Prof. Dr. Ir. Muhtar, S.T., M.T., IPM.
3. Ketua Program Studi Teknik Informatika, Ibu Rosita Yanuarti, S.Kom., M. Cs beserta seluruh dosen dan civitas akademika yang telah memberikan ilmu, bimbingan, serta dukungan selama masa perkuliahan di Universitas Muhammadiyah Jember.
4. Dosen Pembimbing, Bapak Hardian Oktavianto, S.Si., M.Kom dan Bapak Ginanjar Abdurrahman, S.Si., M.Pd. atas bimbingan, arahan, dan dukungan selama penyusunan skripsi ini.
5. Dosen Penguji, Ibu Nur Qodariyah Fitriyah, S.T., M.Kom dan Ibu Luluk Handayani, S.Si., M.Si. atas masukan dan saran yang diberikan.
6. Ibu tersayang serta almarhum Ayah tercinta yang selalu memberikan doa, kasih sayang, dan dukungan.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari kesempurnaan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan. Semoga skripsi ini bermanfaat, khususnya dalam bidang analisis sentimen dan klasifikasi teks.

Jember, 8 Mei 2026

Penulis
MOH. REZA ALFI ALI

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN ORISINALITAS	iv
ABSTRAK	v
ABSTRACT	vi
MOTTO	vii
HALAMAN PERSEMBAHAN	viii
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian	4
1.5 Batasan Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu.....	5
2.2 Kajian Teori.....	7
2.2.1 <i>Google Play Store</i>	7
2.2.2 Aplikasi <i>Getcontact</i>	8
2.2.3 <i>Text Mining</i>	8
2.2.4 Analisis Sentimen.....	9
2.2.5 <i>Web Scraping</i>	11
2.2.6 <i>Preprocessing</i>	11
2.2.7 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	13
2.2.8 <i>Chi-Square</i>	16
2.2.9 <i>Multinomial Naive Bayes (MNB)</i>	18
2.2.10 Evaluasi Kinerja Model.....	20

BAB III METODOLOGI PENELITIAN	23
3.1 Proses dan Langkah Penelitian.....	23
3.2 Pengumpulan Data	23
3.2.1 Sumber Data.....	24
3.2.2 Tahapan Pengambilan Data.....	24
3.2.3 Pelabelan Sentimen Otomatis	25
3.2.4 Komposisi Dataset	26
3.3 <i>Preprocessing</i>	26
3.3.1 <i>Case Folding</i>	27
3.3.2 <i>Normalization</i>	28
3.3.3 <i>Cleaning</i>	29
3.3.4 <i>Tokenizing</i>	30
3.3.5 <i>Stopword Removal</i>	31
3.3.6 <i>Stemming</i>	32
3.3.7 Hasil <i>Preprocessing</i>	32
3.1 Pembagian Data (<i>Data Training dan Data Testing</i>).....	33
3.2 Pembobotan Fitur Menggunakan TF-IDF	34
3.3 Seleksi Fitur Menggunakan <i>Chi-Square</i>	37
3.4 Klasifikasi <i>Multinomial Naive Bayes</i> (MNB).....	40
3.4.1 Prinsip Kerja <i>Multinomial Naive Bayes</i> (MNB).....	40
3.4.2 Proses Pelatihan Model.....	40
3.4.3 Perhitungan Manual <i>Multinomial Naive Bayes</i>	41
3.5 Evaluasi Kinerja Model.....	42
3.5.1 Akurasi	43
3.5.2 Presisi	43
3.5.3 <i>Recall</i>	43
3.5.4 <i>F1-Score</i>	44
3.5.5 <i>Confusion Matrix</i>	44
BAB IV HASIL DAN PEMBAHASAN.....	45
4.1 Deskripsi Dataset	45
4.1.1 Proses Pengumpulan Data.....	45
4.1.2 Struktur Dataset.....	47
4.1.3 <i>Scraping Data Unseen</i>	48
4.1.4 Analisis Dataset.....	49

4.2 <i>Preprocessing</i>	50
4.2.1 Tahapan <i>Preprocessing</i>	50
4.2.2 <i>Case Folding</i>	51
4.2.3 <i>Normalization</i>	53
4.2.4 <i>Cleaning</i>	55
4.2.5 <i>Tokenizing</i>	57
4.2.6 <i>Stopword Removal</i>	58
4.2.7 <i>Stemming</i>	60
4.2.8 Hasil <i>Preprocessing</i>	61
4.3 Pembagian Data (<i>Data Training</i> dan <i>Data Testing</i>).....	62
4.3.1 Metode Pembagian Data	62
4.3.2 Hasil Pembagian Data	63
4.4 Pembobotan Fitur TF-IDF.....	63
4.4.1 Representasi TF-IDF	64
4.4.2 Ekstraksi Nama Fitur.....	64
4.4.3 Transformasi TF-IDF <i>Data Training</i>	65
4.5 Seleksi Fitur <i>Chi-Square</i>	67
4.5.1 Perhitungan Skor <i>Chi-Square</i>	68
4.5.2 Eksperimen Variasi Jumlah Fitur	70
4.6 Pelatihan dan Pengujian <i>Multinomial Naive Bayes</i> (MNB).....	71
4.6.1 Proses Pelatihan dan Pengujian.....	71
4.6.2 Hasil Pengujian Variasi Jumlah Fitur.....	72
4.7 Evaluasi Model pada <i>Data Testing</i> (<i>Internal Validation</i>).....	73
4.7.1 Proses Evaluasi Model	73
4.7.2 Hasil Evaluasi Internal	74
4.8 Evaluasi Model pada <i>Data Unseen</i> (<i>External Validation</i>).....	76
4.8.1 Proses Evaluasi Model pada <i>Data Unseen</i>	76
4.8.2 Hasil Evaluasi Eksternal	78
4.8.3 Analisis Perbandingan Evaluasi Internal dan Eksternal.....	80
BAB V KESIMPULAN DAN SARAN	81
5.1 Kesimpulan	81
5.2 Saran.....	82
DAFTAR PUSTAKA.....	83
LAMPIRAN.....	91

DAFTAR GAMBAR

Gambar 3.1 Bagan Alir (<i>Flowchart</i>).....	23
Gambar 4.1 <i>Confusion Matrix</i> Hasil Klasifikasi.....	76
Gambar 4.2 Hasil <i>Confusion Matrix</i> pada <i>Data Unseen</i>	79



DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....	5
Tabel 2.2 Tabel Kontingensi 2×2 Hubungan <i>Term</i> dan Kelas.....	17
Tabel 2.3 Kategori Performa Klasifikasi.....	21
Tabel 2.4 Tabel <i>Confusion Matrix</i> Klasifikasi	22
Tabel 3.1 Pelabelan Sentimen Otomatis.....	26
Tabel 3.2 Contoh <i>Case Folding</i>	27
Tabel 3.3 Contoh <i>Normalization</i>	28
Tabel 3.4 Contoh <i>Cleaning</i>	29
Tabel 3.5 Contoh <i>Tokenizing</i>	30
Tabel 3.6 Contoh <i>Stopword Removal</i>	31
Tabel 3.7 Contoh <i>Stemming</i>	32
Tabel 3.8 Contoh <i>Text Preprocessing</i>	32
Tabel 3.9 <i>Term Frequency</i> (TF) pada Dokumen Contoh	35
Tabel 3.10 Hasil Perhitungan Bobot TF-IDF Dokumen Contoh	36
Tabel 3.11 Hasil Normalisasi TF-IDF (L2) pada Dokumen Contoh.....	37
Tabel 3.12 Tabel Kontingensi <i>Term</i> “bagus” terhadap Kelas Positif.....	38
Tabel 3.13 Contoh Seleksi Fitur Nilai <i>Chi-Square</i> Tertinggi.....	39
Tabel 3.14 Contoh <i>Confusion Matrix</i>	44
Tabel 4.1 Distribusi Dataset Utama	47
Tabel 4.2 Contoh Hasil dan Struktur Dataset.....	48
Tabel 4.3 Distribusi <i>Data Unseen</i>	49
Tabel 4.4 Contoh Hasil <i>Case Folding</i>	52
Tabel 4.5 Contoh Kamus <i>Normalization</i>	53
Tabel 4.6 Contoh Hasil <i>Normalization</i>	54
Tabel 4.7 Contoh Hasil <i>Cleaning</i>	56
Tabel 4.8 Contoh Hasil <i>Tokenizing</i>	57
Tabel 4.9 Contoh Hasil <i>Stopword Removal</i>	59
Tabel 4.10 Contoh Hasil <i>Stemming</i>	60
Tabel 4.11 Hasil <i>Preprocessing</i> dan Label Sentimen.....	61
Tabel 4.12 Hasil Pembagian Dataset.....	63
Tabel 4.13 100 Fitur Teratas Hasil Ekstraksi <i>Unigram</i> dan <i>Bigram</i>	65
Tabel 4.14 Hasil Representasi TF-IDF dari Beberapa Dokumen <i>Training</i>	66
Tabel 4.15 Hasil Fitur dengan Skor <i>Chi-Square</i> Tertinggi.....	69
Tabel 4.16 Variasi Jumlah Fitur Hasil Seleksi <i>Chi-Square</i>	71
Tabel 4.17 Hasil Pengujian Variasi Jumlah Fitur	72
Tabel 4.18 Hasil Perhitungan Akurasi Model	74
Tabel 4.19 Hasil Evaluasi Model Presisi, <i>Recall</i> , dan <i>F1-Score</i>	75
Tabel 4.20 Hasil Perhitungan Akurasi <i>Data Unseen</i>	78
Tabel 4.21 Hasil Presisi, <i>Recall</i> , dan <i>F1-Score Data Unseen</i>	78
Tabel 4.22 Perbandingan Hasil Evaluasi Model	80

DAFTAR LAMPIRAN

Lampiran 1. Kode *QR Google Drive* Data Penelitian dan Program.....91

