

SKRIPSI

**DETEKSI EMOSI PADA TEKS BERBAHASA INDONESIA DI
APLIKASI X MENGGUNAKAN ALGORITMA *RANDOM*
*FOREST***



INDAH RELISIAWATI

1512622084

PENDIDIKAN TEKNIK INFORMATIKA DAN KOMPUTER

FAKULTAS TEKNIK

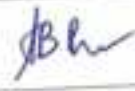
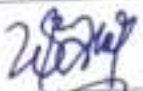
UNIVERSITAS NEGERI JAKARTA

2026

**LEMBAR PENGESAHAN TUGAS AKHIR
BENTUK: SKRIPSI**

Nama Mahasiswa : Indah Relisiawati
NIM : 1512622084
Program Studi/Fakultas : Pendidikan Teknik Informatika dan Komputer / Teknik
Judul Penelitian : Deteksi Emosi Pada Teks Berbahasa Indonesia Di Aplikasi X
Menggunakan Algoritma Random Forest

ini telah dipertahankan dihadapan dewan penguji dan dinyatakan **Lulus** (~~Tidak Lulus~~)*
Ujian Tugas Akhir oleh Tim Penguji pada Tanggal 11 May 2026
Tim Penguji,

Ketua Penguji	Bambang Prasetya Adhi, S.Pd., M.Kom. NIDN. 0025028303	
Anggota Penguji (Penguji Ahli 1)	Wiranti Kusuma Hapsari, S.Kom., M.Cs. NIDN. 0416079403	
Anggota Penguji (Penguji Ahli 2)	Neng Ayu Herawati, S.Kom., M.T. NU/PTK. 125277667723015	
Anggota Penguji (Pembimbing 1)	Dr. Widodo, S.Kom., M.Kom. NIDN. 0025037206	
Anggota Penguji (Pembimbing 2)	Murien Nugraheni, S.T., M.Cs NIDN. 0511108701	

Mengetahui,
Dekan Fakultas Teknik



Prof. Dr. Nengeng Siti Silfi Ambarwati, S.Si., Apt., M.Si.
NIDN. 0029027204

Koordinator Program Studi
Pendidikan Teknik Informatika dan
Komputer



Muchammad Ficky Daskarnaen, M.Sc.
NIDN. 0024097304

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Skripsi ini merupakan Karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Skripsi ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar Pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 22 April 2024

Yang membuat pernyataan,

A handwritten signature in black ink is written over a red official stamp. The stamp is rectangular and contains the text 'UNIVERSITAS NEGERI JAKARTA' at the top, 'METAL TEMPEL' in the center, and 'KASANK28334' at the bottom. There is also a small emblem of the university.

Indah Relisiawati

NIM. 1512622084



KEMENTERIAN PENDIDIKAN TINGGI, SAINS DAN TEKNOLOGI
UNIVERSITAS NEGERI JAKARTA
PERPUSTAKAAN DAN KEARSIPAN

Jalan Rawamangun Muka Jakarta 13220

Telepon/Faksimili: 021-4894221

Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Indah Relisiawati

NIM : 1512622084

Fakultas/Prodi : Fakultas Teknik / Pendidikan Teknik Informatika dan Komputer

Alamat email : iindahrls7@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan dan Kearsipan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☒ Skripsi ☐ Tesis ☐ Disertasi ☐ Lain-lain (.....)

yang berjudul:

DETEKSI EMOSI PADA TEKS BERBAHASA INDONESIA DI APLIKASI X
MENGUNAKAN ALGORITMA RANDOM FOREST

Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan dan Kearsipan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 28 Juli 2026

Penulis



IndahRelisiawati



KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas limpahan rahmat, karunia, dan hidayah Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan tepat waktu. Skripsi ini berjudul "Deteksi Emosi Pada Teks Berbahasa Indonesia Di Aplikasi X Menggunakan Algoritma *Random Forest*", yang disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Pendidikan pada Program Studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta.

Dalam proses penyusunan skripsi ini, penulis telah mendapatkan banyak bantuan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat, penulis menyampaikan ucapan terima kasih kepada:

1. Keluarga tercinta, Ayahku Mansur, S.Sos., sosok panutan dalam kehidupan penulis yang senantiasa mengajarkan arti keteguhan, kerja keras, dan tanggung jawab. Terima kasih atas setiap pengorbanan, doa, serta dukungan yang tak pernah terhenti, yang menjadi kekuatan bagi penulis untuk terus melangkah hingga sampai pada titik ini. Mamahku Empi Sulyati, sosok penuh kasih yang selalu menjadi pelita dalam hidup penulis, yang dengan tulus memberikan doa yang tak pernah putus, cinta yang tak terbatas, serta kehangatan yang selalu menjadi tempat pulang bagi penulis dalam setiap keadaan. Kepada kakak tercinta, Dian Eka Lestari, terima kasih atas perhatian, dukungan, serta semangat yang selalu diberikan, yang menjadi penguat bagi penulis di setiap proses yang tidak mudah, sehingga skripsi ini dapat terselesaikan.
2. Bapak Muhammad Ficky Duskarnaen, S.T., M.Sc., selaku Koordinator Program Studi Pendidikan Teknik Informatika dan Komputer, yang telah memberikan bimbingan, arahan, serta dukungan sehingga proses penyusunan skripsi ini dapat berjalan dengan lebih lancar.
3. Bapak Dr. Widodo, S.Kom., M.Kom., selaku dosen pembimbing I, dan Ibu Murien Nugraheni, S.T., M.Cs., selaku dosen pembimbing II, yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, serta berbagai masukan yang sangat berharga kepada penulis. Dengan penuh

kesabaran dan ketelitian, senantiasa membimbing penulis dalam setiap tahapan penyusunan skripsi ini, sehingga skripsi ini dapat terselesaikan dengan baik.

4. Seluruh teman-teman seperjuangan di Program Studi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta, serta teman-teman di organisasi AIESEC, yang telah menjadi bagian penting dalam perjalanan penulis. Terima kasih atas kebersamaan, dukungan, serta semangat yang saling diberikan, baik dalam berbagi ilmu, pengalaman, maupun dalam menghadapi berbagai tantangan selama proses perkuliahan, kegiatan organisasi, hingga penyusunan skripsi ini.
5. Teman-teman “Radio”, yang telah menjadi bagian berharga dalam perjalanan hidup penulis sejak masa Sekolah Menengah Atas (SMA) hingga saat ini. Terima kasih kepada Fayza, Radit, dan Iqbal, yang senantiasa memberikan dukungan, motivasi, serta semangat kepada penulis dalam berbagai kondisi, sehingga penulis dapat terus bertahan dan menyelesaikan skripsi ini.
6. Teman-teman sejak masa Sekolah Menengah Pertama (SMP), yaitu Syahla, Friska, Widya, dan Kirana, yang telah menjadi sahabat terbaik bagi penulis. Terima kasih atas kebersamaan yang telah terjalin sejak lama, serta dukungan, perhatian, dan semangat yang selalu diberikan hingga saat ini, sehingga penulis dapat melalui berbagai proses dengan lebih kuat.
7. Teman-teman selama masa magang di Nestlé dan Sequis, yang telah kebersamai penulis dalam menjalani proses magang bersamaan dengan kegiatan perkuliahan. Terima kasih atas kebersamaan, dukungan, serta berbagai masukan yang diberikan, sehingga membantu penulis dalam menjalani kedua proses tersebut hingga penyusunan skripsi ini dapat terselesaikan.
8. Serta seluruh pihak yang telah membantu dalam penyusunan skripsi ini, baik secara langsung maupun tidak langsung, yang tidak dapat penulis sebutkan satu per satu, sehingga skripsi ini dapat terselesaikan dengan baik.

ABSTRAK

Indah Relisiawati, Deteksi Emosi Pada Teks Berbahasa Indonesia Di Aplikasi X Menggunakan Algoritma Random Forest. Program Studi Pendidikan Teknik Informatika dan Komputer. Fakultas Teknik. Universitas Negeri Jakarta. 2026. Dosen Pembimbing: Dr. Widodo, M.Kom., dan Murien Nugraheni, S.T., M.Cs.

Peningkatan penggunaan media sosial X di Indonesia menghasilkan data teks emosional yang melimpah, namun penelitian deteksi emosi berbasis teks berbahasa Indonesia masih terbatas. Penelitian ini bertujuan untuk menganalisis kinerja kombinasi algoritma *Random Forest* dengan representasi fitur TF-IDF dalam mendeteksi emosi pada teks berbahasa Indonesia dari media sosial X dan membandingkannya dengan metode *baseline* (SVM dan KNN yang dikombinasikan dengan TF-IDF dan *Word2Vec*). Data yang digunakan sebanyak 7.687 *tweet* berbahasa Indonesia yang mencakup enam kategori emosi, yaitu Senang, Sedih, Marah, Takut, Jijik, dan Kaget. Data melalui tahapan *pre-processing* meliputi *cleansing*, *case folding*, *stopword removal*, tokenisasi, dan *stemming*, kemudian direpresentasikan menggunakan TF-IDF dan *Word2Vec*, lalu diklasifikasikan dengan enam model yang dievaluasi secara sistematis. Hasil penelitian menunjukkan bahwa model utama Random Forest + TF-IDF mencapai akurasi tertinggi sebesar 95,83%, dengan *precision weighted* 95,98%, *recall* 95,83%, dan *F1-Score* 95,86%, mengungguli semua model *baseline* termasuk SVM + TF-IDF (93,47%) sebagai *baseline* terbaik. Kesimpulan penelitian ini adalah kombinasi Random Forest dengan TF-IDF merupakan pendekatan yang efektif untuk deteksi emosi teks berbahasa Indonesia dari media sosial, dengan keunggulan pada mekanisme *ensemble learning* yang *robust* dan representasi fitur TF-IDF yang kaya.

Kata Kunci: Deteksi Emosi, Klasifikasi Teks, Natural Language Processing, Machine Learning, Random Forest, TF-IDF.

ABSTRACT

Indah Relisiawati, *Emotion Detection on Indonesian Language Text in Application X Using Random Forest Algorithm. Informatics and Computer Engineering Education Study Program. Faculty of Engineering. State University of Jakarta. 2026. Supervisor: Dr. Widodo, M.Kom., dan Murien Nugraheni, S.T., M.Cs.*

The growing use of social media X in Indonesia produces an abundance of emotional text data; however, research on text-based emotion detection in Indonesian remains limited. This study aims to analyze the performance of the Random Forest algorithm combined with Term Frequency-Inverse Document Frequency (TF-IDF) feature representation in detecting emotions from Indonesian-language text on social media X, and to compare it with baseline methods (SVM and KNN combined with TF-IDF and Word2Vec). The dataset consisted of 7,687 Indonesian-language tweets covering six emotion categories: Happy, Sad, Angry, Fear, Disgust, and Surprise. Data underwent pre-processing steps including cleansing, case folding, stopword removal, tokenization, and stemming, then represented using TF-IDF and Word2Vec, and classified using six models evaluated systematically. Results show that the main model, Random Forest + TF-IDF, achieved the highest accuracy of 95.83%, with a weighted precision of 95.98%, recall of 95.83%, and F1-Score of 95.86%, outperforming all baseline models including SVM + TF-IDF (93.47%) as the best baseline. This study concludes that the combination of Random Forest with TF-IDF is an effective approach for emotion detection in Indonesian-language text from social media, with advantages in its robust ensemble learning mechanism and the rich feature representation produced by TF-IDF.

Keywords: *Emotion Detection, Text Classification, Natural Language Processing, Machine Learning, Random Forest, TF-IDF.*

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
LEMBAR PERNYATAAN	ii
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS.....	iii
KATA PENGANTAR.....	v
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiv
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Identifikasi Masalah	6
1.3 Pembatasan Masalah.....	6
1.4 Perumusan Masalah	7
1.5 Tujuan Penelitian.....	7
1.6 Manfaat Penelitian	7
1.6.1 Manfaat Teoritis.....	8
1.6.2 Manfaat Praktis	8
BAB II KAJIAN PUSTAKA.....	10
2.1 Konsep dan Teori.....	10
2.1.1 Deteksi Emosi.....	10
2.1.2 <i>Random Forest</i>	11
2.1.3 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	15
2.1.4 <i>K-Nearest Neighbors (KNN)</i>	17
2.1.5 <i>Support Vector Machine (SVM)</i>	17
2.1.6 <i>Word2Vec</i>	18
2.1.7 <i>Pre-processing Teks</i>	21

2.1.8	<i>Confusion Matrix</i>	23
2.1.9	Aplikasi X.....	26
2.2	Penelitian yang Relevan	26
2.3	Kerangka Berpikir	38
BAB III METODOLOGI PENELITIAN		41
3.1	Tempat dan Waktu Penelitian	41
3.2	Alat dan Bahan Penelitian.....	41
3.2.1	Alat Penelitian.....	41
3.2.2	Bahan Penelitian	42
3.3	Langkah-langkah Penelitian	43
3.4	Prosedur Penelitian	46
3.4.1	Pengumpulan Data	46
3.4.2	<i>Text Preprocessing</i>	46
3.4.3	<i>Term Frequency–Inverse Document Frequency (TF-IDF)</i>	51
3.4.4	<i>Word2Vec</i>	53
3.4.5	Pembagian Dataset	55
3.4.6	Pelatihan Model	55
3.4.7	Evaluasi dan Perbandingan Model	56
3.5	Metode Evaluasi Data.....	56
3.5.1	Pendekatan Evaluasi.....	56
3.5.2	Metrik Evaluasi.....	56
3.5.3	Evaluasi Tambahan.....	57
3.5.4	Interpretasi dan Pemilihan Model Terbaik.....	57
BAB IV PEMBAHASAN		59
4.1	Hasil Penelitian	59
4.1.1.	<i>Text Pre-Processing</i>	59
4.1.2.	<i>Term Weighting TF-IDF</i>	63
4.1.3.	<i>Word2Vec</i>	66
4.1.4.	Pembagian Dataset	68
4.1.5.	Evaluasi model <i>Random Forest</i> +TF-IDF	70
4.1.6.	Evaluasi Model <i>Random Forest</i> + <i>Word2Vec</i>	74

4.1.7.	Evaluasi Model KNN	76
4.1.8.	Evaluasi Model SVM	79
4.2	Analisis Hasil Penelitian	81
4.2.1	Analisis Perbandingan Performa Model.....	81
4.2.2	Analisis <i>Confusion Matrix</i> dan <i>Classification Report</i>	83
4.3	Pembahasan	86
4.3.1	Interpretasi Hasil Penelitian.....	86
BAB V KESIMPULAN DAN SARAN		89
5.1	Kesimpulan	89
5.2	Saran	90
DAFTAR PUSTAKA		92
LAMPIRAN		95



DAFTAR GAMBAR

Gambar 2. 1 Kerangka Berpikir	40
Gambar 3. 1 Diagram Alir Penelitian	45
Gambar 3. 2 Proses Cleansing pada Python.....	47
Gambar 3. 3 Proses Case Folding pada Python.....	48
Gambar 3. 4 Proses Stopword Removal pada Python	49
Gambar 3. 5 Proses Tokenization pada Python	49
Gambar 3. 6 Proses Stemming pada Python	50
Gambar 3. 7 Proses TF-IDF	52
Gambar 3. 8 Proses Word2Vec	54
Gambar 3. 9 Hasil Proses Word2Vec	54
Gambar 4. 1 Distribusi Dataset per Kelas Emosi	59
Gambar 4. 2 Distribusi Panjang Teks Sebelum dan Sesudah Pre-Processing	60
Gambar 4. 3 Word Cloud Kata yang Sering Muncul Tiap Emosi	61
Gambar 4. 4 Kode Program TF-IDF	64
Gambar 4. 5 Top 20 TF-IDF Berdasarkan Skor Kumulatif	65
Gambar 4. 6 Kode Program Word2Vec.....	67
Gambar 4. 7 Visualisasi Word2Vec Embedding (t-SNE) - Baseline	68
Gambar 4. 8 Kode Program Train-Test Split.....	69
Gambar 4. 9 Distribusi Data Train dan Test per Kelas Emosi.....	70
Gambar 4. 10 Kode Program Evaluasi Model Random Forest + TF-IDF	71
Gambar 4. 11 Confussion Matrix Perbandingan Performa 6 Model	72
Gambar 4. 12 Feature Importance Top 20.....	73
Gambar 4. 13 Kode Program Random Forest + Word2Vec	74
Gambar 4. 14 Confusion Matric Random Forest + Word2Vec	75
Gambar 4. 15 Kode Program KNN	77
Gambar 4. 16 Confusion Matrix KNN TF-IDF Baseline	78
Gambar 4. 17 Confusion Matrix KNN Word2Vec Baseline	79
Gambar 4. 18 Kode Program SVM	80

Gambar 4. 19 Perbandingan 4 Metrik untuk Model Utama dan Baseline	82
Gambar 4. 20 Confusion Matrix Detail - Random Forest + TF-IDF (Model Utama)	83
Gambar 4. 21 Classification Report Heatmap - Random Forest + TF-IDF (Model Utama).....	85
Gambar 4. 22 Classification Report Heatmap - Random Forest + TF-IDF (Model Utama).....	86



DAFTAR TABEL

Tabel 2. 1 <i>Contoh Proses Text Cleansing</i>	22
Tabel 2. 2 <i>Contoh Proses Stopword Removal</i>	22
Tabel 2. 3 <i>Contoh Proses Case Folding</i>	22
Tabel 2. 4 <i>Contoh Proses Stemming</i>	23
Tabel 2. 5 <i>Contoh Proses Tokenisasi</i>	23
Tabel 2. 6 <i>Confusion Matrix</i>	24
Tabel 2. 7 <i>Penelitian Relevan</i>	32
Tabel 3. 1 <i>Spesifikasi Laptop Lenovo</i>	41
Tabel 3. 2 <i>Spesifikasi Perangkat Lunak</i>	42
Tabel 3. 3 <i>Hasil Proses Cleansing</i>	47
Tabel 3. 4 <i>Hasil Proses Case Folding</i>	48
Tabel 3. 5 <i>Hasil Proses Stopword Removal</i>	49
Tabel 3. 6 <i>Hasil Proses Tokenization</i>	50
Tabel 3. 7 <i>Hasil Proses Stemming</i>	51
Tabel 4. 1 <i>Statistik Panjang Teks sebelum/sesudah Pre-Processing</i>	60
Tabel 4. 2 <i>Ringkasan Akhir Teks Pre-Processing</i>	63
Tabel 4. 3 <i>Parameter TF-IDF</i>	63
Tabel 4. 4 <i>Metrik TF-IDF</i>	65
Tabel 4. 5 <i>Parameter Word2Vec (Baseline)</i>	66
Tabel 4. 6 <i>Statistik Word2Vec Model (Baseline)</i>	67
Tabel 4. 7 <i>Distribusi Train/Test Split</i>	68
Tabel 4. 8 <i>Parameter Random Forest</i>	70
Tabel 4. 9 <i>Hasil Random Forest + TF-IDF (Model Utama) pada Test Set</i>	71
Tabel 4. 10 <i>Classification Report Performa Emosi</i>	73
Tabel 4. 11 <i>Hasil Evaluasi Random Forest + Word2Vec</i>	74
Tabel 4. 12 <i>Classification Report Model Baseline</i>	75
Tabel 4. 13 <i>Parameter KNN</i>	76
Tabel 4. 14 <i>Hasil Evaluasi Model Baseline KNN</i>	77
Tabel 4. 15 <i>Parameter SVM</i>	79
Tabel 4. 16 <i>Hasil Evaluasi Baseline SVM</i>	80

Tabel 4. 17 Ringkasan Hasil Evaluasi Model Utama dan Baseline pada Test Set	81
Tabel 4. 18 Perbandingan Keunggulan Model Utama dan Baseline	82
Tabel 4. 19 Perbandingan Performa Berdasarkan Representasi Fitur	83
Tabel 4. 20 Analisis Pola Kesalahan Klasifikasi Model Utama	84
Tabel 4. 21 Ranking Emosi Berdasarkan FI-Score (Model Utama)	85



DAFTAR LAMPIRAN

Lampiran 1 Dataset X	95
Lampiran 2 Surat Tugas Dosen Pembimbing.....	96
Lampiran 3 Lembar Konsultasi Dosen Pembimbing I	97
Lampiran 4 Lembar Konsultasi Dosen Pembimbing II	99
Lampiran 5 Surat Pernyataan Dosen Pembimbing I.....	101
Lampiran 6 Surat Pernyataan Dosen Pembimbing II	102
Lampiran 7 Pra-Transkrip Akademi.....	103
Lampiran 8 Surat Pernyataan Menyelesaikan Mata Kuliah Wajib	106

