

**KLASIFIKASI KOSAKATA KASAR DAN TIDAK KASAR
PADA KOMENTAR *INSTAGRAM* MENGGUNAKAN
METODE *RANDOM FOREST* BERBASIS
*MACHINE LEARNING***

SKRIPSI



OLEH :

MOHAMMAD HOFAL FAOZAN

NIM : 2137016

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PASIR PENGARAIAN**

2026

PERSETUJUAN PEMBIMBING
KLASIFIKASI KOSAKATA KASAR DAN TIDAK KASAR PADA
KOMENTAR *INSTAGRAM* MENGGUNAKAN METODE *RANDOM*
FOREST* BERBASIS *MACHINE LEARNING

Disetujui Oleh :

Pembimbing I

Pembimbing II



Luth Fimawahib, M.Kom
NIDN. 1013068901



Satria Riki Mustafa, S.Pd, M.Si
NIDN. 1001039301

Diketahui Oleh :

Ketua Program Studi Teknik Informatika



Satria Riki Mustafa, S.Pd., M.Si
NIDN. 1001039301

PERSETUJUAN PENGUJI

Skripsi Ini Telah Diuji Oleh Tim Penguji Ujian Sarjana Komputer

Program Studi Teknik Informatika

Fakultas Ilmu Komputer

Universitas Pasir Pengaraian

Pada Tanggal 02 Oktober 2025

Tim Penguji:

1. Luth Fimawahib, M.Kom

NIDN. 1013068901

Ketua

()

2. Satria Riki Mustafa, S.Pd., M.si

NIDN. 1001039301

Sekretaris

()

3. Rivi Antoni, M.Pd

NIDN. 1003128103

Anggota

()

4. Asep Suprivanto, S.T., M.Kom

NIDN. 1003108903

Anggota

()

5. Erni Rouza, S.T., M.Kom

NIDN. 1009058707

Anggota

()

Mengetahui:

Dekan Fakultas Ilmu Komputer
Universitas Pasir Pengaraian


Luth Fimawahib, M.Kom
NIDN. 1013068901

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi yang berjudul “Klasifikasi Kosakata Kasar Dan Tidak Kasar Pada Komentar *Instagram* Menggunakan Metode *Random Forest* Berbasis *Machine Learning*”, benar hasil penelitian saya dengan arahan Dosen Pembimbing dan belum pernah diajukan dalam bentuk apapun untuk mendapatkan gelar Kesarjanaan. Dalam skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan dalam naskah dengan menyebutkan referensi yang dicantumkan dalam daftar pustaka. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena skripsi ini, serta lainnya sesuai norma yang berlaku di perguruan tinggi.

Pasir Pengaraian, 20 September 2025

Yang Membuat Pernyataan



MOHAMMAD HOFAL FAOZAN

NIM 2137016

KATA PENGANTAR

Assalamu 'alaikum wa rahmatullahi wa barokatuh

Alhamdulillah rabbil Alamin,

Segala puji syukur kehadiran Allah SWT yang senantiasa melimpahkan rahmat serta karunia-Nya, sehingga penulis mampu menyelesaikan Skripsi ini dengan baik. Shalawat serta salam kita terucapkan buat junjungan kita Rasulullah Muhammad SAW, karna jasa beliau yang telah membawa manusia dari zaman kebodohan hingga sampai ke zaman yang penuh dengan ilmu pengetahuan seperti sekarang ini.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana komputer. Banyak sekali pihak yang telah membantu penulis dalam penyusunan Skripsi ini, baik berupa bantuan materi maupun berupa motivasi dan dukungan kepada penulis. Semua itu tentu terlalu banyak bagi penulis untuk membalasnya, namun pada kesempatan ini penulis hanya dapat mengucapkan terima kasih kepada:

1. Bapak Dr. Hardianto, M.Pd, selaku Rektor Universitas Pasir Pengaraian.
2. Bapak Luth Fimawahib, M.Kom, selaku Dekan Fakultas Ilmu Komputer serta selaku Penasehat Akademik dan juga Pembimbing 1 yang telah membantu dan mengarahkan penulis dalam penyusunan skripsi ini.
3. Bapak Satria Riki Mustafa, S.Pd, M.Si, selaku Ketua Program Studi Teknik Informatika Fakultas Ilmu Komputer serta selaku Pembimbing 2 yang telah membimbing dan memberikan arahan dalam penyusunan skripsi ini.
4. Kepada almamater dan seluruh dosen Prodi Teknik Informatika Universitas Pasir Pengaraian yang telah memberikan ilmu dan pengalamannya serta mendidik penulis selama masa perkuliahan.

5. Kepada kedua orang tua tercinta, Bapak Subyani dan Ibu Suharti yang telah menjadi pahlawan dan panutan dihidup penulis. Terima kasih untuk hari-hari yang telah dihabiskan, terima kasih atas segala doa, kasih sayang, dan nasihat yang selalu mengiringi setiap langkah penulis serta dukungan moril maupun materil yang selalu diberikan. Sungguh Gelar Sarjana ini penulis persembahkan untuk Bapak dan Ibu.
6. Kepada kedua adik kandung tersayang, Mohammad Zaky dan Anissa Risky Novianti yang telah menjadi alasan tersendiri bagi penulis untuk menyelesaikan pendidikan perkuliahan ini. Tumbuhlah dewasa dengan baik.
7. Kepada teman-teman seperjuangan Prodi Teknik Informatika angkatan 2021 khususnya teman-teman *non regular* yang telah kebersamai dari awal perkuliahan sampai titik ini, serta memberi dukungan dan semangat yang menjadi sumber motivasi hingga tercapainya tahap ini.
8. Semua pihak yang tidak dapat disebutkan satu persatu, yang selalu setia menemani dan telah banyak memberikan pemikiran maupun saran demi kelancaran dan keberhasilan penyusunan skripsi ini.
9. Terakhir sebelum penulis akhiri, *“Beberapa anak memang terlahir beruntung ditengah keluarga yang berkecukupan. Sisanya lebih beruntung lagi karena diberi hati dan tulang yang kuat untuk berusaha sendiri.”* Kata-kata yang pernah penulis baca dan membuat penulis bisa bertahan hingga titik ini. Terima kasih banyak kepada Mohammad Hofal Faozan selaku penulis, sudah berusaha dan tidak memilih menyerah dalam kondisi apapun. Adapun kurang lebihmu, mari kita rayakan diri sendiri.

Dalam penyusunan skripsi ini, penulis menyadari masih jauh dari kesempurnaan, untuk itu penulis selalu terbuka atas kritik dan saran yang membangun guna penyempurnaan skripsi ini serta penulis berharap semoga skripsi ini dapat bermanfaat bagi pembaca dan peneliti selanjutnya.

Wassalamu'alaikum wa rahmatullahi wa barakatuh

Pasir Pengaraian, 20 September 2025

A handwritten signature in blue ink, appearing to read 'Hofal'.

Mohammad Hofal Faozan
NIM. 2137016

ABSTRACT

The rapid growth of Instagram comments demands fast, consistent moderation of abusive vocabulary. This study develops an automatic classifier to distinguish “abusive” vs “non-abusive” Indonesian comments harvested from the Instagram account of Gibran Rakabuming Raka, the Vice President of Indonesia. The end-to-end pipeline comprises public comment collection, text cleaning and normalization, TF-IDF (unigram–bigram) feature extraction, Random Forest training, evaluation, and an inference UI with confidence scores. Under an 80/20 hold-out split (1,324 validation comments), the best model with 340 trees attains 0.9607 accuracy and 0.9353 macro-F1; for the “abusive” class it yields 0.9651 precision and 0.8340 recall, consistent with a confusion matrix showing low false positives but some missed obfuscated insults. The system passes black-box testing per feature and is deployable as a moderation aid: reducing manual curation workload, preserving decision consistency, and producing an auditable trail (serialized model, logs, and metrics). Practically, the approach improves conversational hygiene while keeping the risk of flagging neutral comments low.

Keywords: Instagram, text classification, Random Forest

ABSTRAK

Pertumbuhan komentar di *Instagram* menuntut moderasi yang cepat dan konsisten atas kosakata kasar. Penelitian ini mengembangkan pengklasifikasi otomatis untuk membedakan komentar “kasar” dan “tidak kasar” berbahasa Indonesia yang diambil dari akun *Instagram* Gibran Rakabuming Raka, Wakil Presiden Republik Indonesia. *Pipeline end-to-end* mencakup pengambilan komentar publik, pembersihan dan normalisasi teks, ekstraksi fitur *TF-IDF*, pelatihan *Random Forest*, evaluasi, serta antarmuka inferensi dengan skor kepercayaan. Pada skema *hold-out* 80/20 (1.324 komentar validasi), model terbaik dengan 340 pohon mencapai akurasi 0,9607 dan *F1-macro* 0,9353, untuk kelas “kasar” diperoleh presisi 0,9651 dan *recall* 0,8340, selaras dengan *confusion matrix* yang menunjukkan *false positive* rendah namun sebagian hinaan tersamarkan masih terlewat. Sistem lulus uji *black-box* per fitur dan layak dioperasikan sebagai alat bantu moderasi dengan menurunkan beban kurasi manual, menjaga konsistensi keputusan, serta menyediakan jejak audit (model serialisasi, *log*, dan metrik). Secara praktis, pendekatan ini meningkatkan higienitas percakapan dengan risiko penandaan keliru pada komentar netral yang rendah.

Kata Kunci: Instagram, klasifikasi teks, *Random Forest*

DAFTAR ISI

PERSETUJUAN PEMBIMBING	Error! Bookmark not defined.
PERSETUJUAN PENGUJI	ii
LEMBAR PERNYATAAN	iii
KATA PENGANTAR	iv
ABSTRACT	vii
ABSTRAK	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Batasan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	6
BAB 2 LANDASAN TEORI	8
2.1. Pembelajaran Mesin (<i>Machine Learning</i>).....	8
2.2. Algoritma <i>Random Forest</i>	9
2.3. <i>Text Classification</i> dan <i>Natural Language Processing (NLP)</i>	10
2.4. Media Sosial dan Komentar di <i>Instagram</i>	12
2.5. Kosakata di Dunia Digital.....	13
2.6. Dataset Bahasa Indonesia di Media Sosial.....	14

2.7.	<i>Preprocessing</i> Data Teks Bahasa Indonesia	15
2.8.	<i>Feature Extraction</i> dan <i>Word Embedding</i>	16
2.9.	<i>Phyton</i>	17
2.10.	<i>Installoader</i>	17
2.11.	Sastrawi.....	18
2.12.	<i>Scikit-learn</i>	19
2.13.	<i>Pandas</i> dan <i>Numpy</i>	19
2.14.	<i>Matplotlib</i> dan <i>Seaborn</i>	20
2.15.	Penelitian Terkait	21
BAB 3 METODOLOGI PENELITIAN		32
3.1	Identifikasi Masalah	33
3.2	Perumusan Masalah	33
3.3	Pengumpulan Data	33
3.4	Analisa Sistem.....	34
3.4.1	Analisa Metode <i>Random Forest</i>	34
3.5	Perancangan Sistem	35
3.6	Implementasi Sistem	35
3.7	Pengujian Sistem.....	36
3.8	Kesimpulan dan Saran.....	37
BAB 4 ANALISA DAN PERANCANGAN		38
4.1	Analisa Sistem	38
4.1.1	Metode <i>Waterfall</i>	38
4.1.2	Analisa Sistem Lama	40
4.1.3	Analisa Sistem Baru	40
4.1.4	Analisa <i>Flowchart</i> Sistem.....	41

4.1.5	Analisa Kebutuhan Sistem.....	43
4.1.6	Analisa Masukan Sistem.....	43
4.1.7	Analisa Keluaran Sistem.....	44
4.2	Tahap Perancangan Model <i>Random Forest</i>	44
4.2.1	Formulasi Masalah & Metrik.....	45
4.2.2	<i>Preview Data</i> Mentah	47
4.2.3	<i>Preview Data</i> Setelah <i>Cleaning</i>	48
4.2.4	<i>Preview Data</i> Setelah <i>Labelling</i>	49
4.2.5	Representasi Fitur & Konstruksi Model	50
4.2.6	Hasil Klasifikasi Model	53
4.2.7	Contoh Prediksi	53
4.2.8	Analisis Hasil.....	55
4.3	Perancangan Sistem	55
4.3.1	<i>Unified Modelling Language (UML)</i>	55
4.3.1.1	<i>Use Case Diagram</i>	56
4.3.1.2	<i>Sequence Diagram</i>	57
4.4	Desain Sistem.....	62
4.4.1	Perancangan Antarmuka Aplikasi	62
4.4.1.1	Perancangan Antarmuka <i>Scrape Data</i>	62
4.4.1.2	Perancangan Antarmuka <i>Clean Data</i>	63
4.4.1.3	Perancangan Antarmuka <i>Manual Labelling</i>	63
4.4.1.4	Perancangan Antarmuka <i>Train Model</i>	64
4.4.1.5	Perancangan Antarmuka <i>Model Management</i>	65
4.4.1.6	Perancangan Antarmuka Test Prediksi	66
BAB 5 IMPLEMENTASI DAN PENGUJIAN		68

5.1	Implementasi Perangkat Lunak.....	68
5.1.1	Batasan Implementasi.....	68
5.1.2	Lingkungan Implementasi	69
5.1.3	Hasil Implementasi	70
5.1.3.1	Tampilan <i>Scrape Data</i>	71
5.1.3.2	Tampilan <i>Clean Data</i>	71
5.1.3.3	Tampilan Manual <i>Labelling</i>	72
5.1.3.4	Tampilan <i>Train Model</i>	73
5.1.3.5	Tampilan <i>Model Management</i>	75
5.1.3.6	Tampilan Test Prediksi (Hasil Tidak Kasar)	75
5.1.3.7	Tampilan Test Prediksi (Hasil Kasar).....	76
5.2	Pengujian Sistem.....	77
5.2.1	Pengujian Dengan Menggunakan <i>Blackbox</i>	77
5.2.2	Evaluasi Performa Model	81
5.3	Kesimpulan Pengujian.....	83
BAB 6 PENUTUP		85
6.1	Kesimpulan	85
6.2	Saran.....	86

DAFTAR GAMBAR

Gambar 3. 1 Tahapan Metodologi Penelitian	32
Gambar 4. 1 Metode <i>Waterfall</i>	39
Gambar 4. 2 <i>Flowchart</i> Sistem	42
Gambar 4. 4 Ilustrasi Aturan <i>Threshold</i> Pohon T_1	52
Gambar 4. 5 Ilustrasi Aturan <i>Threshold</i> Pohon T_2	52
Gambar 4. 6 <i>Use Case Diagram</i> Sistem Klasifikasi Komentar <i>Instagram</i>	56
Gambar 4. 7 <i>Sequence Diagram</i> <i>Scraping/Import Data</i>	57
Gambar 4. 8 <i>Sequence Diagram</i> <i>Data Cleaning</i>	58
Gambar 4. 9 <i>Sequence Diagram</i> <i>Data Labelling</i>	59
Gambar 4. 10 <i>Sequence Diagram</i> <i>Training Model</i>	60
Gambar 4. 11 <i>Sequence Diagram</i> Prediksi (Uji Komentar)	61
Gambar 4. 12 Perancangan Antarmuka <i>Scrape Data</i>	62
Gambar 4. 13 Perancangan Antarmuka <i>Clean Data</i>	63
Gambar 4. 14 Perancangan Antarmuka <i>Manual Labelling</i>	64
Gambar 4. 15 Perancangan Antarmuka <i>Train Model</i>	65
Gambar 4. 16 Perancangan Antarmuka <i>Model Management</i>	66
Gambar 4. 17 Perancangan Antarmuka Test Prediksi	67
Gambar 5. 1 Tampilan <i>Scrape Data</i>	71
Gambar 5. 2 Tampilan <i>Clean Data</i>	72
Gambar 5. 3 Tampilan <i>Manual Labelling</i>	73
Gambar 5. 4 Tampilan <i>Train Model</i>	74
Gambar 5. 5 Tampilan <i>Model Management</i>	75
Gambar 5. 6 Tampilan Test Prediksi (Hasil Tidak Kasar).....	76
Gambar 5. 7 Tampilan Test Prediksi (Hasil Kasar)	76
Gambar 5. 8 <i>Confusion Matrix</i>	81
Gambar 5. 9 Grafik Kinerja <i>Training Model</i>	82
Gambar 5. 10 Riwayat <i>Progress Training Model</i>	83

DAFTAR TABEL

Tabel 2. 1 Penelitian Sebelumnya.....	21
Tabel 4. 1 <i>Preview Data</i> Mentah	48
Tabel 4. 2 <i>Preview Data</i> Setelah <i>Cleaning</i>	49
Tabel 4. 3 <i>Preview Data</i> Setelah <i>Labelling</i>	50
Tabel 4. 4 Deskripsi Aktor Pada <i>Use Case</i>	56
Tabel 5. 1 Uji Fitur <i>Scrape Data</i>	78
Tabel 5. 2 Uji Fitur <i>Clean Data</i>	78
Tabel 5. 3 Uji Fitur <i>Manual Labelling</i>	79
Tabel 5. 4 Uji Fitur <i>Train Model</i>	79
Tabel 5. 5 Uji Fitur <i>Model Management</i>	80
Tabel 5. 6 Uji Fitur Test Prediksi – kasus tidak_kasar	80
Tabel 5. 7 Uji Fitur Test Prediksi – kasus kasar	80