

**Analisis Sentimen Komentar *YouTube* terkait Peluncuran
DANANTARA Menggunakan Metode *Naïve Bayes* dengan
Pendekatan *Random Undersampling* dan *Tomek Link*
untuk Menangani Ketidakseimbangan Data**

TUGAS AKHIR



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK ELEKTRO DAN INFORMATIKA
FAKULTAS TEKNIK DAN TEKNOLOGI KEMARITIMAN
UNIVERSITAS MARITIM RAJA ALI HAJI
TANJUNGPINANG
2026**

**Analisis Sentimen Komentar *YouTube* terkait Peluncuran
DANANTARA Menggunakan Metode *Naïve Bayes* dengan
Pendekatan *Random Undersampling* dan *Tomek Link*
untuk Menangani Ketidakseimbangan Data**

Tugas Akhir

Skema Skripsi

Tugas Akhir diajukan sebagai salah satu persyaratan memperoleh gelar Sarjana

Teknik pada Program Studi Teknik Informatika



Adityo Khorri Ramadhan

2101020045

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK ELEKTRO DAN INFORMATIKA
FAKULTAS TEKNIK DAN TEKNOLOGI KEMARITIMAN
UNIVERSITAS MARITIM RAJA ALI HAJI
TANJUNGPINANG
2026**

PERNYATAAN MENGENAI TUGAS AKHIR DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tugas akhir berjudul “Analisis Sentimen Komentar *YouTube* terkait Peluncuran DANANTARA Menggunakan Metode *Naïve Bayes* dengan Pendekatan *Random Undersampling* dan *Tomek Link* untuk Menangani Ketidakseimbangan Data” adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Apabila di kemudian hari terbukti pernyataan saya tidak benar dan melanggar peraturan yang berlaku dalam karya tulis dan hak intelektual, maka saya bersedia ijazah yang telah saya terima untuk ditarik kembali oleh Universitas Maritim Raja Ali Haji dan menerima sanksi lainnya sesuai dengan peraturan yang berlaku. Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Universitas Maritim Raja Ali Haji.

Tanjungpinang, Desember 2025



Adityo Khorir Ramadhan
NIM 2101020045



© Hak Cipta Milik Universitas Maritim Raja Ali Haji, Tahun 2021

Hak Cipta Dilindungi Undang- Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah; dan pengutipan tersebut tidak merugikan kepentingan Universitas Maritim Raja Ali Haji.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Universitas Maritim Raja Ali Haji.



HALAMAN PERSETUJUAN

Judul : Analisis Sentimen Komentar *YouTube* terkait
Peluncuran DANANTARA Menggunakan Metode
Naïve Bayes dengan Pendekatan *Random*
Undersampling dan *Tomek Link* untuk Menangani
Ketidakseimbangan Data
Nama : Adityo Khorī Ramadhan
NIM : 2101020045
Program Studi : Teknik Informatika
Tanggal Persetujuan : 11 Desember 2025

Disetujui oleh,
Komisi Pembimbing



Nurfalinda, S.T., M.Cs
(Ketua)



Fortia Magfira, M.Kom.
(Anggota)

HALAMAN PENGESAHAN

Judul : Analisis Sentimen Komentar *YouTube* terkait Peluncuran DANANTARA Menggunakan Metode *Naïve Bayes* dengan Pendekatan *Random Undersampling* dan *Tomek Link* untuk Menangani Ketidakseimbangan Data

Nama : Adityo Khorir Ramadhan

NIM : 2101020045

Telah berhasil dipertahankan di hadapan Komisi Penguji sebagai bagian persyaratan yang diperlukan dalam memperoleh gelar Sarjana Teknik pada Program Studi Teknik Elektro dan Informatika, Jurusan Teknik Informatika, Fakultas Teknik dan Teknologi Kemaritiman, Universitas Maritim Raja Ali Haji.

KOMISI PENGUJI

Ketua : Martaleli Bettiza, S.Si., M.Sc.

Anggota I : Muhamad Fadli, M.Kom.

Anggota II : Feri Irawan, M.Kom.

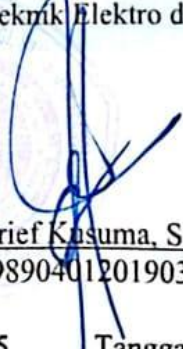
Anggota III : Nurfalinda, S.T., M.Cs

Anggota IV : Fortia Magfira, M.Kom.

(*Bert*)
(*Fadli*)
(*Feri*)
(*Nurfalinda*)
(*Fortia*)

Mengetahui,

Ketua Jurusan Teknik Elektro dan Informatika


Hollanda Arief Kusuma, S.IK., M.Si
NIP. 198904012019031016

Tanggal Ujian: 19 Desember 2025 Tanggal Lulus:

PRAKATA

Puji dan syukur dipanjatkan kepada Allah *subhanahu wa ta'ala* atas segala karunia-Nya sehingga laporan tugas akhir ini berhasil diselesaikan. Tema yang dipilih dalam tugas akhir yang akan dilaksanakan pada bulan Agustus 2025 ini ialah analisis sentimen, dengan judul **“Analisis Sentimen Komentar *YouTube* terkait Peluncuran DANANTARA Menggunakan Metode *Naïve Bayes* dengan Pendekatan *Random Undersampling* dan *Tomek Link* untuk Menangani Ketidakseimbangan Data**”.

Terima kasih penulis ucapkan kepada Ibu Nurfalinda, S.T., M.Cs dan Ibu Fortia Magfira, M.Kom. yang telah banyak memberi saran. Ungkapan terima kasih juga disampaikan kepada ayah, ibu, serta seluruh keluarga, atas segala doa dan kasih sayangnya.

Semoga laporan tugas akhir ini bermanfaat.

Tanjungpinang, 12 Desember 2025



Adityo Khorir Ramadhan

DAFTAR ISI

HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
ABSTRAK	vi
ABSTRACT.....	vii
PRAKATA.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN.....	1
A. Latar Belakang	1
B. Perumusan Masalah	4
C. Batasan Penelitian	5
D. Tujuan Penelitian	7
E. Manfaat Penelitian	7
BAB II KAJIAN LITERATUR	9
A. Tinjauan Pustaka	9
B. Landasan Teori.....	12
1. DANANTARA	12
2. Analisis Sentimen	13
3. <i>Text Mining</i>	14
4. Klasifikasi Sentimen dengan <i>Naïve Bayes</i>	16
5. <i>Class Imbalance</i>	17
6. <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	18
7. <i>Random Undersampling (RUS)</i>	19

8. Tomek Link	20
9. K-Fold Cross Validation	21
10. Confusion Matrix	21
11. Tesaurus (Kamus Sinonim dan Antonim Bahasa Indonesia).....	22
BAB III METODE PENELITIAN.....	23
A. Waktu dan Tempat Penelitian	23
B. Bahan dan Alat	23
C. Prosedur Pelaksanaan Penelitian.....	24
D. Perancangan Sistem	27
E. Analisis Data	48
BAB IV HASIL DAN PEMBAHASAN	78
A. Output Analysis Phase	78
1. Hasil Crawling Data YouTube	78
2. Hasil Preprocessing Data	78
3. Hasil Labelling dengan Inset Lexicon	79
4. Hasil Tingkat Akurasi dan Performa.....	80
B. Perbandingan Hasil dengan Penelitian Terdahulu	88
BAB V SIMPULAN DAN SARAN	91
A. Simpulan	91
B. Saran.....	93
DAFTAR PUSTAKA	94
LAMPIRAN.....	99

DAFTAR TABEL

Tabel 1. Tabel <i>labelling</i> data sampel	49
Tabel 2. Tabel Data Sampel setelah proses prapemrosesan data	51
Tabel 3. Tabel menghitung <i>Term Frequency (TF)</i>	53
Tabel 4. Tabel menghitung <i>Inverse Document Frequency (IDF)</i>	55
Tabel 5. Tabel menghitung TF-IDF	59
Tabel 6. Tabel data <i>Test Fold-1</i>	62
Tabel 7. Tabel data <i>train Fold-1</i>	62
Tabel 8. Tabel proses perhitungan <i>Tomek Link</i>	64
Tabel 9. Tabel hasil distribusi data	68
Tabel 10. <i>Dataset</i> baru hasil <i>Balancing</i>	70
Tabel 11. Tabel Distribusi kelas.....	71
Tabel 12. Tabel perhitungan <i>Prior Probabilities</i>	71
Tabel 13. Tabel TFIDF untuk contoh data <i>Test</i>	71
Tabel 14. Hasil klasifikasi data uji-1.....	73
Tabel 15. Hasil <i>Confusion Matrix</i>	74
Tabel 16. Perhitungan <i>Precision</i>	75
Tabel 17. Tabel perhitungan <i>Recall</i>	75
Tabel 18. Tabel perhitungan <i>F1-Score</i>	75
Tabel 19. Tabel perhitungan <i>Support</i>	76
Tabel 20. Tabel hasil kinerja lengkap model dalam <i>5-Fold</i>	76
Tabel 21. Tabel hasil akurasi dan performa <i>Imbalanced Data</i>	80
Tabel 22. Tabel hasil akurasi dan performa <i>Balanced Data</i>	84

DAFTAR GAMBAR

Gambar 1. Skema Alur Penelitian.....	25
Gambar 2. Alur Perancangan Sistem	28
Gambar 3. <i>Screenshot</i> video <i>YouTube</i> sebagai sumber data.....	29
Gambar 4. <i>Screenshot</i> video <i>YouTube</i> lainnya sebagai sumber data	29
Gambar 5. <i>Flowchart</i> proses <i>Labelling Data</i>	33
Gambar 6. Alur pembobotan kata (TFIDF)	34
Gambar 7. <i>Flowchart</i> alur <i>K-Fold Cross Validation</i>	36
Gambar 8. <i>Flowchart</i> alur <i>Tomek Link</i> dan <i>Random Undersampling</i>	37
Gambar 9. Diagram DFD <i>level 0</i>	39
Gambar 10. Diagram DFD <i>level 1</i>	41
Gambar 11. Gambar Perancangan ERD.....	42
Gambar 12. Tampilan <i>Login page</i>	43
Gambar 13. Tampilan Fitur <i>Crawling Dataset</i> Komentar <i>YouTube</i>	44
Gambar 14. Tampilan Fitur <i>Preprocessing Dataset</i>	44
Gambar 15. Tampilan Fitur <i>Labelling Dataset</i>	45
Gambar 16. Tampilan Fitur <i>Feature Extraction (TFIDF)</i>	46
Gambar 17. Tampilan Fitur <i>Training Model</i>	46
Gambar 18. Tampilan Fitur Prediksi Sentimen (<i>Input Teks</i>).....	47
Gambar 19. Hasil <i>Confusion Matrix</i> perhitungan data sampel.....	77
Gambar 20. Hasil <i>Crawling Data YouTube</i>	78
Gambar 21. Hasil <i>Preprocessing data</i>	79
Gambar 22. Diagram batang hasil <i>labelling data Preprocessing</i>	79
Gambar 23. Diagram batang hasil akurasi setiap <i>Fold imbalanced data</i> .	82
Gambar 24. Hasil <i>Confusion Matrix</i> data <i>Imbalanced</i>	83
Gambar 25. Diagram batang hasil akurasi setiap <i>Fold balanced data</i>	85
Gambar 26. Hasil <i>Confusion Matrix</i> data <i>Balanced</i>	86