



**ANALISIS SENTIMEN PENYAKIT PNEUMONIA DI MEDIA SOSIAL
MENGUNAKAN EKSTRAKSI *TERM FREQUENCY-RELEVANCE*
FREQUENCY (TF-RF) PADA KLASIFIKASI *BOOSTING***

PROPOSAL SKRIPSI

Untuk memenuhi persyaratan
melakukan penelitian dalam rangka penyusunan skripsi

Oleh

NUR HASANAH

NIM 2211016220011

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2026



**ANALISIS SENTIMEN PENYAKIT PNEUMONIA DI MEDIA SOSIAL
MENGUNAKAN EKSTRAKSI *TERM FREQUENCY-RELEVANCE*
FREQUENCY (TF-RF) PADA KLASIFIKASI *BOOSTING***

SKRIPSI

**Untuk memenuhi Persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer**

**Oleh
NUR HASANAH
NIM 2211016220011**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2026**

SKRIPSI

ANALISIS SENTIMEN PENYAKIT PNEUMONIA DI MEDIA SOSIAL MENGGUNAKAN EKSTRAKSI *TERM FREQUENCY-RELEVANCE FREQUENCY* (TF-RF) PADA KLASIFIKASI *BOOSTING*

Oleh:

NUR HASANAH

NIM. 2211016220011

Telah dipertahankan di depan Dosen Penguji pada tanggal 25 Juni 2026.

Susunan Dosen Penguji:

Pembimbing I



Rudy Herteno, S.Kom., M.Kom.

NIP. 198809252022031003

Dosen Penguji I



Muliadi, S.Kom., M.Cs.

NIP. 197804222010121002

Pembimbing II



Dodon Turianto Nugrahadi, S.Kom., M.Eng

NIP. 198001122009121002

Dosen Penguji II



Dwi Kartini, S.Kom., M.Kom.

NIP. 198704212012122003

Banjarbaru, 14 Juli 2026

Koordinator Program Studi Ilmu Komputer



Dwi Kartini, S.Kom., M.Kom.

NIP. 198704212012122003

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 01 Mei 2026



Nur Hasanah

NIM. 2211016220011

ABSTRAK

ANALISIS SENTIMEN PENYAKIT PNEUMONIA DI MEDIA SOSIAL MENGGUNAKAN EKSTRAKSI *TERM FREQUENCY-RELEVANCE FREQUENCY* (TF-RF) PADA KLASIFIKASI *BOOSTING* (Oleh: Nur Hasanah; Pembimbing: Rudy Herteno, S.Kom.,M.Kom. dan Dodon Turianto Nugrahadi, S.Kom., M.Eng.; 2026; 67 halaman)

Pneumonia adalah salah satu penyakit infeksi yang menyerang saluran nafas bawah yang merupakan penyebab kematian utama pada anak. Di era digital, media sosial seperti TikTok menjadi sarana bagi masyarakat untuk berbagi informasi, pengalaman, serta opini terkait penyakit pneumonia. Penelitian ini bertujuan untuk mengetahui dan membandingkan tingkat akurasi analisis sentimen penyakit pneumonia di media sosial TikTok menggunakan ekstraksi fitur *Term Frequency-Relevance Frequency* (TF-RF) pada klasifikasi *boosting* dengan algoritma AdaBoost, Gradient Boosting, dan *Extreme Gradient Boosting* (XGBoost). Data penelitian berupa 3.030 komentar TikTok yang diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Tahapan penelitian meliputi preprocessing, ekstraksi fitur menggunakan TF-RF, dan pembagian data menggunakan Holdout Validation dengan rasio 80:20. Selanjutnya, proses klasifikasi menggunakan tiga algoritma boosting, yaitu AdaBoost, Gradient Boosting, dan XGBoost. Hasil penelitian menunjukkan bahwa algoritma XGBoost memperoleh performa terbaik dibandingkan AdaBoost dan Gradient Boosting. Berdasarkan hasil evaluasi menggunakan metrik accuracy, precision, recall, F1-score, dan confusion matrix, XGBoost menghasilkan akurasi sebesar 88,12%. Sementara itu, Gradient Boosting memperoleh akurasi sebesar 84,82% dan AdaBoost memperoleh akurasi sebesar 67,33%. Hasil tersebut menunjukkan bahwa XGBoost memiliki kemampuan yang lebih baik dalam mengklasifikasikan sentimen komentar mengenai penyakit pneumonia di media sosial TikTok menggunakan ekstraksi fitur TF-RF.

Kata Kunci: Pneumonia, Media Sosial, Analisis Sentimen, AdaBoost, Gradient Boosting, XGBoost, TF-RF

ABSTRACT

SENTIMENT ANALYSIS OF PNEUMONIA ON SOCIAL MEDIA USING TERM FREQUENCY–RELEVANCE FREQUENCY (TF-RF) FEATURE EXTRACTION IN BOOSTING CLASSIFICATION (By: Nur Hasanah; Supervisors: Rudy Herteno, S.Kom., M.Kom. and Dodon Turianto Nugrahadi, S.Kom., M.Eng.; 2026; 67 pages)

Pneumonia is an infectious disease that affects the lower respiratory tract and is one of the leading causes of death among children. In the digital era, social media platforms such as TikTok have become a medium for people to share information, experiences, and opinions related to pneumonia. This study aims to determine and compare the accuracy of sentiment analysis on pneumonia-related comments on TikTok using the Term Frequency–Relevance Frequency (TF-RF) feature extraction method with boosting classification algorithms, namely AdaBoost, Gradient Boosting, and Extreme Gradient Boosting (XGBoost). The dataset consisted of 3,030 TikTok comments, which were classified into three sentiment categories: positive, neutral, and negative. The research process included data preprocessing, feature extraction using TF-RF, and data splitting using Holdout Validation with an 80:20 ratio. Subsequently, sentiment classification was performed using the three boosting algorithms. The results showed that XGBoost achieved the best performance compared to AdaBoost and Gradient Boosting. Based on the evaluation using accuracy, precision, recall, F1-score, and confusion matrix, XGBoost achieved an accuracy of 88.12%. Meanwhile, Gradient Boosting and AdaBoost achieved accuracies of 84.82% and 67.33%, respectively. These findings indicate that XGBoost is more effective in classifying sentiment in TikTok comments related to pneumonia using the TF-RF feature extraction method.

Keywords: *Pneumonia, Social Media, Sentiment Analysis, AdaBoost, Gradient Boosting, XGBoost, TF-RF*

PRAKATA

Puji Syukur penulis panjatkan ke Tuhan Yang Maha Esa karena atas berkat Rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul **“ANALISIS SENTIMEN PENYAKIT PNEUMONIA DI MEDIA SOSIAL MENGGUNAKAN EKSTRAKSI *TERM FREQUENCY-RELEVANCE FREQUENCY* (TF-RF) PADA KLASIFIKASI *BOOSTING*”** untuk memenuhi syarat dalam menyelesaikan Pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, Adapun yang dimaksud Adalah sebagai berikut:

1. Orang tua, adik, serta seluruh keluarga besar yang selalu memberikan dukungan, doa, motivasi, dan kasih sayang dalam proses penyelesaian skripsi ini.
2. Bapak Rudy Herteno, S.Kom., M.Kom. selaku dosen pembimbing utama yang telah memberikan bimbingan, membantu dan meluangkan waktu dalam proses penyelesaian skripsi ini.
3. Dodon Turianto Nugrahadi, S.Kom., M.Eng. selaku dosen pembimbing pendamping yang telah memberikan bimbingan, membantu, dan meluangkan waktu dalam proses penyelesaian skripsi ini.
4. Bapak Muliadi, S.Kom., M.Cs. dan Ibu Dwi Kartini, S.Kom., M.Kom. selaku dosen penguji dan dosen anggota penguji yang telah memberikan masukan-masukan untuk penyajian skripsi.
5. Bapak Rudy Herteno, S.Kom., M.Kom. selaku Dosen Pembimbing Akademik dan Ibu Dwi Kartini, S.Kom., M.Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
6. Seluruh dosen dan staf Program Studi Ilmu Komputer FMIPA ULM yang telah memberikan ilmu yang sangat bermanfaat selama penulis menempuh pendidikan.

7. Rudi yang senantiasa memberikan dukungan, semangat, dan bantuan selama masa perkuliahan serta dengan tulus meluangkan waktu, tenaga, dan perhatian untuk menemani setiap proses yang dilalui hingga penyusunan skripsi ini dapat diselesaikan.
8. Seluruh sahabat terdekat yaitu Helma, Lia, Dijah, Ismi, Arifah, dan Nisa yang telah menjadi bagian dari perjalanan perkuliahan, senantiasa memberikan dukungan, bantuan, motivasi, dan menemani hingga proses penyelesaian skripsi ini.
9. Seluruh teman dan rekan mahasiswa yang telah memberikan banyak dukungan, bantuan, dan doa selama masa studi hingga penyelesaian skripsi ini.
10. Semua pihak yang tidak bisa disebutkan satu persatu, yang turut berperan dan membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penyusunan skripsi ini masih terdapat berbagai kekurangan dan belum mencapai kesempurnaan yang diharapkan. Oleh karena itu, penulis mengharapkan bantuan berupa kritik dan saran yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 19 Juni 2026



Nur Hasanah

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK.....	iv
<i>ABSTRACT</i>	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Kajian terdahulu.....	6
2.2 Landasan Teori.....	11
2.2.1 Pneumonia.....	11
2.2.2 Media Sosial.....	12
2.2.3 Analisis Sentimen.....	12
2.2.4 Pemrosesan Data.....	13
2.2.5 <i>Term Frequency-Relevance Frequency (TF-RF)</i>	14
2.2.6 <i>Boosting</i>	14
2.2.7 <i>Adaptive Boosting (AdaBoost)</i>	15

2.2.8	<i>Gradient Boosting</i>	15
2.2.9	<i>Extreme Gradient Boosting (XGBoost)</i>	16
BAB III METODE PENELITIAN.....		20
3.1	Alat Penelitian	20
3.2	Bahan Penelitian	20
3.3	Prosedur Penelitian	21
BAB IV HASIL DAN PEMBAHASAN		24
4.1	Hasil Penelitian.....	24
4.1.1	Pengumpulan Data	24
4.1.2	Pelabelan Data.....	25
4.1.3	Preprocessing Data	25
4.1.4	Pembagian Data.....	30
4.1.5	Ekstraksi Fitur Menggunakan TF-RF	31
4.1.6	Klasifikasi.....	38
4.1.7	AdaBoost.....	39
4.1.9	Gradient Boosting	39
4.1.10	XGBoost.....	40
4.1.11	Evaluasi Model.....	44
4.2	Pembahasan	53
BAB V PENUTUP.....		61
5.1	Kesimpulan.....	61
5.2	Saran	61
DAFTAR PUSTAKA		62

DAFTAR TABEL

Tabel 1. Keaslian Penelitian.....	9
Tabel 2. Perancangan Penelitian	11
Tabel 3. Contoh Data Tiktok.....	24
Tabel 4. Pelabelan Data.....	25
Tabel 5. Data <i>Cleansing</i>	26
Tabel 6. Data <i>Case Folding</i>	27
Tabel 7. Data <i>Tokenization</i>	28
Tabel 8. Data <i>Stopword Removal</i>	29
Tabel 9. Data <i>Stemming</i>	30
Tabel 10. Hasil pembagian data (<i>Train-Test Split</i>).	31
Tabel 11. Contoh hasil ekstraksi TF-RF	32
Tabel 12. Tabel <i>Term</i> Komentar	32
Tabel 13. Tabel TF Komentar	32
Tabel 14. Tabel Nilai b dan c Kata “alhamdulillah”.	33
Tabel 15. Tabel Nilai b dan c Kata “sehat”.	33
Tabel 16. Tabel Nilai b dan c Kata “nak”.	33
Tabel 17. Tabel Perbandingan Hasil Nilai RF Kata “alhamdulillah”.	34
Tabel 18. Tabel Perbandingan Hasil Nilai RF Kata “sehat”.	35
Tabel 19. Tabel Perbandingan Hasil Nilai RF Kata “nak”.	36
Tabel 20. Tabel Hasil Akhir Nilai TF-RF.	37
Tabel 21. Data Hasil TF-RF	41
Tabel 22. Tabel Skor Prediksi Seluruh Kelas	41
Tabel 23. Tabel <i>Gradient</i>	42
Tabel 24. Tabel <i>Hessian</i>	42
Tabel 25. Tabel <i>Split</i> Fitur “alhamdulillah”	42
Tabel 26. Tabel Skor Prediksi Iterasi Pertama.....	43
Tabel 27. Hasil Akurasi Ketiga Algoritma	45
Tabel 28. Hasil Evaluasi Model AdaBoost.....	46

Tabel 29. Tabel <i>Confusion Matrix</i> AdaBoost	47
Tabel 30. Hasil Evaluasi Model Gradient Boosting.....	48
Tabel 31. Tabel <i>Confusion Matrix</i> Gradient Boosting.....	49
Tabel 32. Hasil Evaluasi Model XGBOOST	50
Tabel 33. Tabel <i>Confusion Matrix</i> XGBoost	52

DAFTAR GAMBAR

Gambar 1. Prosedur Penelitian.....	21
Gambar 2. Jumlah Distribusi Data	54
Gambar 3. Hasil Perbandingan Ketiga Akurasi	56
Gambar 4. <i>Confussion Matrix</i> Ketiga Algoritma <i>Boosting</i>	59