

**OPTIMIZATION OF INDOBERT FOR SENTIMENT ANALYSIS
OF FOMO ON SOCIAL MEDIA THROUGH
FINE-TUNING AND HYBRID LABELING**

LAPORAN NON-REGULER SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

NADHIF FAUZILADHIM

22.11.5035

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**OPTIMIZATION OF INDOBERT FOR SENTIMENT ANALYSIS
OF FOMO ON SOCIAL MEDIA THROUGH
FINE-TUNING AND HYBRID LABELING**

LAPORAN NON-REGULER SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

NADHIF FAUZIL ADHIM

22.11.5035

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

JALUR NON-REGULER

OPTIMIZATION OF INDOBERT FOR SENTIMENT ANALYSIS OF FOMO ON
SOCIAL MEDIA THROUGH FINE-TUNING AND HYBRID LABELING

yang disusun dan diajukan oleh

Nadhif Fauzil Adhim

22.11.5035

telah disetujui oleh Dosen Pembimbing
pada tanggal 27 Januari 2026

Dosen Pembimbing,



Nuri Cahyono, M.Kom
NIK. 190302278

HALAMAN PENGESAHAN

JALUR NON-REGULER

OPTIMIZATION OF INDOBERT FOR SENTIMENT ANALYSIS OF FOMO ON
SOCIAL MEDIA THROUGH FINE-TUNING AND HYBRID LABELING

yang disusun dan diajukan oleh

Nadhif Fauzil Adhim
22.11.5035

Telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arifivanto Hadinegoro, S.Kom., M.T.
NIK. 190302289

Subektiningsih, S.Kom., M.Kom.
NIK. 190302413

Nuri Cahyono, M.Kom.
NIK. 190302278

Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nadhif Fauzil Adhim

NIM : 22.11.5035

Menyatakan bahwa Laporan dengan judul berikut:

OPTIMIZATION OF INDOBERT FOR SENTIMENT ANALYSIS OF FOMO ON SOCIAL MEDIA THROUGH FINE-TUNING AND HYBRID LABELING

Dosen Pembimbing : Nuri Cahyono, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 27 Januari 2026

Yang Menyatakan,



Nadhif Fauzil Adhim

HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati dan rasa syukur yang mendalam, skripsi ini dipersembahkan kepada :

1. Kedua Orang Tua Tercinta Kepada Bapak dan Ibu saya, yang senantiasa memberikan kasih sayang tak terhingga, dukungan materil, serta lantunan doa yang tidak pernah putus di setiap sujudnya. Terima kasih atas segala pengorbanan dan kesabaran dalam membesarkan serta mendidik saya hingga dapat menyelesaikan pendidikan ini.
2. Dosen Pembimbing Kepada Bapak Nuri Cahyono, M.Kom, selaku dosen pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan arahan, bimbingan, serta ilmu yang sangat berharga selama proses penyusunan penelitian ini hingga selesai.
3. Sahabat Seperjuangan Kepada sahabat-sahabat terdekat yang selalu ada untuk mendengarkan keluh kesah, memberikan semangat, serta menjadi tempat bertukar pikiran di saat sulit maupun senang selama masa perkuliahan dan penyelesaian laporan ini.

KATA PENGANTAR

Segala puji dan syukur penulis haturkan ke hadirat Allah SWT. Berkat limpahan rahmat, hidayah, serta karunia-Nya, penulis dimampukan untuk merampungkan penyusunan Laporan Non-Reguler ini dengan judul "Optimization of IndoBERT for Sentiment Analysis of FOMO on Social Media Through Fine-Tuning and Hybrid Labeling". Laporan ini disusun sebagai salah satu syarat guna mencapai gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

Penulis menyadari bahwa selesainya laporan ini tidak lepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, dengan kerendahan hati, penulis ingin menghaturkan terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta, yang telah memberikan kesempatan dan fasilitas bagi penulis untuk menimba ilmu di universitas ini.
2. Bapak Nuri Cahyono, M.Kom, selaku Dosen Pembimbing yang senantiasa meluangkan waktu untuk memberikan arahan, saran yang membangun, serta bimbingan penuh kesabaran hingga laporan ini dapat terselesaikan dengan baik.
3. Kedua Orang Tua tercinta, yang senantiasa mengalirkan doa di setiap sujudnya, memberikan kasih sayang tanpa batas, serta dukungan moril maupun materiil yang menjadi motivasi terbesar bagi penulis.
4. Rekan-rekan seperjuangan serta seluruh pihak yang tidak dapat penulis sebutkan satu per satu, yang telah menjadi teman diskusi, tempat berbagi semangat selama masa studi hingga penyusunan laporan ini.

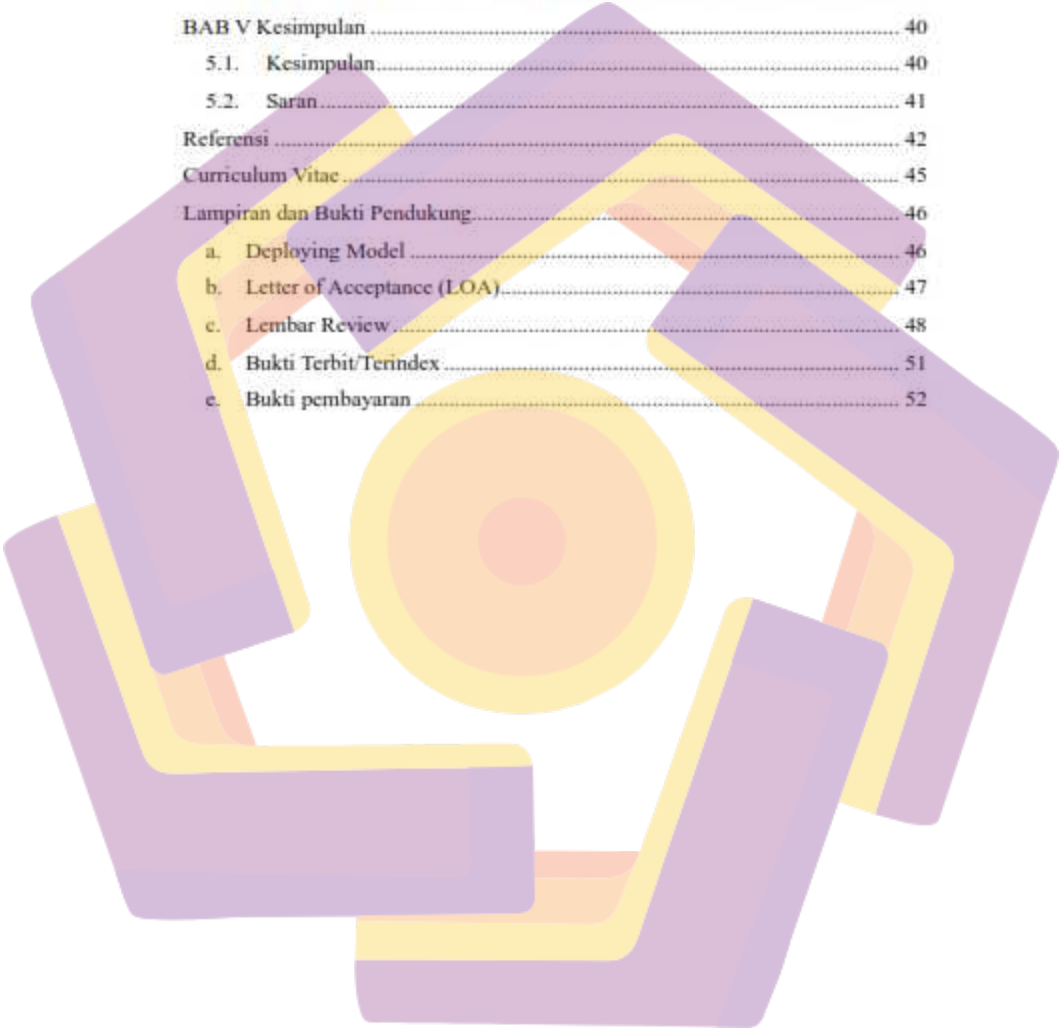
Yogyakarta, 27 Januari 2026

Nadhif Fauzil Adhim

DAFTAR ISI

Halaman Persetujuan.....	ii
Halaman Pengesahan.....	iii
Halaman Pernyataan Keaslian Karya.....	iv
Halaman Persembahan.....	v
Kata Pengantar.....	vi
Daftar Isi.....	vii
DAFTAR TABEL.....	x
Daftar Gambar.....	xi
Daftar Lampiran.....	xii
Daftar Lambang dan Singkatan.....	xiii
Daftar Istilah.....	xiv
Intisari.....	xvii
<i>Abstract</i>	xviii
Bab I Pendahuluan.....	1
1.1. Gambaran Umum.....	1
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah.....	3
1.4. Tujuan.....	3
Bab II Tinjauan Pustaka.....	5
2.1. Studi Literatur.....	5
2.2. Landasan Teori.....	7
2.2.1 Media Sosial dan Fenomena FOMO.....	7
2.2.2 Analisis Sentimen.....	7
2.2.3 Natural Language Processing.....	8
2.2.4 Arsitektur Transformer.....	8
2.2.5 Hybrid Labeling.....	9
2.2.6 Ketidakseimbangan Kelas dan Class Weighting.....	9
2.2.7 Support Vector Machine (SVM).....	9
2.2.8 Naïve Bayes.....	10
2.2.9 Bidirectional Long Short-Term Memory.....	11

2.2.10 BERT dan IndoBERT	12
2.2.11 Term Frequency-Inverse Document Frequency (TF-IDF)	13
2.2.12 Grid Search	14
2.2.13 Bayesian Optimization	14
2.2.14 Confusion Matrix	14
BAB III Metode Penelitian	17
3.1. Metode	17
3.1.1 Pengumpulan Data	18
3.1.2 Preprocessing Data	18
3.1.3 Hybrid Labeling	19
3.1.4 Splitting Data	20
3.1.5 Class Weight	20
3.1.6 Implementasi Support Vector Machine (SVM)	21
3.1.7 Implementasi Naïve Bayes	21
3.1.8 Implementasi Bidirectional Long Short-Term Memory (BiLSTM)	21
3.1.10 Grid Search	22
3.1.11 Bayesian Optimalization	22
3.1.12 Evaluasi	22
BAB IV Pembahasan	23
4.1. Pengumpulan Data	23
4.2. Preprocessing Data	24
4.3. Hybrid Labeling	26
4.3.1 Pelabelan Otomatis	26
4.3.2 Ambang Batas dan Validasi Hibrida	26
4.3.3 Anotasi Manual	28
4.3.4 Intra-Annotator Consistency Test	29
4.3.5 Datasets Final	30
4.4. Split Data	31
4.5. Class Weighting	32
4.6. Hyperparameter Tuning	32
4.6.1 GridSearchCV untuk Model Statistik	32
4.6.2 Optuna untuk Model Deep Learning	33



4.6.3 Hasil Pencarian Hyperparameter Tuning	33
4.7. Evaluasi Model dan Analisis Hasil	34
BAB V Kesimpulan	40
5.1. Kesimpulan	40
5.2. Saran	41
Referensi	42
Curriculum Vitae	45
Lampiran dan Bukti Pendukung	46
a. Deploying Model	46
b. Letter of Acceptance (LOA)	47
c. Lembar Review	48
d. Bukti Terbit/Terindex	51
e. Bukti pembayaran	52

DAFTAR TABEL

Tabel 4. 1 Distribusi Frekuensi Kata.....	23
Tabel 4. 2 Deskripsi Fitur Dataset.....	24
Tabel 4. 3 Tahapan Preprocessing Data	25
Tabel 4. 4 Aturan Pelabelan Hibrida	27
Tabel 4. 5 Pedoman Standar Anotasi Tabel.....	28
Tabel 4. 6 Distribusi Kelas Pelatihan	31
Tabel 4. 7 Hasil Perhitungan Bobot Kelas	32
Tabel 4. 8 Ruang Pencarian Hiperparameter (GridSearchCV)	33
Tabel 4. 9 Ruang Pencarian Hiperparameter (Optuna)	33
Tabel 4. 10 Hasil Optimasi Hyperparameter.....	34
Tabel 4. 11 Hasil Evaluasi pada Datasets Awal.....	35
Tabel 4. 12 Hasil evaluasi pada Datasets Final	35

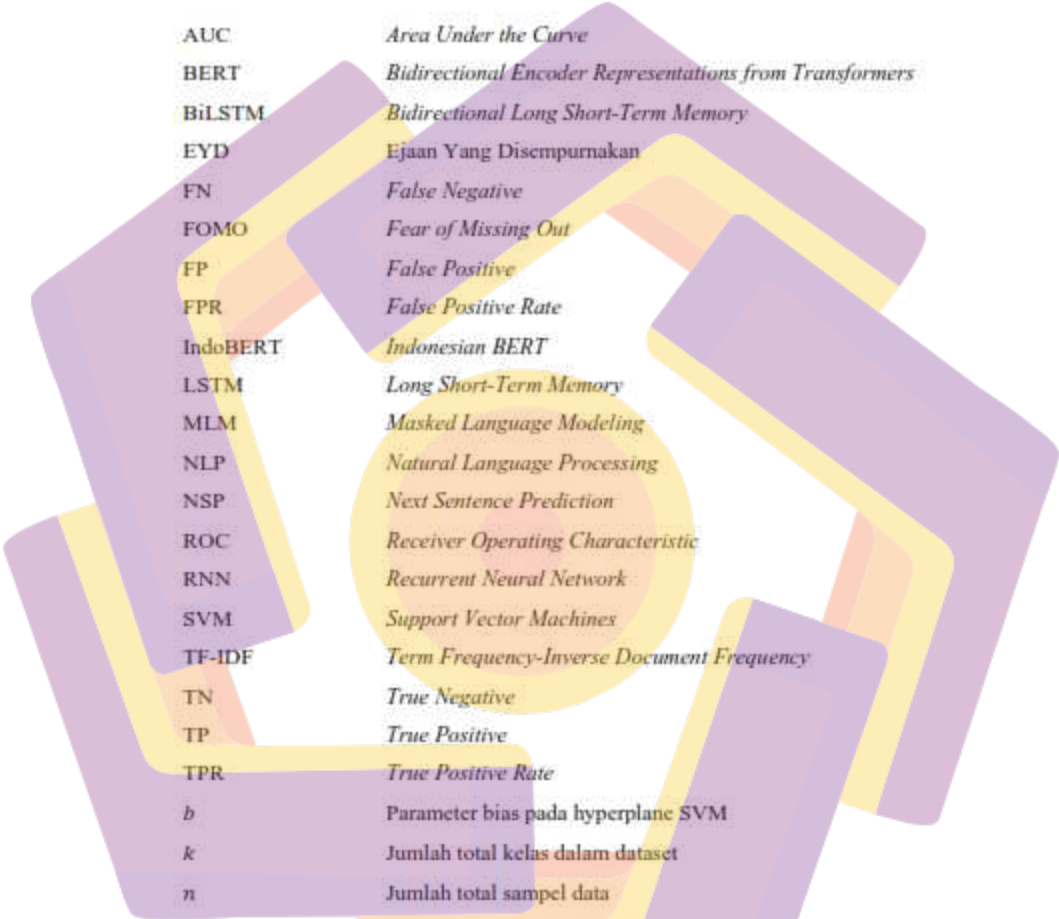
DAFTAR GAMBAR

Gambar 2. 1 Arsitektur BERT	12
Gambar 3. 1 Alur Penelitian	17
Gambar 4. 1 Word cloud pada data sampling	24
Gambar 4. 2 Distribusi Label Sebelum Hybrid Labeling	26
Gambar 4. 3 Distribusi Confidence Score dari Model	27
Gambar 4. 4 Distribusi Label pada Data Ambigu	28
Gambar 4. 5 Distribusi Sentimen Hasil Anotasi Manual	29
Gambar 4. 6 Confusion Matrix Intra-Anotator Consistency	30
Gambar 4. 7 Distribusi Final	31
Gambar 4. 8 Confusion Matrix Baseline	37
Gambar 4. 9 Confusion Matrix Optimized	37
Gambar 4. 10 Perbandingan Kurva ROC pada Datasets Final	38

DAFTAR LAMPIRAN

Lampiran 1 Deploy Model.....	46
Lampiran 2 Letter Of Acceptance (LOA).....	47
Lampiran 3 Lembar Review.....	50
Lampiran 4 Bukti Terbit.....	51
Lampiran 5 Lembar Review.....	52

DAFTAR LAMBANG DAN SINGKATAN



AUC	<i>Area Under the Curve</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
BiLSTM	<i>Bidirectional Long Short-Term Memory</i>
EYD	<i>Ejaan Yang Disempurnakan</i>
FN	<i>False Negative</i>
FOMO	<i>Fear of Missing Out</i>
FP	<i>False Positive</i>
FPR	<i>False Positive Rate</i>
IndoBERT	<i>Indonesian BERT</i>
LSTM	<i>Long Short-Term Memory</i>
MLM	<i>Masked Language Modeling</i>
NLP	<i>Natural Language Processing</i>
NSP	<i>Next Sentence Prediction</i>
ROC	<i>Receiver Operating Characteristic</i>
RNN	<i>Recurrent Neural Network</i>
SVM	<i>Support Vector Machines</i>
TF-IDF	<i>Term Frequency-Inverse Document Frequency</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
TPR	<i>True Positive Rate</i>
b	Parameter bias pada hyperplane SVM
k	Jumlah total kelas dalam dataset
n	Jumlah total sampel data
n_j	Jumlah sampel pada kelas $kc - j$
$P(c d)$	Probabilitas posterior dokumen d termasuk ke kelas c
$P(c)$	Probabilitas awal (<i>prior probability</i>) kelas c
W_j	Bobot kelas (<i>class weight</i>) untuk kelas $kc-j$
w	Vektor normal pada hyperplane SVM

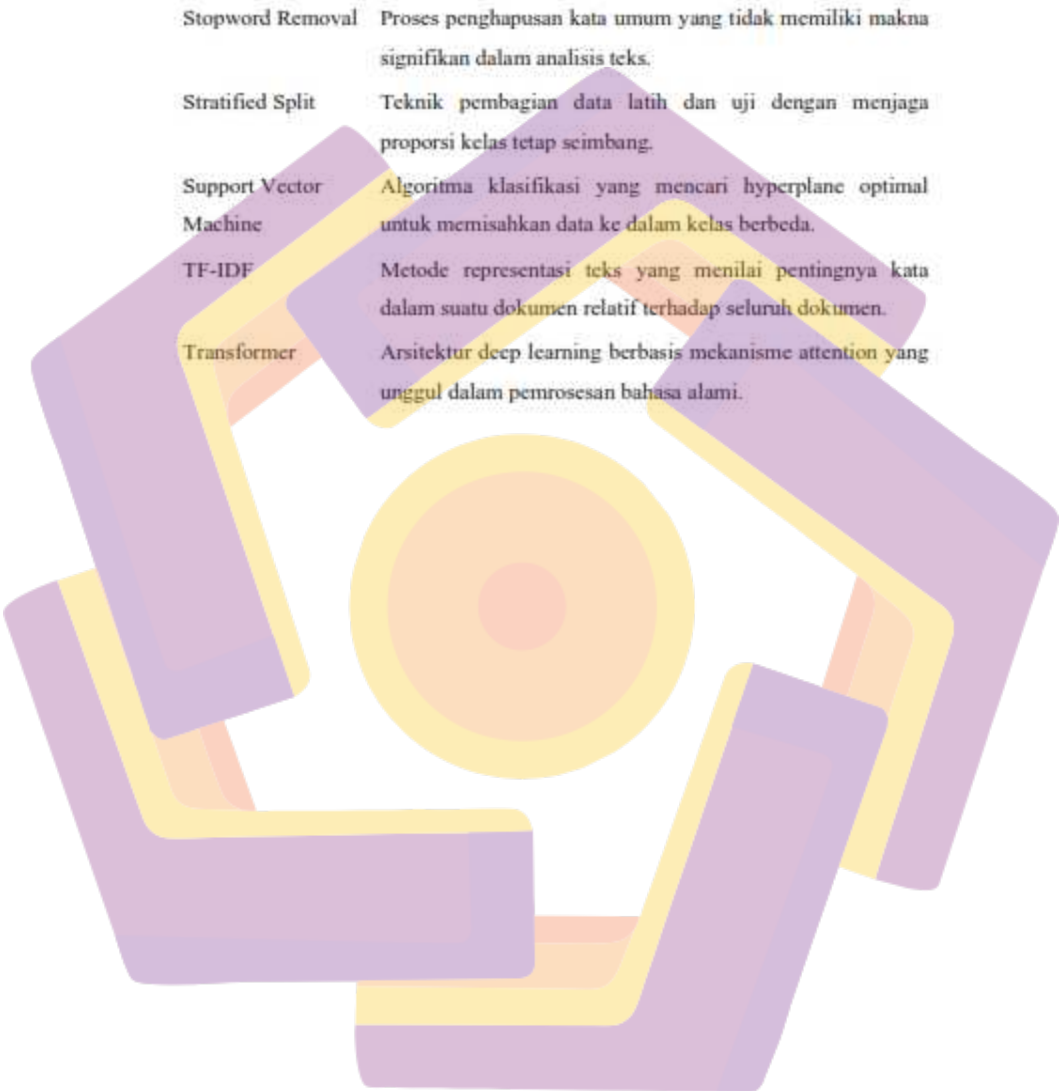
DAFTAR ISTILAH



Accuracy	Ukuran kinerja model yang menunjukkan proporsi prediksi benar terhadap seluruh data uji.
Ambiguous Data	Data teks yang memiliki konflik sentimen antara hasil prediksi model dan validasi leksikon sehingga memerlukan anotasi manual.
Analisis Sentimen	Proses mengidentifikasi dan mengklasifikasikan opini atau emosi dalam teks ke dalam sentimen positif, negatif, atau netral.
Anotasi	Proses pemberian label sentimen secara manual oleh anotator manusia berdasarkan pedoman tertentu.
Bayesian Optimization	Metode optimasi hiperparameter berbasis probabilistik untuk menemukan konfigurasi parameter terbaik secara efisien.
BiLSTM	Arsitektur jaringan saraf berulang yang memproses data teks dari dua arah untuk menangkap konteks secara menyeluruh.
Class Weight	Bobot yang diberikan pada setiap kelas untuk mengatasi ketidakseimbangan data saat pelatihan model.
Confusion Matrix	Matriks evaluasi yang menunjukkan perbandingan antara kelas aktual dan kelas hasil prediksi model.
Dataset	Kumpulan data teks yang digunakan dalam proses pelatihan dan pengujian model.
Deep Learning	Pendekatan pembelajaran mesin berbasis jaringan saraf dalam untuk menangkap pola kompleks dalam data.
F1-Score	Rata-rata harmonik antara precision dan recall yang digunakan untuk mengevaluasi performa model klasifikasi.
FOMO	Fenomena psikologis berupa rasa cemas atau takut tertinggal informasi, aktivitas, atau tren sosial.



Fine-Tuning	Proses pelatihan ulang model pra-latih menggunakan dataset spesifik untuk meningkatkan performa pada tugas tertentu.
Grid Search	Metode pencarian hiperparameter dengan mencoba seluruh kombinasi parameter yang telah ditentukan.
Hybrid Labeling	Pendekatan pelabelan data yang menggabungkan prediksi model otomatis, validasi leksikon, dan anotasi manual.
Hyperparameter	Parameter eksternal model yang ditentukan sebelum pelatihan dan tidak dipelajari secara otomatis.
IndoBERT	Model bahasa berbasis Transformer yang dilatih khusus menggunakan korpus bahasa Indonesia.
Leksikon Sentimen	Kumpulan kata yang memiliki nilai polaritas sentimen tertentu seperti positif atau negatif.
Macro AUC	Nilai rata-rata Area Under Curve dari seluruh kelas sentimen pada klasifikasi multikelas.
Naïve Bayes	Algoritma klasifikasi berbasis probabilitas yang mengasumsikan independensi antar fitur.
Natural Language Processing	Bidang ilmu yang berfokus pada pemrosesan dan analisis bahasa alami oleh komputer.
Preprocessing Data	Tahap pembersihan dan normalisasi teks sebelum digunakan dalam pemodelan.
Precision	Metrik evaluasi yang mengukur ketepatan prediksi positif yang dihasilkan model.
Recall	Metrik evaluasi yang mengukur kemampuan model dalam menemukan seluruh data positif yang relevan.
ROC Curve	Kurva yang menggambarkan hubungan antara True Positive Rate dan False Positive Rate.
Self-Attention	Mekanisme pada Transformer yang memungkinkan model memahami hubungan antar kata dalam satu kalimat.



Stopword Removal	Proses penghapusan kata umum yang tidak memiliki makna signifikan dalam analisis teks.
Stratified Split	Teknik pembagian data latih dan uji dengan menjaga proporsi kelas tetap seimbang.
Support Vector Machine	Algoritma klasifikasi yang mencari hyperplane optimal untuk memisahkan data ke dalam kelas berbeda.
TF-IDF	Metode representasi teks yang menilai pentingnya kata dalam suatu dokumen relatif terhadap seluruh dokumen.
Transformer	Arsitektur deep learning berbasis mekanisme attention yang unggul dalam pemrosesan bahasa alami.

INTISARI

Pertumbuhan pesat media sosial di Indonesia telah memunculkan fenomena sosial seperti Fear of Missing Out (FOMO). Ekspresi FOMO pada platform seperti X (sebelumnya Twitter) sering kali ditulis secara informal, dipenuhi dengan singkatan, bahasa gaul, dan nuansa emosional, sehingga menjadi tantangan bagi metode Natural Language Processing (NLP) tradisional. Penelitian ini bertujuan untuk mengembangkan model klasifikasi sentimen yang dioptimalkan untuk unggahan terkait FOMO dengan melakukan fine-tuning pada arsitektur IndoBERT serta menerapkan strategi peningkatan kualitas data yang komprehensif. Studi ini memperkenalkan tiga inovasi utama: (1) normalisasi teks secara sistematis untuk menangani ekspresi informal, (2) kerangka hybrid labeling yang menggabungkan prediksi otomatis model, validasi berbasis leksikon, dan anotasi manual untuk membangun data ground-truth berkualitas tinggi, dan (3) hyperparameter tuning menggunakan GridSearchCV untuk model machine learning tradisional serta Bayesian Optimization (Optuna) untuk model deep learning guna memaksimalkan kinerja. Hasil eksperimen menunjukkan bahwa IndoBERT yang telah dioptimalkan mencapai kinerja unggul dengan Accuracy sebesar 94,50%, F1-Score sebesar 94,52%, dan Macro AUC sebesar 0,987. Hasil ini secara signifikan melampaui model pembandingan, termasuk BiLSTM (Accuracy 86,60%), Support Vector Machine (88,06%), dan Naive Bayes (80,73%). Temuan ini menegaskan bahwa pengintegrasian hybrid labeling dan fine-tuned IndoBERT secara signifikan meningkatkan performa klasifikasi sentimen. Penelitian ini berkontribusi pada pengembangan sistem analisis sentimen yang lebih andal untuk mendeteksi dinamika kecemasan sosial serta mendukung riset computational social science dalam konteks Indonesia.

Kata kunci: Sentimen Analisis, FOMO, IndoBERT, Fine-Tuning, Hybrid Labeling

ABSTRACT

The rapid growth of social media in Indonesia has given rise to social phenomena such as Fear of Missing Out (FOMO). Expressions of FOMO on platforms like X (previously Twitter) often written informally, filled with abbreviations, slang, and emotional nuances, posing challenges for traditional Natural Language Processing (NLP) methods. This research aims to develop an optimized sentiment classification model for FOMO-related posts by fine-tuning the IndoBERT architecture and applying comprehensive data enhancement strategies. The study introduces three key innovations: (1) systematic text normalization to handle informal expressions, (2) a hybrid labeling framework combining automated model prediction, lexicon-based validation, and manual annotation to construct high-quality ground-truth data, and (3) hyperparameter tuning using both GridSearchCV for traditional machine learning models and Bayesian Optimization (Optuna) for deep learning models to maximize performance. The experimental results demonstrate that the optimized IndoBERT achieved superior performance with an Accuracy of 94.50%, F1-Score of 94.52%, and Macro AUC of 0.987. These results significantly surpass comparative models, including BiLSTM (Accuracy 86.60%), Support Vector Machine (88.06%), and Naive Bayes (80.73%). These results confirm that integrating hybrid labeling and fine-tuned IndoBERT significantly enhances sentiment classification performance. The findings contribute to developing reliable sentiment analysis systems for detecting social anxiety dynamics and computational social science research in Indonesian contexts.

Keyword: Sentiment Analysis, FOMO, IndoBERT, Fine-tuning, Hybrid Labeling