

PERAMALAN SENTIMEN PUBLIK TERHADAP MAGANG NASIONAL PADA MEDIA SOSIAL X MENGGUNAKAN LSTM DENGAN PELABELAN *LARGE LANGUAGE MODEL*

HAFIDZ CANDRA RAMADHAN



PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Peramalan Sentimen Publik terhadap Magang Nasional pada Media Sosial X Menggunakan LSTM dengan Pelabelan *Large Language Model*” adalah karya saya dengan arahan dari dosen 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.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2026

Hafidz Candra Ramadhan
G1401221079



ABSTRAK

HAFIDZ CANDRA RAMADHAN. Peramalan Sentimen Publik terhadap Magang Nasional pada Media Sosial X Menggunakan LSTM dengan Pelabelan *Large Language Model*. Dibimbing oleh ERFIANI dan SEPTIAN RAHARDIANTORO.

Media sosial menjadi ruang utama bagi publik untuk mengekspresikan opini terhadap kebijakan pemerintah, termasuk program Magang Nasional Kementerian Ketenagakerjaan. Penelitian ini bertujuan mengevaluasi pelabelan otomatis GPT-4o mini dengan strategi *zero-shot* dan *few-shot* terhadap pelabelan manual, membandingkan performa peramalan Long Short-Term Memory (LSTM) berdasarkan sumber pelabelan terbaik dengan dua representasi input berupa persentase sentimen positif harian dan persentase sentimen positif harian yang dihaluskan dengan moving average 7 hari (MA-7), serta memproyeksikan sentimen positif publik terhadap program tersebut. Sebanyak 7.240 *tweet* berbahasa Indonesia dikumpulkan dan dilabeli secara manual oleh dua *annotator* dengan Cohen's Kappa sebesar 0,7545. Strategi *few-shot* mengungguli *zero-shot* dengan *balanced accuracy* 0,9012, Cohen's Kappa 0,8818, dan *macro F1_{score}* 0,9199. Pada tahap peramalan, input MA-7 secara konsisten mengungguli input harian mentah dengan penurunan RMSE uji sebesar 81,1% pada pelabelan manual dan 61,2% pada pelabelan *few-shot*. Proyeksi 30 hari ke depan menunjukkan kecenderungan melemah pada akhir periode. Temuan ini mengindikasikan bahwa pelabelan *few-shot* dapat menjadi alternatif pelabelan manual yang efisien dan *smoothing* MA-7 bisa meningkatkan kinerja peramalan pada data sentimen harian yang berfluktuasi tinggi.

Kata kunci: *Large Language Model*, LSTM, moving average, pelabelan otomatis, peramalan sentimen



ABSTRACT

HAFIDZ CANDRA RAMADHAN. Forecasting Public Sentiment towards National Internship on Social Media X Using LSTM with Large Language Model Labeling. Supervised by ERFIANI and SEPTIAN RAHARDIANTORO.

Social media has become the primary space for the public to express opinions on government policies, including the Ministry of Manpower's National Internship program. This study aims to evaluate the automatic labeling of GPT-4o mini using zero-shot and *few-shot* strategies against manual labeling, compare the forecasting performance of Long Short-Term Memory (LSTM) based on the best labeling source with two input representations namely the daily positive sentiment percentage and the daily positive sentiment percentage smoothed with a 7-day moving average (MA-7), and project public positive sentiment toward the program. A total of 7,240 Indonesian-language tweets were collected and manually labeled by two *annotators*, achieving a Cohen's Kappa of 0.7545. The *few-shot* strategy outperformed the zero-shot strategy, yielding a balanced accuracy of 0.9012, a Cohen's Kappa of 0.8818, and a macro $F1_{score}$ of 0.9199. In the forecasting stage, the MA-7 input consistently outperformed the raw daily input, showing a decrease in test RMSE by 81.1% for manual labeling and 61.2% for *few-shot* labeling. A 30-day ahead projection indicates a weakening trend toward the end of the period. These findings indicate that *few-shot* labeling can serve as an efficient alternative to manual labeling, and MA-7 *smoothing* can enhance forecasting performance on highly fluctuating daily sentiment data.

Keywords: automatic labeling, Large Language Model, LSTM, moving average, sentiment forecasting.



@Hak cipta milik IPB University

IPB University



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026¹
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 IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PERAMALAN SENTIMEN PUBLIK TERHADAP MAGANG NASIONAL PADA MEDIA SOSIAL X MENGGUNAKAN LSTM DENGAN PELABELAN *LARGE LANGUAGE MODEL*

HAFIDZ CANDRA RAMADHAN

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:
Gerry Alfa Dito, S.Si., M.Si



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Peramalan Sentimen Publik terhadap Magang Nasional pada
Media Sosial X Menggunakan LSTM Dengan Pelabelan *Large Language Model*

Nama : Hafidz Candra Ramadhan
NIM : G1401221079

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Dr. Ir. Erfiani, M.Si

Pembimbing 2:

Dr. Septian Rahardiantoro, S.Stat., M.Si

Diketahui oleh

Ketua Program Studi:

Dr. Bagus Sartono, S.Si., M.Si.
NIP 197804112005011002

Tanggal Ujian:
21 Juli 2026

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan November 2025 sampai bulan Juni 2026 ini ialah Peramalan Sentimen, dengan judul “Peramalan Sentimen Publik terhadap Magang Nasional pada Media Sosial X Menggunakan LSTM Dengan Pelabelan *Large Language Model*”.

Penulisan karya ilmiah ini tidak luput dari bantuan, doa, dan dukungan berbagai pihak sehingga karya ilmiah ini sehingga bisa terselesaikan dengan baik. Oleh karena itu, penulis dengan tulus mengucapkan terima kasih kepada :

1. Kedua orang tua tercinta, yaitu Bapak Syukri dan Ibu Isnaeni Ari Santi serta saudara kandung penulis, Ihfazhi Fityatan Imani, atas curahan doa, kasih sayang, motivasi, dan dukungan yang tiada henti selama penulis menjalani masa perkuliahan.
2. Bunda Dr. Ir. Erfiani, M.Si. dan Bapak Dr. Septian Rahardiantoro, S.Stat., M.Si. selaku komisi pembimbing yang telah memberikan bimbingan, arahan, masukan, dan dukungan selama proses penyusunan karya ilmiah ini.
3. Bapak Dr. Agus Mohamad Soleh, S.Si., M.T., Bapak La Ode Abdul Rahman, S.Si., M.Si, dan Ibu Aulia Rizki Firdawanti, S.Stat., M.Si selaku moderator dalam kolokium dan seminar hasil yang turut memberikan saran untuk karya ilmiah ini.
4. Bapak Gerry Alfa Dito, S.Si., M.Si selaku penguji dalam sidang skripsi penulis yang turut memberikan saran untuk karya ilmiah ini.
5. Seluruh dosen dan tenaga pendidik Program Studi Statistika dan Sains Data SSMI IPB yang telah memberikan ilmu dan wawasan yang sangat bermanfaat bagi penulis serta turut membantu penulis dalam kelancaran administrasi selama perkuliahan hingga penulisan tugas akhir.
6. Mahib, Abyan, Fahrezi, Raihan, Yafi, Fauzan, Aji, Dzakwan, Haikal, Ali, dan Wardit yang telah kebersamaan penulis selama penyusunan karya ilmiah ini.
7. Adinda Maudy Maulidya yang telah menjadi sumber semangat bagi penulis dalam menyelesaikan karya ilmiah ini.
8. Seluruh teman-teman Marinestic Statistika 59 yang telah menjadi bagian tak terpisahkan dari perjalanan penulis dalam menempuh pendidikan bersama.
9. Semua pihak yang telah mendukung dan tidak dapat penulis sebutkan namanya satu per satu.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juli 2026

Hafidz Candra Ramadhan



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat	3
II TINJAUAN PUSTAKA	4
2.1 Program Magang Nasional	4
2.2 Analisis Sentimen	4
2.3 Pelabelan Data	4
2.4 Cohen's Kappa	5
2.5 Long Short Term Memory	6
2.6 Metrik Evaluasi Klasifikasi	9
2.7 Moving Average	10
2.8 Metriks Evaluasi Peramalan	11
III METODE	12
3.1 Data	12
3.2 Prosedur Analisis Data	12
IV HASIL DAN PEMBAHASAN	15
4.1 Pengumpulan dan Eksplorasi Data	15
4.2 Pelabelan Manual	15
4.3 Pelabelan Otomatis dengan Strategi <i>Zero-shot</i>	17
4.4 Pelabelan Otomatis dengan Strategi <i>Few-shot</i>	18
4.5 Evaluasi Kinerja Pelabelan	19
4.6 Pembentukan Input deret Waktu	20
4.7 Pemodelan Long Short Term Memory dengan input Persentase Positif Harian	21
4.8 Pemodelan Long Short Term Memory dengan Input Persentase Positif Harian MA-7	23
4.9 Evaluasi dan Perbandingan Kinerja Model	25
4.10 Proyeksi Tren Sentimen Positif 30 Hari ke Depan	26
V SIMPULAN DAN SARAN	28
5.1 Simpulan	28
5.2 Saran	28
DAFTAR PUSTAKA	29
LAMPIRAN	32
RIWAYAT HIDUP	36

DAFTAR TABEL

1	Interpretasi Kappa	6
2	Daftar <i>hyperparameter</i> yang digunakan	8
3	<i>Confusion matrix</i> untuk klasifikasi multi kelas	9
4	Skenario <i>Hyperparameter</i>	14
5	Evaluasi kinerja pelabelan otomatis terhadap pelabelan manual	19
6	Kombinasi <i>hyperparameter</i> terbaik	21
7	Evaluasi kinerja LSTM dengan input persentase positif harian	23
8	Kombinasi <i>hyperparameter</i> terbaik	23
9	Evaluasi kinerja LSTM dengan input persentase positif harian MA-7	25
10	Perbandingan kinerja kedua input model LSTM	26

DAFTAR GAMBAR

1	Arsitektur LSTM	6
2	Frekuensi sebaran <i>tweet</i> awal	15
3	Frekuensi sebaran <i>tweet</i> data final	16
4	<i>Wordcloud</i> keseluruhan <i>tweet</i>	16
5	Distribusi sentimen hasil pelabelan manual	17
6	<i>Wordcloud</i> tiap kategori sentimen hasil pelabelan manual	17
7	Distribusi sentimen hasil pelabelan otomatis <i>zero-shot</i>	18
8	<i>Wordcloud</i> tiap kategori sentimen hasil pelabelan otomatis <i>zero-shot</i>	18
9	Distribusi sentimen hasil pelabelan otomatis <i>few-shot</i>	19
10	<i>Wordcloud</i> tiap kategori sentimen hasil pelabelan otomatis <i>few-shot</i>	19
11	Plot deret waktu (a) input persentase positif harian manual (b) input persentase positif harian <i>few-shot</i> (c) input persentase positif harian MA-7 manual (d) input persentase positif harian MA-7 <i>few-shot</i>	20
12	Plot prediksi (a) pelabelan manual terhadap pelabelan aktual (b) pelabelan <i>few-shot</i> terhadap pelabelan aktual (c) pelabelan manual terhadap <i>ground truth</i> (d) pelabelan <i>few-shot</i> terhadap <i>ground truth</i>	22
13	Plot prediksi (a) pelabelan manual terhadap pelabelan aktual (b) pelabelan <i>few-shot</i> terhadap pelabelan aktual (c) pelabelan manual terhadap <i>ground truth</i> (d) pelabelan <i>few-shot</i> terhadap <i>ground truth</i> (e) pelabelan manual terhadap <i>ground truth</i> harian (f) pelabelan <i>few-shot</i> terhadap pelabelan <i>few-shot</i> harian	24
14	Plot proyeksi sentimen positif 30 titik waktu ke depan (a) pelabelan manual (b) pelabelan <i>few-shot</i>	27

DAFTAR LAMPIRAN

1	Panduan <i>Annotator</i>	33
---	--------------------------	----



2
3

Prompt untuk Few-shot
Prompt untuk Zero-shot

33
35

@Hak cipta milik IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.