



SKRIPSI

***FORECASTING TOPIK, HASHTAG, DAN KATA
BERBASIS LIGHTGBM DENGAN KLASIFIKASI
INTERAKSI TWEET MENGGUNAKAN MLP
UNTUK OPTIMALISASI MARKETING CAMPAIGN
DI X***

DEANNISA SYAFIRA PUTRI
NPM 22083010062

DOSEN PEMBIMBING

Amri Muhaimin, S.Stat., M.Stat., M.S.
Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

**FORECASTING TOPIK, HASHTAG, DAN KATA BERBASIS LIGHTGBM
DENGAN KLASIFIKASI INTERAKSI TWEET MENGGUNAKAN MLP
UNTUK OPTIMALISASI MARKETING CAMPAIGN DI X**

Oleh:
DEANNISA SYAFIRA PUTRI
NPM. 22083010062

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada Tanggal 20 April 2026:

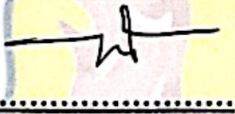
Menyetujui,

Amri Muhaimin, S.Stat., M.Stat., M.S.
NIP. 19950723 202406 1 002



..... (Pembimbing I)

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 19830310 202121 1 006



..... (Pembimbing II)

Aviolla Terza Damaliana, S.Si., M.Stat.
NIP. 19940802 202203 2 015



..... (Ketua Penguji)

Shindi Shella May Wara, M. Stat.
NIP. 19960518 202406 2 003



..... (Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer



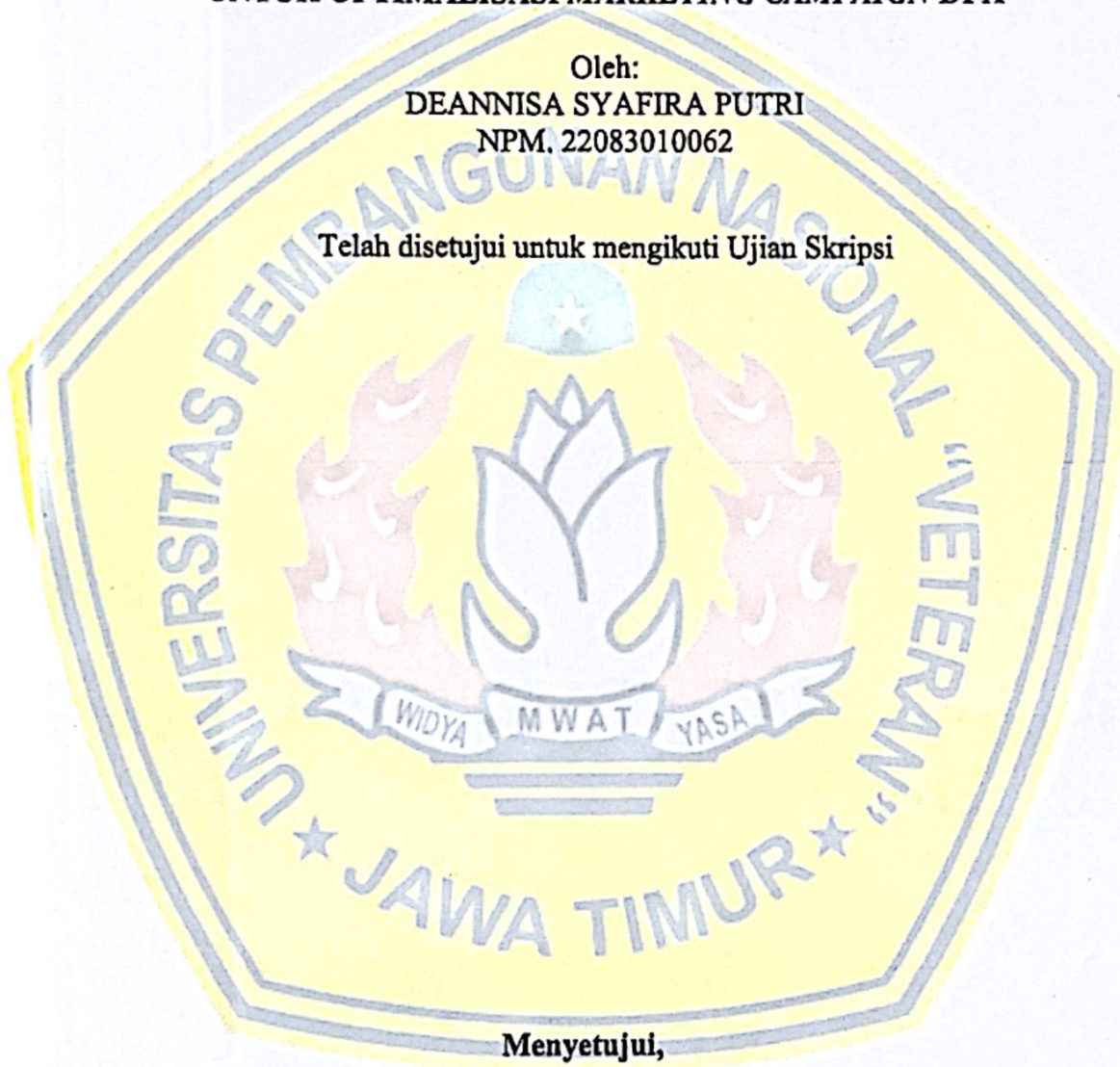
Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**FORECASTING TOPIK, HASHTAG, DAN KATA BERBASIS LIGHTGBM
DENGAN KLASIFIKASI INTERAKSI TWEET MENGGUNAKAN MLP
UNTUK OPTIMALISASI MARKETING CAMPAIGN DI X**

Oleh:
DEANNISA SYAFIRA PUTRI
NPM. 22083010062

Telah disetujui untuk mengikuti Ujian Skripsi



Menyetujui,

**Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., ASEAN Eng.
NIP. 19801205 200501 1 002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Deannisa Syafira Putri
NPM : 22083010062
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 12 Mei 2026
Yang Membuat Pernyataan,



DEANNISA SYAFIRA PUTRI
NPM. 22083010062

ABSTRAK

Nama Mahasiswa / NPM : Deannisa Syafira Putri / 22083010062
Judul Proposal Skripsi : *Forecasting* Topik, *Hashtag*, dan Kata Berbasis LightGBM dengan Klasifikasi Interaksi *Tweet* Menggunakan MLP untuk Optimalisasi *Marketing Campaign* di X
Dosen Pembimbing : 1. Amri Muhaimin, S.Stat., M.Stat., M.S.
2. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

Media sosial, khususnya platform X, memiliki pola penyebaran konten yang cepat dan luas sehingga membuka peluang besar untuk mengoptimalkan strategi kampanye pemasaran berbasis data. Penelitian ini mengembangkan sistem *forecasting* frekuensi topik, *hashtag*, dan kata di platform X sebagai dasar optimalisasi *social media marketing campaign*, didukung model klasifikasi tingkat interaksi *tweet* sebagai analisis pendukung. Data *tweet* berbahasa Indonesia dikumpulkan melalui *scraping* selama 18 bulan dan menghasilkan 3.110 data bersih. *Forecasting* dilakukan menggunakan BERTopic untuk ekstraksi topik dan LightGBM yang dioptimalkan dengan Optuna sebagai model *forecasting*. Model klasifikasi tingkat interaksi dibangun dengan mengintegrasikan IndoBERTweet, RoBERTa, dan metadata numerik melalui *Multi-Layer Perceptron* (MLP). Hasil evaluasi *forecasting* menunjukkan performa baik pada data *hashtag* (RMSE 0,5470; MAE 0,3683; RMSSE 0,7554) dan topik (RMSE 1,5999; MAE 0,5783; RMSSE 0,7754), namun kurang optimal pada data kata (RMSE 3,2833; MAE 1,3928; RMSSE 1,4273). Model klasifikasi mencapai akurasi 0,8358, *precision* 0,8351, *recall* 0,8358, dan *F1-score* 0,8355. Integrasi kedua model menghasilkan *insight* berbasis data yang mendukung perancangan strategi konten yang lebih terarah dan optimal.

Kata Kunci: *Social Media Marketing Campaign, Trending Topik, Kata Populer, BERTopic, LightGBM, Optuna, Prediksi Viralitas, IndoBERTweet, RoBERTa, MLP.*

ABSTRACT

Student Name / NPM : Deannisa Syafira Putri / 22083010062
Undergraduate thesis title : *LightGBM-Based Forecasting of Topics, Hashtags, and Words with MLP-Based Tweet Interaction Classification for Marketing Campaign Optimization on X*
Advisors : 1. Amri Muhaimin, S.Stat., M.Stat., M.S.
2. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

Social media, particularly platform X, exhibits rapid and widespread content dissemination patterns that create significant opportunities for data-driven marketing campaign optimization. This study develops a topic, hashtag, and word frequency forecasting system on platform X as the foundation for social media marketing campaign optimization, supported by a tweet interaction level classification model as supplementary analysis. Indonesian-language tweet data was collected through scraping over 18 months, yielding 3,110 clean records after preprocessing. Forecasting was performed using BERTopic for topic extraction and LightGBM optimized with Optuna as the forecasting model. The interaction level classification model was built by integrating IndoBERTweet, RoBERTa, and numerical metadata through a Multi-Layer Perceptron (MLP). Forecasting evaluation results demonstrate strong performance on hashtag data (RMSE 0.5470; MAE 0.3683; RMSSE 0.7554) and topic data (RMSE 1.5999; MAE 0.5783; RMSSE 0.7754), though performance remains suboptimal on word data (RMSE 3.2833; MAE 1.3928; RMSSE 1.4273). The classification model achieves accuracy of 0.8358, precision of 0.8351, recall of 0.8358, and F1-score of 0.8355. Integration of both models produces data-driven insights that support more targeted and effective content strategy development.

Keywords: *Social Media Marketing Campaign, Trending Topics, Popular Words, BERTopic, LightGBM, Optuna, Virality Prediction, IndoBERTweet, RoBERTa, MLP.*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Forecasting Topik, Hashtag, dan Kata Berbasis LightGBM dengan Klasifikasi Interaksi Tweet Menggunakan MLP untuk Optimalisasi Marketing Campaign di X”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Bapak Amri Muhaimin, S.Stat., M.Stat., M.S. selaku Dosen Pembimbing pertama dan Bapak Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T. selaku Dosen Pembimbing kedua yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Dan penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Bapak Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., ASEAN Eng., selaku Ketua Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Dosen-dosen Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan ilmu dan dukungan selama masa perkuliahan.
4. Kedua orang tua penulis yang senantiasa memberikan doa, dukungan moral, dan semangat tanpa henti.
5. Diri sendiri yang telah berjuang, bertahan, dan berusaha semaksimal mungkin dalam menyelesaikan seluruh proses perkuliahan hingga penyusunan skripsi ini.
6. Teman-teman seperjuangan di Program Studi Sains Data angkatan 2022 yang selalu memberikan bantuan, kebersamaan, serta motivasi selama proses perkuliahan hingga penyusunan skripsi ini.

7. Teman-teman satu dosen pembimbing yang telah saling berbagi ilmu, pengalaman, dan dukungan selama proses penelitian.
8. Teman-teman kelompok “Bismillah IPS 4” dan Amirah Rizky Ramadhanti yang telah menjadi sumber semangat dan kebersamaan selama perjalanan akademik.
9. Awita Dori Jatiningtyas dan Balqis Nur Syifa yang telah menjadi teman baik penulis dan selalu memberikan dukungan hingga saat ini.
10. Muhammad Rusydi Athaillah Akbar yang telah menjadi salah satu support system terbaik, selalu memberikan dukungan, semangat, perhatian, serta motivasi yang tiada henti kepada penulis selama proses penyusunan skripsi ini.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, April 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xvi
DAFTAR LAMPIRAN	xviii
DAFTAR NOTASI	xix
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	6
1.3. Batasan Masalah	7
1.4. Tujuan Penelitian	7
1.5. Manfaat Penelitian	8
BAB II TINJAUAN PUSTAKA	9
2.1. Penelitian Terdahulu	9
2.2. Landasan Teori.....	19
2.2.1. Media Sosial dan <i>Digital Marketing</i>	19
2.2.2. Twitter/X.....	20
2.2.3. <i>Natural Language Processing (NLP)</i>	22
2.2.4. <i>Topic Modeling</i>	24
2.2.5. <i>Time Series Forecasting</i>	25
2.2.6. <i>Light Gradient Boosting Machine (LightGBM)</i>	27
2.2.7. <i>Hyperparameter Tuning</i>	31
2.2.8. Klasifikasi	34
2.2.9. <i>Transformer Models</i>	36

2.2.9.	<i>Gaussian Mixture Model (GMM)</i>	43
2.2.10.	<i>Multi-Layer Perceptron (MLP)</i>	46
2.2.11.	Metrik Evaluasi.....	49
2.2.12.	Streamlit.....	56
BAB III	METODOLOGI	57
3.1.	Variabel Penelitian dan Sumber Data.....	57
3.2.	Langkah Analisis	62
3.2.1.	Pengumpulan Data.....	62
3.2.2.	<i>Preprocessing Data</i>	63
3.2.3.	Analisis Deskriptif Data	69
3.2.4.	Ekstraksi Topik, <i>Hashtag</i> , dan Kata.....	69
3.2.5.	Hitung Frekuensi Kemunculan.....	71
3.2.6.	<i>Feature Engineering Time Series</i>	71
3.2.7.	Pembangunan Model <i>Forecasting</i>	73
3.2.8.	Hasil Evaluasi Model <i>Forecasting</i>	77
3.2.9.	Analisis Pendukung Klasifikasi Tingkat Interaksi <i>Tweet</i>	78
3.2.10.	Membangun GUI (<i>Website</i> berbasis Streamlit).....	85
3.2.11.	Interpretasi Hasil.....	86
3.3.	Desain Sistem	86
3.3.1.	Halaman Utama / <i>Homepage</i>	88
3.3.2.	Halaman <i>Data Preprocessing</i>	88
3.3.3.	Halaman <i>Forecasting</i> Frekuensi.....	88
3.3.4.	Halaman Analisis Pendukung.....	89
BAB IV	HASIL DAN PEMBAHASAN	91
4.1.	Pengumpulan Data.....	91
4.2.	<i>Preprocessing Data</i>	96
4.2.1.	<i>Merge Data</i>	97
4.2.2.	Analisis Statistika Deskriptif.....	99
4.2.3.	Menghapus Kolom yang tidak Relevan, Mengubah Nama dan Format Kolom.....	101
4.2.4.	Penghapusan Data Duplikat.....	103

4.2.5.	Menambah Kolom Daftar <i>Hashtag</i>	104
4.2.6.	Menambahkan Kolom Jumlah <i>Mention</i> dan Jumlah <i>Hashtag</i>	106
4.2.7.	Menghapus URL (Tautan), Karakter Non-alfabet, dan Normalisasi Emoji	108
4.2.8.	<i>Filter</i> Bahasa Indonesia dan Inggris dalam Teks <i>Tweet</i>	111
4.2.9.	Menerjemahkan <i>Tweet</i> Berbahasa Inggris ke Bahasa Indonesia	115
4.2.10.	Menghitung Panjang <i>Tweet</i>	119
4.2.11.	<i>Handling Missing Value</i>	121
4.3.	Analisis Deskriptif Data.....	123
4.4.	Ekstraksi Topik, <i>Hashtag</i> , dan Kata	131
4.4.1.	Ekstraksi Topik (<i>Topic Modeling</i>).....	132
4.4.2.	Ekstraksi <i>Hashtag</i> dan Kata.....	144
4.5.	Perhitung Frekuensi Kemunculan dan Pemilihan <i>Top 10</i> Topik, <i>Hashtag</i> , dan Kata.....	146
4.6.	<i>Feature Engineering Time Series</i>	152
4.7.	Pembangunan Model <i>Forecasting</i>	158
4.7.1.	Hasil <i>Forecasting</i> Topik	166
4.7.2.	Hasil <i>Forecasting Hashtag</i>	175
4.7.3.	Hasil <i>Forecasting</i> Kata	183
4.8.	Hasil Evaluasi Model <i>Forecasting</i>	192
4.9.	Analisis Pendukung Klasifikasi Tingkat Interaksi <i>Tweet</i>	195
4.9.1.	<i>Labeling</i> dan <i>Splitting</i> Data	196
4.9.2.	Ekstraksi Fitur dan Arsitektur MLP.....	202
4.9.3.	Pelatihan Model	213
4.9.4.	Hasil Evaluasi Model.....	223
4.9.5.	Hasil Model <i>Ablation (Feature Importance)</i>	226
4.9.6.	Hasil Prediksi pada <i>Tweet</i> Baru	230
4.10.	Tampilan Antar Muka (<i>Graphical User Interface</i>)	234
4.10.1.	Halaman <i>Home</i>	234
4.10.2.	Halaman <i>Data & Preprocessing</i>	235
4.10.3.	Halaman Model <i>Forecasting</i>	238

4.10.4. Halaman Analisis Pendukung.....	241
BAB V PENUTUP.....	247
5.1. Kesimpulan.....	247
5.2. Saran Pengembangan.....	249
DAFTAR PUSTAKA	251
LAMPIRAN	261

DAFTAR GAMBAR

Gambar 2.1. Perubahan Logo Twitter ke X	22
Gambar 2.2. Struktur LightGBM	28
Gambar 2.3. Arsitektur BERT	37
Gambar 2.4. Struktur BERTopic	41
Gambar 3.1. Diagram Alir Penelitian	62
Gambar 3.2. Diagram Alir Tahap <i>Data Preprocessing</i>	63
Gambar 3.3. Diagram Alir Analisis Pendukung Klasifikasi Tingkat Interaksi <i>Tweet</i>	78
Gambar 3.4. Arsitektur MLP yang Digunakan	81
Gambar 3.5. Rancangan Desain Sistem	87
Gambar 4.1. Distribusi <i>Engagement</i> : (a) Jumlah <i>Like</i> (b) Jumlah <i>Retweet</i>	125
Gambar 4.2. Distribusi <i>Engagement</i> : (a) Jumlah <i>Reply</i> (b) Jumlah <i>Quote</i>	126
Gambar 4.3. Distribusi Panjang <i>Tweet</i>	127
Gambar 4.4. <i>Word Cloud Most Common Words</i>	128
Gambar 4.5. <i>Time Series Plot</i>	131
Gambar 4.6. Hasil <i>Forecasting</i> Frekuensi Topic_0	166
Gambar 4.7. Hasil <i>Forecasting</i> Frekuensi Topic_1	167
Gambar 4.8. Hasil <i>Forecasting</i> Frekuensi Topic_2	168
Gambar 4.9. Hasil <i>Forecasting</i> Frekuensi Topic_4	169
Gambar 4.10. Hasil <i>Forecasting</i> Frekuensi Topic_3	170
Gambar 4.11. Hasil <i>Forecasting</i> Frekuensi Topic_5	170
Gambar 4.12. Hasil <i>Forecasting</i> Frekuensi Topic_6	171
Gambar 4.13. Hasil <i>Forecasting</i> Frekuensi Topic_7	172
Gambar 4.14. Hasil <i>Forecasting</i> Frekuensi Topic_8	173
Gambar 4.15. Hasil <i>Forecasting</i> Frekuensi Topic_9	174
Gambar 4.16. Hasil <i>Forecasting</i> Frekuensi <i>Hashtag</i> #waktuindonesiabelanja	175
Gambar 4.17. Hasil <i>Forecasting</i> Frekuensi <i>Hashtag</i> #lazadagiveaway	176
Gambar 4.18. Hasil <i>Forecasting</i> Frekuensi <i>Hashtag</i> #lazadaid	177
Gambar 4.19. Hasil <i>Forecasting</i> Frekuensi <i>Hashtag</i> #termurahdishopee	177
Gambar 4.20. Hasil <i>Forecasting</i> Frekuensi <i>Hashtag</i> #cekshopeedulu	178

Gambar 4.21. Hasil <i>Forecasting</i> Frekuensi <i>Hashtag</i> #promoguncangtokopedia	179
Gambar 4.22. Hasil <i>Forecasting</i> Frekuensi <i>Hashtag</i> #tokopediachingu	180
Gambar 4.23. Hasil <i>Forecasting</i> Frekuensi <i>Hashtag</i> #smartfren	180
Gambar 4.24. Hasil <i>Forecasting</i> Frekuensi <i>Hashtag</i> #100persenuntukindonesia	181
Gambar 4.25. Hasil <i>Forecasting</i> Frekuensi <i>Hashtag</i> #zonauang	182
Gambar 4.26. Hasil <i>Forecasting</i> Frekuensi Kata “shopee”	183
Gambar 4.27. Hasil <i>Forecasting</i> Frekuensi Kata “cek”	184
Gambar 4.28. Hasil <i>Forecasting</i> Frekuensi Kata “sekarang”	185
Gambar 4.29. Hasil <i>Forecasting</i> Frekuensi Kata “gratis”	186
Gambar 4.30. Hasil <i>Forecasting</i> Frekuensi Kata “paket”	186
Gambar 4.31. Hasil <i>Forecasting</i> Frekuensi Kata “voucher”	187
Gambar 4.32. Hasil <i>Forecasting</i> Frekuensi Kata “buat”	188
Gambar 4.33. Hasil <i>Forecasting</i> Frekuensi Kata “ongkir”	189
Gambar 4.34. Hasil <i>Forecasting</i> Frekuensi Kata “belanja”	190
Gambar 4.35. Hasil <i>Forecasting</i> Frekuensi Kata “hari”	190
Gambar 4.36. Grafik <i>Training Loss</i> , <i>Validation Loss</i> dan <i>Validation Metrics</i> ..	222
Gambar 4.37. Tampilan Halaman <i>Home</i>	235
Gambar 4.38. Tampilan Utama Halaman <i>Data & Preprocessing</i>	236
Gambar 4.39. Tampilan <i>Tab Data Integration & Overview</i> : (a) <i>Upload</i> data dan (b) Setelah Proses <i>Preprocessing</i>	236
Gambar 4.40. Tampilan <i>Tab Preprocessing Steps</i>	237
Gambar 4.41. Tampilan <i>Tab Data Upload & Process Data</i> : (a) Sebelum <i>Upload</i> data dan (b) Setelah Proses <i>Preprocessing</i>	237
Gambar 4.42. Tampilan <i>Tab Visualisasi</i> : (a) Sebelum <i>Upload</i> data dan (b) Hasil Visualisasi Data	238
Gambar 4.43. Tampilan Utama Halaman Model <i>Forecasting</i>	239
Gambar 4.44. Tampilan Halaman Model <i>Forecasting</i> Setelah <i>Upload</i> Data	239
Gambar 4.45. Tampilan Halaman Model <i>Forecasting</i> : (a) Hasil Proses <i>Topic</i> <i>Modeling</i> dan (b) Contoh <i>Tweet</i> per Topik	240

Gambar 4.46. Tampilan Halaman Model <i>Forecasting</i> Bagian Analisis Frekuensi	240
Gambar 4.47. Tampilan Halaman Model <i>Forecasting</i> : (a) Proses <i>Forecasting</i> dan (b) Hasil <i>Forecasting</i>	241
Gambar 4.48. Tampilan Utama Halaman Model Klasifikasi	242
Gambar 4.49. Tampilan Hasil Pelabelan pada Tab Upload & Pelabelan.....	242
Gambar 4.50. Tampilan <i>Tab Training</i> Model: (a) Sebelum Proses dijalankan dan (b) Sesudah Proses Dijalankan.....	243
Gambar 4.51. Tampilan <i>Tab Evaluasi & Perbandingan</i> : (a) Hasil Evaluasi Model dan (b) Hasil Evaluasi Model Prediksi <i>Engagement</i>	243
Gambar 4.52. Tampilan <i>Tab Feature Importance</i> : (a) Sebelum Proses dan (b) Sesudah Proses.....	244
Gambar 4.53. Tampilan <i>Tab Prediksi Tweet</i> Baru: (a) Tampilan Awal dan (b) Bagian Masukkan Detail <i>Tweet</i>	244
Gambar 4.54. Tampilan Hasil Prediksi <i>Tweet</i> Baru: (a) Hasil Prediksi Viralitas dan (b) Hasil Prediksi <i>Engagement</i>	245

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu	9
Tabel 3.1. Struktur Variabel Data	58
Tabel 3.2. Struktur Data Model <i>Forecasting</i>	60
Tabel 3.3. Struktur Data pada Model Analisis Pendukung	61
Tabel 3.4. Konfigurasi Umum Proses <i>Forecasting</i>	74
Tabel 3.5. Konfigurasi Optimasi <i>Hyperparameter</i> Optuna.....	74
Tabel 3.6. Rentang <i>Hyperparameter</i> LightGBM yang Dioptimasi oleh Optuna .	75
Tabel 3.7. Konfigurasi Proses Pelatihan Model	76
Tabel 3.8. Konfigurasi pada arsitektur model MLP	82
Tabel 3.9. <i>Hyperparameter</i> Pelatihan Model.....	83
Tabel 4.1. Hasil Proses <i>Scraping</i> Data.....	93
Tabel 4.2. Hasil Pengumpulan Data.....	95
Tabel 4.3. Hasil Statistika Deskriptif Sebelum <i>Preprocessing</i>	100
Tabel 4.4. Hasil Penambahan Kolom Daftar Hashtag	105
Tabel 4.5. Hasil Penambahan Kolom Jumlah <i>Mention</i> dan Jumlah <i>Hashtag</i>	107
Tabel 4.6. Hasil Teks <i>Tweet</i> Setelah Dibersihkan dan Dinormalisasi.....	109
Tabel 4.7. Hasil Teks <i>Tweet</i> yang telah Diterjemahkan	117
Tabel 4.8. Hasil Penambahan Kolom Panjang <i>Tweet</i>	120
Tabel 4.9. Hasil Pengecekan <i>Missing Value</i>	121
Tabel 4.10. Hasil Statistika Deskriptif Sesudah <i>Preprocessing</i>	128
Tabel 4.11. Hasil Penentuan Jumlah Topik Terbaik	135
Tabel 4.12. Hasil Distribusi Topik yang Terbentuk.....	138
Tabel 4.13. Hasil Statistik Probabilitas	142
Tabel 4.14. Hasil Distribusi Tingkat Kepercayaan pada Topik yang Terbentuk	143
Tabel 4.15. <i>Top 10</i> Topik yang Sering Muncul	147
Tabel 4.16. <i>Top 10 Hashtag</i> yang Sering Muncul	149
Tabel 4.17. <i>Top 10</i> Kata yang Sering Muncul	150
Tabel 4.18. Contoh Data untuk Perhitungan <i>Lag Features</i>	152
Tabel 4.19. Contoh Data dalam Bentuk <i>Time Series</i>	153
Tabel 4.20. Hasil Contoh Pehitungan Pembentukan Fitur	155

Tabel 4.21. Struktur Data <i>Forecasting</i> Setelah <i>Feature Engineering</i>	157
Tabel 4.22. Contoh Hasil <i>Trial</i> Optuna	160
Tabel 4.23. Rata-rata Hasil Evaluasi Model <i>Forecasting</i> pada Topik.....	193
Tabel 4.24. Rata-rata Hasil Evaluasi Model <i>Forecasting</i> pada <i>Hashtag</i>	193
Tabel 4.25. Rata-rata Hasil Evaluasi Model <i>Forecasting</i> pada Kata.....	194
Tabel 4.26. Rata-rata Fitur Setiap Label	200
Tabel 4.27. Contoh Hasil Token ID	203
Tabel 4.28. Contoh Fitur Numerik.....	208
Tabel 4.29. Hasil Pelatihan Model.....	221
Tabel 4.30. Hasil Evaluasi Model Klasifikasi	224
Tabel 4.31. Hasil Evaluasi Model Prediksi <i>Engagement</i>	225
Tabel 4.32. Hasil Evaluasi Model <i>Ablation</i>	228

DAFTAR LAMPIRAN

Lampiran 1. Dataset Penelitian	261
Lampiran 2. Kode Penelitian.....	262
Lampiran 3. Kode GUI	263
Lampiran 4. LoA Publikasi Jurnal Sinta	264

DAFTAR NOTASI

x	:	Data atau input pada model
$P(x)$:	Probabilitas data x menggunakan model Gaussian Mixture Model (GMM)
K	:	Jumlah komponen Gaussian dalam GMM
π_K	:	Bobot komponen ke- k
$N(x \mu_k, \Sigma_k)$:	Distribusi normal dengan mean dan kovarians untuk komponen ke- k
Σ_k	:	Kovarians ke- k
Q, K, V	:	<i>Query, Key, dan Value</i> dalam mekanisme <i>self-attention Transformer</i>
d_k	:	Dimensi vektor <i>Key</i> pada <i>Transformer</i>
W_1, W_2	:	Matriks bobot pada lapisan linear <i>Feed Forward Neural Network</i> (FFNN)
b_1, b_2	:	Bias pada lapisan linear FFNN
$ReLU$:	Fungsi aktivasi <i>Rectified Linear Unit</i>
p_i	:	Probabilitas prediksi kelas ke- i dari fungsi Softmax
L_{MLM}	:	<i>Loss function</i> pada <i>Masked Language Modeling</i> (MLM)
L_{NSP}	:	<i>Loss function</i> pada <i>Next Sentence Prediction</i> (NSP)
M_x	:	Himpunan posisi token yang di- <i>mask</i>
x_i	:	Token <i>input</i> yang di- <i>mask</i>
$y(t)$:	Nilai deret waktu pada waktu t
$g(t)$:	Komponen tren (<i>trend component</i>)
$s(t)$:	Komponen musiman (<i>seasonality component</i>)
$a(t)$:	Komponen autoregresif / efek <i>lag</i>
ϵ_t	:	Komponen <i>noise</i> residual
k, m	:	<i>Slope</i> dan <i>intercept</i> pada fungsi tren dasar
δ_j	:	Perubahan kemiringan pada titik perubahan (<i>change point</i>) ke- j

$a_j(t)$:	Indikator perubahan tren
P	:	Periode musiman
N	:	Jumlah harmonisasi (komponen <i>Fourier</i>)
α_n, β_n	:	Koefisien <i>Fourier</i> untuk <i>seasonality</i>
L	:	Orde <i>lag</i> maksimal
w_l	:	Bobot untuk lag ke- l
λ	:	Parameter peluruhan eksponensial (<i>decay parameter</i>)
t	:	Waktu prediksi saat ini
\mathcal{L}	:	<i>Loss function</i> (umumnya <i>Mean Squared Error</i>)
Ω	:	Fungsi regularisasi kompleksitas model
G_L, G_R	:	Jumlah gradien pada sisi kiri dan kanan saat pemisahan <i>node</i> di LightGBM
H_L, H_R	:	Jumlah <i>hessian</i> pada sisi kiri dan kanan
λ_1, λ_2	:	Parameter regulasi pada LightGBM
BP/TP	:	Jumlah dari <i>True Positive</i>
BN/TN	:	Jumlah dari <i>True Negative</i>
SP/FP	:	Jumlah dari <i>False Positive</i>
SN/FN	:	Jumlah dari <i>False Negative</i>
n	:	Data ke- n
y_i	:	Nilai aktual
\hat{y}_i	:	Nilai prediksi
Acc	:	Akurasi model
$Prec$:	<i>Precision</i> (ketepatan prediksi positif)
Rec	:	<i>Recall</i> (tingkat keberhasilan menangkap kelas positif)
$F1$:	<i>F1-score</i> (rata-rata harmonis antara <i>precision</i> dan <i>recall</i>)
$RMSE$:	<i>Root Mean Squared Error</i>
MAE	:	<i>Mean Absolute Error</i>
$MAPE$:	<i>Mean Absolute Percentage Error</i>