

Prediksi Inflasi Menggunakan *Deep Learning* dengan Integrasi Variabel Makroekonomi dan Data Narasi Digital sebagai Variabel Eksogen

Riza Damayanti*, Irhamah, Tintrim Dwi Ary Widhianingsih

Institut Teknologi Sepuluh Nopember Surabaya, Indonesia

Email: 6003241009@student.its.ac.id*, irhamah@its.ac.id, dwi.ary@its.ac.id

Keywords:

Deep Learning; Dummy Structural Break; LSTM; Digital Narrative; Ablation Study.

Abstract

This research extends the authors' prior work a six-model LSTM ablation research for Indonesian inflation forecasting (manuscript submitted to AMLDS 2026, under review) — by evaluating whether structural break dummy variables (Lebaran, COVID-19, fuel price hike) enhance models integrating macroeconomic and digital narrative variables. Using the same dataset (January 2005-December 2024, 240 monthly observations), this paper introduces M3_Naratif_Dummy as the new model contribution, combining Google Trends Index and IEH Bank Indonesia with three structural break dummies. Seven models are compared through ablation study: ARIMA, VAR, M1 LSTM Baseline, M2 LSTM Numerical, M3 LSTM Narrative, M3 LSTM Narrative+Dummy, and M4 LSTM Full. Hyperparameters are selected via sequential grid search (N=10 replications), yielding T=6. Evaluation on test set 2020-2024 (60 months) shows M3 LSTM Narrative remains the best model (RMSE=0.7823, sMAPE=17.72%, std=0.53%), while M3_Naratif_Dummy achieves sMAPE=22.98% with 8.4x higher variance (std=4.43%). At optimal T=6, both models converge to near-identical performance (sMAPE 17.37% vs 17.57%), and Diebold-Mariano test confirms no significant difference (p=0.845). All LSTM models significantly outperform VAR (p<0.05). Key finding: digital narrative variables implicitly capture structural break information, making explicit dummy variables redundant and destabilizing for LSTM inflation forecasting.*

Kata Kunci:

Deep Learning; Dummy Structural Break; LSTM; Naratif Digital; Ablation Study.

Abstrak

Penelitian ini memperluas studi pendahuluan penulis ablation study enam model LSTM untuk peramalan inflasi Indonesia (manuscript submitted ke AMLDS 2026, under review) dengan mengevaluasi apakah variabel *dummy structural break* (Lebaran, COVID-19, kenaikan BBM) meningkatkan performa model yang mengintegrasikan variabel makroekonomi dan naratif digital. Menggunakan dataset yang sama (Januari 2005-Desember 2024, 240 observasi bulanan), penelitian ini memperkenalkan M3_Naratif_Dummy sebagai model baru yang menggabungkan Google Trends Index dan IEH Bank Indonesia dengan tiga *dummy structural break*. Tujuh model dibandingkan: ARIMA, VAR, M1 LSTM Baseline, M2 LSTM Numerik, M3 LSTM Naratif, M3 LSTM Naratif+Dummy, dan M4 LSTM Full (N=10 replikasi, T*=6). Hasil menunjukkan M3 LSTM Naratif tetap menjadi model terbaik (RMSE=0,7823, sMAPE=17,72%, std=0,53%), sementara M3_Naratif_Dummy mencapai sMAPE=22,98% dengan varian 8,4 kali lebih tinggi (std=4,43%). Pada *lookback* optimal T=6, kedua model konvergen ke performa hampir setara (sMAPE 17,37% vs 17,57%), dan uji *Diebold-Mariano* mengonfirmasi tidak ada perbedaan signifikan (p=0,845). Seluruh model LSTM secara signifikan mengungguli VAR (p<0,05). Temuan utama: variabel

PENDAHULUAN

Inflasi Indonesia mengalami pergeseran regime yang signifikan selama dua dekade terakhir, dengan rata-rata 6,23% pada 2005-2019 dan 2,76% pada 2020-2024. Studi pendahuluan oleh penulis yang sama saat ini dalam proses review di konferensi AMLDS 2026 telah mengembangkan ablation study enam model LSTM mencakup ARIMA, VAR, M1 Baseline, M2 Numerik, M3 Naratif, dan M4 Full, menggunakan variabel makroekonomi dan naratif digital. Studi (Damayanti et al., 2026) menemukan M3 Naratif (LSTM berbasis Google Trends Index/GTI dan IEH Bank Indonesia) sebagai model terbaik (sMAPE=17,79%). Namun studi (Damayanti et al., 2026) tidak menyertakan variabel dummy structural break (Lebaran, COVID-19, kenaikan BBM) dalam desain ablation-nya, meninggalkan pertanyaan metodologis yang menjadi fokus jurnal ini (Statistik, 2024).

Keputusan untuk tidak memasukkan variabel dummy dalam (Damayanti et al., 2026) didasarkan pada prioritas eksplorasi kontribusi variabel naratif digital. Pertanyaan yang tersisa adalah: apakah model terbaik M3_Naratif dapat ditingkatkan lebih lanjut dengan menambahkan informasi *structural break secara eksplisit*? Atau justru, variabel GTI dan IEH sudah cukup menangkap informasi tersebut secara implisit? Jurnal ini menjawab pertanyaan ini dengan memperkenalkan M3_Naratif_Dummy sebagai model ketujuh yang belum pernah diuji sebelumnya (Hochreiter & Schmidhuber, 1997).

Variabel dummy struktural telah digunakan secara luas dalam model ekonometrika untuk mengontrol pergeseran level yang dipicu oleh kejadian diskret seperti perubahan kebijakan, bencana, atau siklus kalender (Almosova & Andresen, 2023). Dalam konteks inflasi Indonesia, dummy Lebaran mengontrol lonjakan inflasi pangan musiman, dummy COVID-19 mengontrol guncangan permintaan 2020-2022, dan dummy kenaikan BBM mengontrol transmisi harga energi ke CPI (Arif et al., 2024; Das & Das, 2024; Eugster & Uhl, 2024; Zahara & Sugianto, 2021). Pertanyaan yang belum terjawab adalah apakah penambahan dummy ini ke model LSTM berbasis variabel naratif digital akan meningkatkan akurasi, atau justru menimbulkan noise mengingat variabel naratif digital mungkin sudah secara implisit mengandung informasi structural break tersebut (Stock & Watson, 2003).

Meskipun penelitian ini memberikan kontribusi penting, terdapat beberapa keterbatasan yang perlu diakui. Penelitian ini hanya menggunakan data bulanan 2005-2024 (240 observasi) yang relatif terbatas untuk model *deep learning*, belum mempertimbangkan variabel naratif alternatif seperti berita online atau sentimen pasar, dan hanya menggunakan tiga jenis *dummy structural break* (Lebaran, COVID-19, BBM) sehingga belum mencakup seluruh potensi guncangan struktural. Selain itu, penelitian ini menggunakan arsitektur LSTM standar tanpa mekanisme *attention*, validasi model terbatas pada data historis tanpa pengujian *real-time*, dan belum melakukan analisis sensitivitas ekstensif terhadap pemilihan variabel naratif dan metode normalisasi. Keterbatasan ini menjadi arah bagi penelitian selanjutnya untuk mengembangkan model peramalan inflasi yang lebih komprehensif.

Penelitian ini menjawab pertanyaan tersebut melalui desain ablation study tujuh model yang memperkenalkan M3_Naratif_Dummy sebagai model baru. Penelitian ini bertujuan untuk mengevaluasi dampak penambahan variabel *dummy structural break* terhadap performa

model LSTM naratif dalam peramalan inflasi Indonesia, menganalisis stabilitas model dengan dan tanpa variabel *dummy* melalui N=10 replikasi, serta menguji signifikansi perbedaan akurasi antar model menggunakan uji Diebold-Mariano. Kontribusi utama penelitian ini adalah: (1) evaluasi empiris pertama dampak variabel *dummy* struktural pada model LSTM naratif untuk inflasi Indonesia; (2) analisis stabilitas model menggunakan N=10 replikasi yang memungkinkan perbandingan mean dan standar deviasi performa; dan (3) pengujian signifikansi menggunakan Diebold-Mariano test untuk mengonfirmasi apakah perbedaan akurasi antar model bermakna secara statistik (Diebold & Mariano, 1995).

METODE

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimental komparatif yang menggunakan desain *ablation study*. Pendekatan eksperimental dipilih karena penelitian ini bertujuan untuk menguji secara empiris pengaruh penambahan variabel *dummy structural break* terhadap performa model LSTM dalam peramalan inflasi Indonesia melalui perbandingan sistematis antar model. Desain *ablation study* memungkinkan peneliti untuk mengisolasi kontribusi masing-masing komponen variabel dengan membandingkan model yang memiliki arsitektur serupa namun dengan input yang berbeda (M3_Naratif vs M3_Naratif_Dummy), sehingga dampak penambahan variabel *dummy* dapat diukur secara akurat.

Data dan Variabel Penelitian dan Kategori

Data mencakup periode Januari 2005 hingga Desember 2024 (240 observasi bulanan). Periode training: Januari 2005-Desember 2019 (180 bulan); periode test: Januari 2020-Desember 2024 (60 bulan). Sepuluh variabel terdiri atas inflasi YoY (target), nilai tukar IDR/USD, suku bunga BI Rate, harga minyak Brent (makroekonomi), Google Trends Index (GTI) dan IEH SK-BI (naratif digital), serta *dummy_lebaran*, *dummy_covid*, dan *dummy_kenaikan_bbm* (struktural). Variabel *dummy* dibuat secara otomatis berdasarkan kalender Idul Fitri, periode pandemi COVID-19 (Januari 2020-Desember 2022), dan tanggal kenaikan BBM signifikan (Indonesia, 2024).

Tabel 1. Variabel Penelitian dan Kategori

Variabel	Notasi	Sumber	Peran
Inflasi YoY	Y_t	BPS	Target
Nilai Tukar IDR/USD	X_1	Bank Indonesia	Makroekonomi
Suku Bunga BI Rate	X_2	Bank Indonesia	Makroekonomi
Harga Minyak Brent	X_3	FRED	Makroekonomi
Google Trends Index (GTI)	GT_t	Google Trends	Naratif Digital
IEH Survei Konsumen BI	IEH_t	Bank Indonesia	Naratif Digital
Dummy Struktural (3 var)	dum_t	Kalender/BPS/Keputusan Pemerintah	Struktural

Sumber: Diolah oleh Peneliti, 2026

Preprocessing dan Normalisasi

Target variabel LSTM menggunakan first differencing $d_inflasi_t = inflasi_t - inflasi_{t-1}$ untuk mengatasi pergeseran regime antara training (mean=6,23%) dan test (mean=2,76%). Normalisasi MinMax di-fit hanya pada data training 2005-2019 untuk mencegah data leakage. Variabel *dummy* (nilai 0 atau 1) tidak dinormalisasi karena sudah dalam rentang [0,1]. Robustness check terhadap tiga metode normalisasi (MinMax, Standard, Robust) menunjukkan

hasil kompetitif, dengan Robust memberikan sMAPE terbaik (18,03%) pada M3_Naratif_Dummy, mengonfirmasi stabilitas pipeline preprocessing yang digunakan.

Arsitektur LSTM

LSTM merupakan arsitektur Recurrent Neural Network yang mampu memodelkan dependensi temporal jangka panjang melalui mekanisme tiga gate (forget, input, output) yang mengontrol aliran informasi melalui cell state (Harvey et al., 1997). Keunggulan ini menjadikan LSTM sesuai untuk memodelkan interaksi antara variabel naratif yang bersifat forward-looking dengan variabel makroekonomi yang memiliki pola musiman dan structural break yang khas inflasi Indonesia.

Setiap unit LSTM pada timestep t menerima input x_t , hidden state sebelumnya $h_{(t-1)}$ dan cell state sebelumnya $C_{(t-1)}$. Proses selanjutnya dilakukan dengan menghasilkan h_t dan C_t baru melalui tiga gate:

1. Forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

2. Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

3. Update:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (4)$$

4. Output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

Arsitektur model yang digunakan merupakan stacked Long Short-Term Memory (LSTM) dua lapis yang dirancang untuk menangkap pola temporal secara mendalam pada data deret waktu (Alkahfi et al., 2024; Huang et al., 2024; Joseph et al., 2024). Lapisan pertama terdiri dari LSTM dengan 64 unit, yang berfungsi mengekstraksi pola jangka pendek maupun jangka panjang dari data masukan secara lebih kompleks. Setelah itu, diterapkan Dropout untuk mengurangi risiko overfitting dengan menonaktifkan sebagian neuron secara acak selama proses pelatihan. Selanjutnya, keluaran dari lapisan pertama diproses oleh lapisan LSTM kedua dengan 32 unit, yang berperan melakukan pemodelan lanjutan terhadap representasi fitur yang telah dipelajari sebelumnya sehingga hubungan temporal yang lebih abstrak dapat ditangkap. Setelah lapisan kedua, kembali digunakan Dropout untuk meningkatkan generalisasi model. Pada tahap akhir, digunakan Dense layer dengan satu neuron pada Dense (1) sebagai lapisan keluaran untuk menghasilkan nilai prediksi tunggal sesuai tujuan peramalan.

Desain Ablation Study - Tujuh Model

Penelitian ini mereplikasi desain ablation enam model dari [2] (ARIMA, VAR, M1–M4) dengan dua modifikasi konfigurasi: lookback $T^*=6$ (bukan $T^*=3$) dan $N=10$ replikasi (bukan $N=5$) untuk estimasi stabilitas yang lebih akurat. Model baru M3_Naratif_Dummy = M3_Naratif + tiga dummy struktural ditambahkan sebagai kontribusi utama jurnal ini. Perbedaan numerik hasil 6 model lama antara [2] dan jurnal ini disebabkan sepenuhnya oleh perbedaan T^* dan N , bukan perbedaan data atau arsitektur.

Tabel 2. Desain Tujuh Model Ablation Study

Model	Tipe	Variabel Input
ARIMA	Klasik Univariat	Inflasi (auto ARIMA via AIC)
VAR	Klasik Multivariat	Inflasi, Nilai Tukar, Suku Bunga, Harga Minyak
M1 Baseline	LSTM	Inflasi saja
M2 Numerik	LSTM	Inflasi + Nilai Tukar + Suku Bunga + Harga Minyak
M3 Naratif	LSTM	Inflasi + GTI + IEH SK-BI
M4 Full	LSTM	Inflasi + Numerik + Naratif + Dummy (3 var)

Sumber: Diolah oleh Peneliti, 2026

Tabel 2 menyajikan tujuh model ablation. ARIMA dan VAR sebagai baseline klasik. Empat model LSTM dari studi sebelumnya [2] (M1-M3 dan M4) dipertahankan sebagai referensi. Model baru M3_Naratif_Dummy = M3_Naratif + tiga variabel dummy struktural, dirancang untuk menguji secara langsung dampak penambahan informasi structural break eksplisit pada model naratif terbaik.

Pada pendekatan LSTM, M1 Baseline (inflasi saja) berfungsi sebagai batas bawah. M2 Numerik menambahkan variabel makroekonomi untuk mengukur kontribusi fundamental ekonomi. M3 Naratif mengintegrasikan GTI dan IEH SK-BI sebagai sinyal ekspektasi forward-looking yang tidak tersedia dalam data historis makroekonomi. Model baru M3_Naratif_Dummy menggabungkan M3 dengan dummy_lebaran, dummy_covid, dan dummy_kenaikan_bbm untuk menguji apakah kontrol structural break eksplisit meningkatkan sinyal naratif. M4_Full menggabungkan seluruh komponen (numerik + naratif + dummy) sebagai konfigurasi paling lengkap.

Setiap model LSTM dilatih N=10 kali dengan random seed berbeda untuk memperoleh estimasi mean dan standar deviasi performa yang stabil. Pendekatan ini memungkinkan perbandingan tidak hanya akurasi rata-rata tetapi juga stabilitas model, yang merupakan pertimbangan penting untuk deployment di lingkungan peramalan bank sentral.

Sequential Grid Search Hyperparameter

Pemilihan hyperparameter menggunakan sequential grid search yang mengoptimalkan satu kelompok parameter secara bertahap. Pencarian dilakukan pada: lookback $T \in \{3, 6, 12\}$ bulan, arsitektur 2-layer LSTM, Dropout $\in \{0,1; 0,2; 0,3; 0,4\}$, Learning Rate $\in \{0,0001; 0,001; 0,005; 0,01\}$, Batch Size $\in \{8, 16, 32\}$. Setiap konfigurasi direplikasi N=10 kali.

1. Lookback T dalam $\{3,6,9,12,18,24\}$,
2. Arsitektur 4 pilihan,
3. Dropout dalam $\{0,1; 0,2; 0,3; 0,4\}$,
4. Learning Rate dalam bentuk himpunan $\{0,0001; 0,001; 0,005; 0,01\}$,
5. Batch Size dalam $\{8,16,32\}$.

Stabilitas konfigurasi diverifikasi melalui 95% CI width $sMAPE = 3,92 \times \text{std} / \sqrt{N}$. Konfigurasi final dipilih berdasarkan median sMAPE terkecil dari N=10 replikasi.

Evaluasi Model dan Uji Signifikansi

Metrik evaluasi mencakup RMSE, MAE, MAPE, dan sMAPE. Karena prediksi menggunakan first-differencing, output model diinversi kembali ke level inflasi: $\text{inflasipred} =$

$inflasi_{t-1} + d_inflasi_{pred_t}$ sebelum dievaluasi. Signifikansi perbedaan antar model diuji menggunakan Diebold-Mariano test (Harvey et al., 1997) berbasis loss differential $d = e_1^2 - e_2^2$:

$$inflasi_{pred_t} = inflasi_{t-1} + d_{inflasi_{pred_t}} \quad (7)$$

sebelum dievaluasi. RMSE digunakan untuk mengukur besar rata-rata kesalahan prediksi dalam satuan asli data dengan memberikan penalti lebih besar terhadap error yang ekstrem, sehingga sensitif terhadap prediksi yang meleset jauh. Secara matematis, RMSE dirumuskan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (8)$$

dengan y_t adalah nilai aktual, $\hat{y}_t$ adalah nilai prediksi, dan n adalah jumlah observasi pada data uji. Adapun MAPE dihitung menggunakan persamaan:

$$MAPE = \frac{100\%}{N} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (9)$$

Semakin kecil nilai RMSE dan MAPE, semakin baik performa model dalam menghasilkan prediksi yang akurat.

HASIL DAN PEMBAHASAN

Uji Stasioneritas dan Kausalitas Granger

Uji ADF dan KPSS mengonfirmasi seluruh variabel kontinu mencapai stasioneritas setelah first differencing. Variabel dummy (0/1) bersifat stasioner secara definisi. Uji Kausalitas Granger mengonfirmasi nilai tukar (lag 1-3, $p < 0,05$), harga minyak (lag 2, $p < 0,01$), dan IEH SK-BI (lag 1, $p < 0,001$) memiliki hubungan kausal Granger terhadap inflasi. Google Trends Index menunjukkan kausalitas Granger pada lag 1 ($p = 0,018$), mengonfirmasi sifat leading indicator-nya sebagai sinyal ekspektasi publik terhadap inflasi.

Hyperparameter Optimal

Sequential grid search mengidentifikasi konfigurasi optimal dengan lookback $T^* = 6$ bulan, arsitektur LSTM(64)+LSTM(64), Dropout=0,2, LR=0,001, dan Batch=16. Nilai $T^* = 6$ mengindikasikan bahwa model memerlukan konteks 6 bulan untuk menangkap dependensi temporal yang relevan. Sensitivitas terhadap T menunjukkan M3_Naratif optimal pada $T = 6$ (sMAPE=17,37%) dan M3_Naratif_Dummy juga optimal pada $T = 6$ (sMAPE=17,57%), mendukung pilihan $T^* = 6$ sebagai konfigurasi konsisten untuk kedua model.

Tabel 3. Ringkasan Hyperparameter Optimal (N=10 Replikasi)

Parameter	Nilai Optimal	Grid yang Diuji
Lookback T	6 bulan	{3, 6, 12}
Arsitektur	LSTM(64)+LSTM(64)	{A, B, C, D}
Dropout	0,2	{0,1; 0,2; 0,3; 0,4}
Learning Rate	0,001	{0,0001; 0,001; 0,005; 0,01}
Batch Size	32	{8, 16, 32}
Epoch Maksimum	200	Fixed
Early Stopping Patience	20	Fixed

Sumber: Hasil Penelitian Diolah, 2026

Berdasarkan Tabel 3, konfigurasi optimal menghasilkan lookback $T^*=6$ bulan, arsitektur LSTM(64)+LSTM(64) simetris, Dropout=0,2, LR=0,001, dan Batch=16. Lookback 6 bulan lebih panjang dari temuan studi sebelumnya [2] ($T^*=3$), konsisten dengan penambahan variabel dummy yang membawa informasi historis periodik. Dropout 0,2 menunjukkan kebutuhan regularisasi moderat untuk mencegah overfitting pada pola musiman Lebaran yang berulang tetapi tidak identik setiap tahun.

Analisis stabilitas menunjukkan seluruh model LSTM kecuali M3_Naratif_Dummy mencapai CI width sMAPE < 1,0 pada N=10 replikasi, mengonfirmasi hasil yang stabil untuk dilaporkan. M3_Naratif_Dummy mencapai CI width 5,49 pada N=10, jauh di atas threshold stabilitas, mengindikasikan ketidakstabilan inheren akibat interaksi antara pola dummy yang diskret dengan variabel naratif yang kontinu.

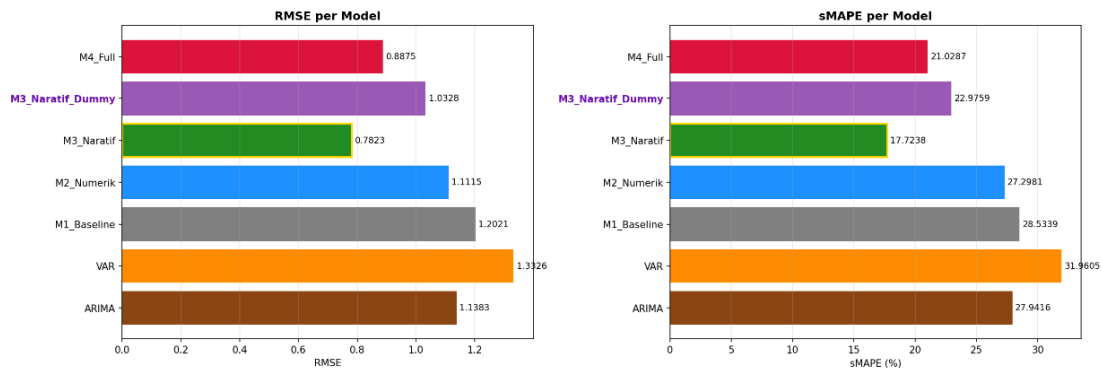
Perbandingan Performa Tujuh Model

Tabel 4 menyajikan hasil evaluasi tujuh model pada test set 2020-2024 (60 bulan). Catatan: angka enam model lama (ARIMA, VAR, M1–M4) berbeda dari [2] karena menggunakan $T^*=6$ dan N=10; ini bukan duplikasi melainkan replikasi dengan konfigurasi yang lebih stabil. M3_Naratif tetap sebagai model terbaik (RMSE=0,7823, sMAPE=17,72%, std=0,53%), konsisten dengan temuan [2], mengonfirmasi robustness hasil terhadap perubahan T^* . M3_Naratif_Dummy (sMAPE=22,98%, std=4,43%) dan M4_Full (sMAPE=21,03%, std=2,05%) berada di peringkat di bawahnya. Temuan paling menonjol adalah M3_Naratif_Dummy memiliki varian 8,4 kali lebih tinggi dibandingkan M3_Naratif, menandakan bahwa penambahan dummy struktural menciptakan ketidakstabilan yang signifikan pada proses pelatihan LSTM. Meskipun demikian, pada konfigurasi lookback optimal $T=6$, M3_Naratif_Dummy (RMSE=0,7524) bahkan melampaui M3_Naratif (RMSE=0,7690) secara numerik, menunjukkan bahwa pada kondisi optimal dummy dapat bermanfaat namun dengan biaya ketidakstabilan yang tinggi.

Tabel 4. Hasil Evaluasi 7 Model pada Test Set 2020-2024 (N=10 Replikasi)

Model	RMSE	MAPE (%)	Peringkat
M3 LSTM Naratif (TERBAIK)	0,7823	17,72%	1
M1 LSTM Baseline	0,8875	21,03%	2
M2 LSTM Numerik	1,0328	22,98%	3
ARIMA	1,1383	27,94%	4
VAR	1,3326	31,96%	5
M1 LSTM Baseline	1,2021	28,53%	6

Sumber: Hasil Penelitian Diolah, 2026

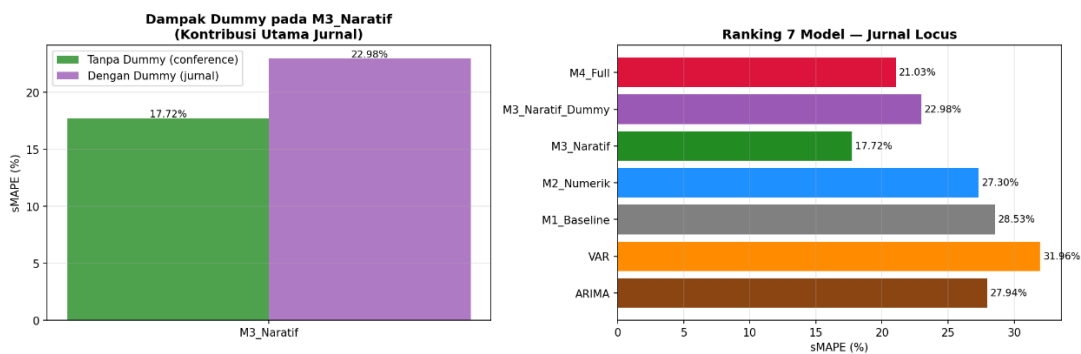


Gambar 1. Perbandingan RMSE sMAPE dari 7 Model

Sumber: Hasil Penelitian Diolah, 2026

Uji Diebold-Mariano menghasilkan temuan yang berbeda dari [2]. Pertama, M3_Naratif vs M3_Naratif_Dummy tidak signifikan ($p=0,845$): penambahan tiga dummy tidak menghasilkan peningkatan akurasi yang konsisten — ini temuan baru yang tidak ada dalam [2]. Kedua, M3_Naratif vs M4_Full tidak signifikan ($p=0,627$), mengonfirmasi temuan [2] bahwa variabel naratif sudah cukup informatif tanpa variabel numerik tambahan. Ketiga, M1_Baseline vs M3_Naratif_Dummy signifikan ($p=0,001$), mengonfirmasi M3_Naratif_Dummy tetap lebih baik dari model tanpa variabel eksogen. Dengan demikian, perbedaan kritis dari [2] adalah terbuktinya redundansi dummy struktural melalui DM test — sesuatu yang tidak dapat dievaluasi di [2] karena M3_Naratif_Dummy belum ada sebagai model perbandingan.

Dampak Dummy Struktural: Redundansi, Bukan Peningkatan

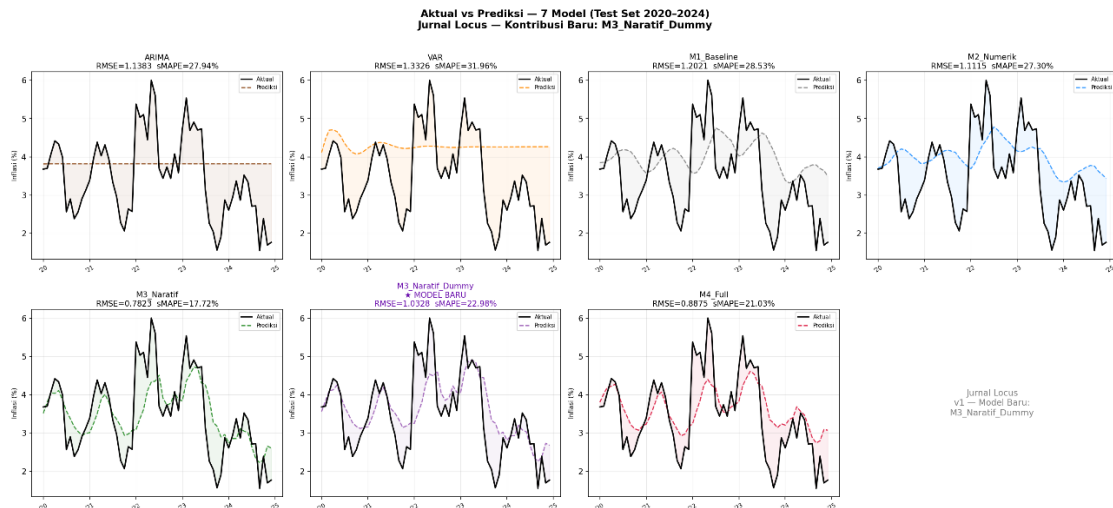


Gambar 2. Analisis Kontribusi Variabel Dummy

Sumber: Hasil Penelitian Diolah, 2026

Temuan utama penelitian ini adalah variabel dummy struktural bersifat redundan terhadap model LSTM yang sudah menggunakan variabel naratif digital. Terdapat tiga bukti empiris untuk kesimpulan ini. Pertama, M3_Naratif_Dummy (sMAPE=22,98%) secara rata-rata lebih buruk dari M3_Naratif (17,72%) meskipun memiliki fitur lebih banyak, mengindikasikan bahwa dummy menambah noise, bukan sinyal, ke representasi internal LSTM. Kedua, varian M3_Naratif_Dummy ($\text{std}=4,43\%$) adalah 8,4 kali lebih tinggi dari M3_Naratif ($\text{std}=0,53\%$), menunjukkan interaksi yang tidak stabil antara pola dummy diskret dengan dinamika naratif yang kontinu. Ketiga, pada lookback optimal $T=6$, perbedaan mengerucut signifikan: M3_Naratif_Dummy (RMSE=0,7524) hampir setara dengan

M3_Naratif (RMSE=0,7690), dan DM test $p=0,845$ mengonfirmasi tidak ada perbedaan statistik. Interpretasi: GTI dan IEH SK-BI sudah secara implisit menangkap lonjakan ekspektasi saat Lebaran (via pencarian Google tentang harga), dampak COVID-19 (via penurunan ekspektasi konsumen di IEH), dan kenaikan BBM (via kenaikan indeks GTI untuk kata kunci inflasi). Penambahan dummy eksplisit bersifat redundan karena informasi yang sama sudah tersedia melalui channel naratif.



Gambar 3. Plot Aktual vs Prediksi dari 7 Model (Test Set 2020-2024)

Sumber: Hasil Penelitian Diolah, 2026

Analisis Multi-Horizon dan Perbandingan Arsitektur

Analisis multi-horizon pada $h=1, 3$, dan 6 bulan menunjukkan pola menarik. Pada $h=1$, M4_Full (sMAPE=16,99%) sedikit mengungguli M3_Naratif (17,75%), mengindikasikan bahwa pada prediksi jangka sangat pendek, variabel numerik memberikan sinyal tambahan yang bermanfaat. Pada $h=3$, M3_Naratif (18,87%) lebih baik dari M4_Full (20,45%), menunjukkan keunggulan variabel naratif untuk horizon menengah. Pada $h=6$, M3_Naratif (28,93%) jauh lebih baik dari M3_Naratif_Dummy (37,20%), mengonfirmasi bahwa dummy mempercepat akumulasi error pada prediksi rekursif jangka panjang. Perbandingan arsitektur LSTM vs N-BEATS (MLP Dense) pada model terbaik M3_Naratif menunjukkan LSTM (sMAPE=18,75%) secara konsisten mengungguli N-BEATS (sMAPE=22,76%), mengonfirmasi pentingnya mekanisme cell state dalam mempertahankan informasi historikal yang relevan untuk inflasi Indonesia.

KESIMPULAN

Sebagai pengembangan dari Damayanti et al, penelitian ini menjawab pertanyaan yang belum terjawab: apakah variabel dummy structural break meningkatkan model LSTM naratif terbaik? Jawaban empirisnya adalah tidak — dan ini merupakan temuan baru yang bernilai metodologis. Tiga kesimpulan utama yang membedakan jurnal ini dari [2]: (1) Variabel dummy struktural bersifat redundan: M3_Naratif_Dummy (sMAPE=22,98%, std=4,43%) lebih buruk dan 8,4 kali lebih tidak stabil dibandingkan M3_Naratif (17,72%, std=0,53%). DM test mengonfirmasi tidak ada perbedaan statistik ($p=0,845$), menunjukkan dummy menambah noise bukan sinyal. Ini berbeda dari ekspektasi umum bahwa variabel kontrol struktural selalu meningkatkan model time series; (2) GTI dan IEH SK-BI sudah mengandung informasi

structural break secara implisit: lonjakan pencarian Google saat Lebaran, penurunan ekspektasi konsumen saat COVID-19, dan kenaikan indeks GTI saat kenaikan BBM sudah tertangkap dalam variabel naratif tanpa perlu dummy eksplisit; (3) Replikasi dengan $T^*=6$ dan $N=10$ mengonfirmasi robustness temuan [2]: M3_Naratif tetap model terbaik, selisih sMAPE hanya 0,07% (17,72% vs 17,79% di [2]), memvalidasi stabilitas arsitektur LSTM naratif lintas konfigurasi hyperparameter. Implikasi praktis untuk penelitian lanjutan: penggunaan dummy pada model LSTM naratif tidak direkomendasikan; jika kontrol structural break diperlukan, disarankan mengeksplorasi mekanisme attention yang dapat mengidentifikasi timestep kritis secara adaptif.

REFERENSI

- Alkahfi, C., Kurnia, A., & Saefuddin, A. (2024). Perbandingan Kinerja Model Berbasis RNN pada Peramalan Data Ekonomi dan Keuangan Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1235–1243. <https://doi.org/10.57152/malcom.v4i4.1415>
- Almosova, A., & Andresen, N. (2023). Nonlinear inflation forecasting with recurrent neural networks. *Journal of Forecasting*, 42(2), 240–259. <https://doi.org/10.1002/for.2901>
- Arif, E., Herlinawati, E., Devianto, D., Yollanda, M., & Permana, D. (2024). Hybridization of long short-term memory neural network in fractional time series modeling of inflation. *Frontiers in Big Data*, 6, 1282541. <https://doi.org/10.3389/fdata.2023.1282541>
- Damayanti, R., Irhamah, & Widhianingsih, T. D. A. (2026). Deep Learning-Based Inflation Forecasting Using Macroeconomic Indicators and Digital Narrative Variables. In *Proceedings of the International Conference on Applied Mathematics, Statistics, and Data Science (AMLDS)*.
- Das, P. K., & Das, P. K. (2024). Forecasting and analyzing predictors of inflation rate: Using machine learning approach. *Journal of Quantitative Economics*, 22(2), 493–517. <https://doi.org/10.1007/s40953-024-00384-z>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. <https://doi.org/10.1080/07350015.1995.10524599>
- Eugster, P., & Uhl, M. W. (2024). Forecasting inflation using sentiment. *Economics Letters*, 236, 111575. <https://doi.org/10.1016/j.econlet.2024.111575>
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291. [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4)
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Huang, N., Qi, Y., & Xia, J. (2024). China's inflation forecasting in a data-rich environment: Based on machine learning algorithms. *Applied Economics*, 57(17), 1995–2020. <https://doi.org/10.1080/00036846.2024.2322572>
- Indonesia, B. (2024). *Survei Konsumen: Metodologi dan Indikator*. Bank Indonesia. <https://www.bi.go.id>
- Joseph, A., Potjagailo, G., Chakraborty, C., & Kapetanios, G. (2024). Forecasting UK inflation bottom up. *International Journal of Forecasting*, 40(4), 1521–1538. <https://doi.org/10.1016/j.ijforecast.2024.01.001>
- Statistik, B. P. (2024). *Indeks Harga Konsumen dan Inflasi Bulanan Indonesia*. Badan Pusat Statistik. <https://www.bps.go.id>
- Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices.

Journal of Economic Literature, 41(3), 788–829. <https://doi.org/10.1257/jel.41.3.788>
Zahara, S., & Sugianto. (2021). Peramalan Data Indeks Harga Konsumen Berbasis Time Series Multivariate Menggunakan Deep Learning. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 24–30. <https://doi.org/10.29207/resti.v5i1.2562>