

Deteksi Anomali Temporal dan Aktivitas Terkoordinasi pada Komentar Produk Parfum Menggunakan Long Short-Term Memory (LSTM)

Agus Irvan Maulana*, Tacbir Hendro Pudjiantoro, Dea Destiani

Universitas Jenderal Achmad Yani, Indonesia

Email: aim.22@si.unjani.ac.id*, tacbir.hendro@lecture.unjani.ac.id,

dea.destiani@lecture.unjani.ac.id

Abstract

The development of social media has led to increased user interaction in the form of comments on product posts. However, high comment activity also has the potential to give rise to inauthentic activity and coordinated comments that can influence users' perception of a product. This research aims to detect temporal anomalies and identify indications of coordinated comment activity on perfume product posts on Instagram. The research object was chosen because perfume product posts have a high level of interaction, which has the potential to cause spikes in comment activity at certain times. The dataset used consists of 448 comments collected through an Instagram scraping process during the period from January 19, 2026 to May 3, 2026. The data were then processed through preprocessing stages, time-series transformation, and text representation using the IndoBERT model. Long Short-Term Memory (LSTM) method was used to learn the temporal patterns of comments and detect anomalies based on residual values and an anomaly threshold, while Cosine Similarity was used to measure the degree of similarity between comments at anomalous time points. The model evaluation results show an MAE of 0.3318, an RMSE of 1.9483, and a MAPE of 16.37%, indicating that the model's predictive performance falls into the good category. The anomaly detection process successfully identified four anomalous time points with significant spikes in comment activity. The cosine similarity analysis revealed groups of comments with a high degree of textual similarity, including identical comments and repeated use of certain words. This research contributes to the fields of social media analytics and anomaly detection by integrating temporal analysis and text similarity analysis to identify abnormal comment activity patterns and indications of coordinated comment activity on social media.

Keywords: Bidirectional LSTM; Anomaly Detection; Cosine Similarity; IndoBERT; Instagram Comments.

Abstrak

Perkembangan media sosial menyebabkan peningkatan interaksi pengguna dalam bentuk komentar pada unggahan produk. Namun, tingginya aktivitas komentar juga berpotensi memunculkan aktivitas tidak autentik dan komentar terkoordinasi yang dapat memengaruhi persepsi pengguna terhadap suatu produk. Penelitian ini bertujuan untuk mendeteksi anomali temporal dan mengidentifikasi indikasi aktivitas komentar terkoordinasi pada unggahan produk parfum di Instagram. Objek penelitian dipilih karena unggahan produk parfum memiliki tingkat interaksi yang tinggi sehingga berpotensi menimbulkan lonjakan aktivitas komentar pada waktu tertentu. Dataset yang digunakan terdiri dari 448 komentar yang dikumpulkan melalui proses scraping Instagram pada periode 19 Januari 2026 hingga 3 Mei 2026. Data kemudian diproses melalui tahapan *preprocessing*, transformasi *time-series*, serta representasi teks menggunakan model IndoBERT. Metode *Long Short-Term Memory* (LSTM) digunakan untuk mempelajari pola temporal komentar dan mendeteksi anomali berdasarkan nilai residual dan *threshold anomaly*, sedangkan *Cosine Similarity* digunakan untuk mengukur tingkat kemiripan antar komentar pada waktu anomali. Hasil evaluasi model menunjukkan nilai MAE sebesar 0,3318, RMSE sebesar 1,9483, dan MAPE sebesar 16,37%, yang menunjukkan kemampuan prediksi model berada pada kategori baik. Proses deteksi anomali berhasil menemukan empat waktu anomali dengan lonjakan komentar yang signifikan. Analisis cosine similarity menunjukkan adanya kelompok komentar dengan tingkat kemiripan teks yang tinggi, termasuk komentar identik dan pengulangan kata tertentu secara berulang. Penelitian ini memberikan kontribusi pada bidang *social media analytics* dan *anomaly detection* melalui integrasi analisis temporal dan analisis kemiripan teks untuk mengidentifikasi pola aktivitas komentar yang tidak normal serta indikasi aktivitas komentar terkoordinasi pada media sosial.

Kata Kunci: Bidirectional LSTM; Deteksi Anomali; Cosine Similarity; IndoBERT; Komentar Instagram.

Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah mendorong pertumbuhan penggunaan media sosial secara signifikan dalam beberapa tahun terakhir. Media sosial tidak lagi hanya berfungsi sebagai sarana komunikasi, tetapi juga menjadi platform utama untuk berbagi informasi, membangun komunitas, serta mendukung aktivitas pemasaran digital (Permadi, 2022; Rachmad, 2024). Berdasarkan laporan Digital 2025, jumlah pengguna media sosial di dunia telah mencapai lebih dari 5 miliar pengguna dengan rata-rata waktu penggunaan lebih dari dua jam per hari. Tingginya aktivitas pengguna tersebut menghasilkan volume data yang sangat besar dalam bentuk unggahan, komentar, ulasan, maupun interaksi lainnya yang dapat dimanfaatkan untuk berbagai keperluan analisis.

Dalam konteks pemasaran digital, media sosial telah menjadi salah satu saluran utama bagi perusahaan untuk memperkenalkan produk dan berinteraksi dengan konsumen. Komentar dan ulasan yang diberikan pengguna pada suatu produk sering kali menjadi sumber informasi yang lebih dipercaya dibandingkan iklan konvensional. Fenomena ini dikenal sebagai *electronic word-of-mouth* (eWOM), yaitu penyebaran opini dan pengalaman konsumen secara daring yang dapat memengaruhi persepsi, sikap, dan keputusan pembelian pengguna lainnya (Ismagilova et al., 2021; Ngo et al., 2024). Oleh karena itu, kualitas dan keaslian komentar yang muncul pada media sosial menjadi faktor penting dalam membangun kepercayaan konsumen terhadap suatu produk.

Salah satu industri yang sangat bergantung pada interaksi pengguna di media sosial adalah industri parfum. Dalam beberapa tahun terakhir, industri parfum mengalami pertumbuhan yang pesat seiring meningkatnya tren belanja daring dan pemasaran berbasis media sosial. Platform seperti Instagram menjadi media promosi yang efektif karena memungkinkan produsen parfum untuk menampilkan konten visual yang menarik sekaligus menerima umpan balik langsung dari konsumen (Alfiaturrakhmania & Fakhri, 2025). Sebelum membeli parfum, calon konsumen sering kali membaca komentar pengguna lain untuk memperoleh informasi mengenai aroma, ketahanan, kualitas, dan pengalaman penggunaan produk. Dengan demikian, komentar pada unggahan produk parfum memiliki peran penting dalam membentuk persepsi dan keputusan pembelian konsumen.

Namun, tingginya pengaruh komentar pengguna juga memunculkan berbagai tantangan terkait keaslian informasi yang beredar di media sosial. Sejumlah pihak memanfaatkan fitur komentar untuk melakukan manipulasi opini melalui spam, penggunaan akun otomatis (*bot*), maupun aktivitas komentar yang dilakukan secara terkoordinasi (Akhtar et al., 2024). Aktivitas tersebut bertujuan untuk meningkatkan atau menurunkan citra suatu produk secara artifisial sehingga dapat memengaruhi persepsi konsumen. Komentar yang tidak autentik sering kali dibuat menyerupai perilaku pengguna normal sehingga sulit dibedakan secara manual (Wani et al., 2024). Kondisi ini tidak hanya berpotensi menyesatkan konsumen, tetapi juga dapat menciptakan persaingan yang tidak sehat serta menurunkan tingkat kepercayaan terhadap platform digital.

Salah satu karakteristik aktivitas komentar yang tidak autentik adalah munculnya lonjakan komentar dalam waktu tertentu yang tidak sesuai dengan pola aktivitas normal pengguna (Lee et al., 2024). Fenomena tersebut dikenal sebagai anomali temporal, yaitu kondisi ketika suatu aktivitas menunjukkan penyimpangan dari pola historis yang umumnya

terjadi (Zamanzadeh Darban et al., 2024). Pada media sosial, anomali temporal dapat berupa peningkatan jumlah komentar secara mendadak dalam waktu singkat, distribusi komentar yang tidak wajar, maupun aktivitas berulang yang terjadi secara bersamaan (Celik et al., 2025). Analisis terhadap pola temporal menjadi penting karena dapat digunakan sebagai indikator awal untuk mendeteksi aktivitas yang mencurigakan.

Deteksi anomali temporal pada data deret waktu (*time-series*) telah banyak dikembangkan menggunakan pendekatan *machine learning* dan *deep learning*. Salah satu metode yang banyak digunakan adalah *Long Short-Term Memory* (LSTM), yaitu pengembangan dari *Recurrent Neural Network* (RNN) yang dirancang untuk mempelajari hubungan jangka panjang pada data berurutan (Lindemann et al., 2021; Ngo et al., 2024). Kemampuan LSTM dalam memahami pola temporal membuat metode ini banyak diterapkan pada berbagai bidang seperti prediksi lalu lintas jaringan, analisis aktivitas pengguna, deteksi penipuan, dan deteksi anomali. Pada penelitian ini digunakan *Long Short-Term Memory* (LSTM), yang mampu mempelajari pola temporal dan hubungan jangka panjang pada data deret waktu sehingga efektif dalam mendeteksi penyimpangan pola aktivitas komentar (Jiang et al., 2023).

Meskipun analisis temporal mampu mengidentifikasi kapan suatu aktivitas yang tidak biasa terjadi, pendekatan tersebut belum cukup untuk menjelaskan apakah aktivitas tersebut merupakan hasil koordinasi antar pengguna (Pacheco et al., 2021). Dalam konteks komentar media sosial, aktivitas terkoordinasi umumnya ditandai dengan kemunculan komentar yang memiliki kemiripan isi, struktur kalimat, maupun konteks pembahasan yang sama (Song et al., 2026). Oleh karena itu, diperlukan analisis tambahan untuk mengukur tingkat kemiripan antar komentar.

Salah satu metode yang umum digunakan untuk mengukur kemiripan teks adalah *cosine similarity*. Metode ini bekerja dengan menghitung sudut antara dua representasi vektor sehingga dapat menunjukkan tingkat kesamaan semantik antar dokumen atau kalimat (Han et al., 2021; Iskandar & Kurniawati, 2025). Untuk memperoleh representasi teks yang mampu menangkap makna secara lebih baik, penelitian ini memanfaatkan IndoBERT sebagai model *language representation*. IndoBERT merupakan model berbasis Transformer yang telah dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar sehingga mampu menghasilkan representasi teks yang kaya secara semantik (Abidin & Pamungkas, 2025; Song et al., 2026). Dengan menggabungkan IndoBERT dan *cosine similarity*, kemiripan antar komentar dapat diukur secara lebih akurat dibandingkan metode berbasis pencocokan kata sederhana (Maulidya Prastita Syah et al., 2025).

Penelitian terdahulu telah mengkaji deteksi anomali menggunakan berbagai pendekatan. Usmani et al. (2024) melakukan kajian komprehensif terhadap metode *deep learning* untuk deteksi anomali pada data *time-series* dan merekomendasikan penggunaan model *hybrid* serta *attention mechanism*. Namun penelitian tersebut hanya bersifat tinjauan literatur dan belum mengimplementasikan model pada kasus nyata aktivitas media sosial. Silalahi et al. (2023) menerapkan LSTM dan Transformer untuk mendeteksi anomali pada data log sistem berbasis urutan waktu, tetapi tidak mempertimbangkan kemiripan isi data yang dianalisis. Sementara itu, Becker (2025) menggunakan kombinasi *Graph Neural Network* dan GRU untuk mendeteksi aktivitas penipuan pada *platform e-commerce*. Pendekatan tersebut mampu menghasilkan performa yang baik, namun memerlukan pemodelan relasi antar entitas

yang kompleks sehingga kurang sesuai untuk analisis komentar media sosial yang tidak memiliki struktur relasi pengguna yang lengkap.

Berdasarkan kajian penelitian terdahulu, terdapat kesenjangan penelitian (*research gap*) yang menunjukkan bahwa sebagian besar penelitian hanya berfokus pada salah satu aspek, yaitu analisis temporal atau analisis teks. Pendekatan tersebut menyebabkan kemampuan sistem dalam mendeteksi aktivitas komentar tidak autentik menjadi kurang optimal. Padahal, aktivitas komentar terkoordinasi umumnya ditandai oleh dua karakteristik utama, yaitu lonjakan aktivitas pada waktu tertentu dan kemiripan isi komentar yang muncul secara bersamaan. Oleh karena itu, diperlukan pendekatan yang mampu mengintegrasikan kedua aspek tersebut dalam satu kerangka analisis.

Berdasarkan permasalahan dan kesenjangan penelitian tersebut, penelitian ini mengusulkan pendekatan deteksi anomali temporal dan aktivitas komentar terkoordinasi pada komentar produk parfum di Instagram menggunakan kombinasi *Long Short-Term Memory* (LSTM) dan *cosine similarity* berbasis IndoBERT. LSTM digunakan untuk mendeteksi anomali temporal berdasarkan pola distribusi waktu komentar, sedangkan *cosine similarity* digunakan untuk mengukur tingkat kemiripan antar komentar yang muncul pada waktu anomali. Kombinasi kedua metode tersebut diharapkan mampu memberikan kemampuan deteksi yang lebih komprehensif dalam mengidentifikasi aktivitas komentar yang terindikasi tidak autentik.

Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis pola temporal aktivitas komentar pada unggahan produk parfum di Instagram menggunakan metode *Long Short-Term Memory* (LSTM). Melalui analisis tersebut, penelitian ini berupaya mendeteksi anomali temporal yang ditandai dengan adanya lonjakan aktivitas komentar yang menyimpang dari pola normal. Selain itu, penelitian ini juga bertujuan mengukur tingkat kemiripan antar komentar menggunakan metode *cosine similarity* berbasis representasi teks IndoBERT guna mengidentifikasi kesamaan konteks dan isi komentar. Hasil analisis temporal dan kemiripan teks kemudian dikombinasikan untuk mengidentifikasi indikasi aktivitas komentar terkoordinasi pada media sosial Instagram. Selanjutnya, kinerja model yang diusulkan dievaluasi menggunakan metrik *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE) untuk mengetahui kemampuan model dalam mempelajari pola temporal serta mendeteksi anomali pada data komentar.

Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat teoritis maupun praktis. Secara teoritis, penelitian ini berkontribusi pada pengembangan bidang *social media analytics*, *anomaly detection*, dan *natural language processing* melalui penerapan metode LSTM dan *cosine similarity* pada data komentar media sosial. Secara praktis, penelitian ini dapat membantu pelaku usaha, pengelola media sosial, dan peneliti dalam mengidentifikasi aktivitas komentar yang tidak normal sehingga dapat mendukung pengambilan keputusan yang lebih tepat dalam pengelolaan reputasi digital.

Implikasi Penelitian

Penelitian ini memiliki implikasi terhadap pengembangan metode analisis media sosial yang mampu mendeteksi aktivitas tidak autentik berdasarkan aspek temporal dan semantik secara bersamaan. Hasil penelitian dapat digunakan sebagai dasar pengembangan sistem

monitoring otomatis untuk mendeteksi spam, *bot*, maupun aktivitas komentar terkoordinasi pada berbagai platform media sosial. Selain itu, penelitian ini juga dapat mendukung strategi pemasaran digital, pengawasan reputasi merek, serta pengembangan sistem deteksi anomali pada lingkungan media sosial secara *real-time*.

Metode Penelitian

Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan model development yang bertujuan untuk mendeteksi anomali temporal dan aktivitas komentar terkoordinasi pada komentar produk parfum di media sosial Instagram. Pendekatan kuantitatif digunakan karena penelitian berfokus pada pengolahan data numerik, pengukuran performa model, serta analisis pola komentar berdasarkan distribusi waktu dan kemiripan teks. Pengembangan model dilakukan dengan mengintegrasikan metode *Long Short-Term Memory* (LSTM) untuk deteksi anomali temporal dan *Cosine Similarity* berbasis IndoBERT untuk mengidentifikasi kemiripan komentar.

Desain Penelitian

Penelitian diawali dengan pengumpulan data komentar Instagram menggunakan teknik web scraping. Data yang diperoleh kemudian melalui tahap *preprocessing* yang meliputi *cleaning* data, *case folding*, normalisasi kata, *tokenizing*, dan *stemming*. Selanjutnya data timestamp diubah menjadi bentuk time-series untuk proses pelatihan model *Long Short-Term Memory* (LSTM). Hasil prediksi model digunakan untuk menghitung residual dan menentukan anomali temporal berdasarkan nilai threshold. Komentar yang berada pada waktu anomali kemudian dianalisis menggunakan IndoBERT dan *Cosine Similarity* untuk mengidentifikasi tingkat kemiripan teks. Hasil analisis temporal dan analisis kemiripan teks diintegrasikan untuk mengidentifikasi indikasi aktivitas komentar yang tidak autentik atau terkoordinasi.

Sumber Data

Sumber data penelitian berasal dari komentar pengguna pada unggahan produk parfum di platform Instagram. Data diperoleh melalui proses *web scraping* yang menghasilkan atribut utama berupa teks komentar, nama pengguna, dan *timestamp* komentar. Pemilihan objek penelitian didasarkan pada tingginya tingkat interaksi pengguna pada unggahan produk parfum yang memungkinkan munculnya pola komentar berulang maupun lonjakan aktivitas komentar dalam waktu tertentu. Data yang terkumpul digunakan sebagai dasar analisis temporal dan analisis kemiripan teks untuk mendeteksi aktivitas komentar yang tidak normal.

Teknik Pengumpulan Data

Teknik pengumpulan data dilakukan menggunakan metode *web scraping* pada unggahan produk parfum di Instagram. Data yang dikumpulkan meliputi teks komentar, username, dan waktu publikasi komentar (*timestamp*). Seluruh data kemudian disimpan dalam format tabular untuk memudahkan proses preprocessing dan analisis lebih lanjut.

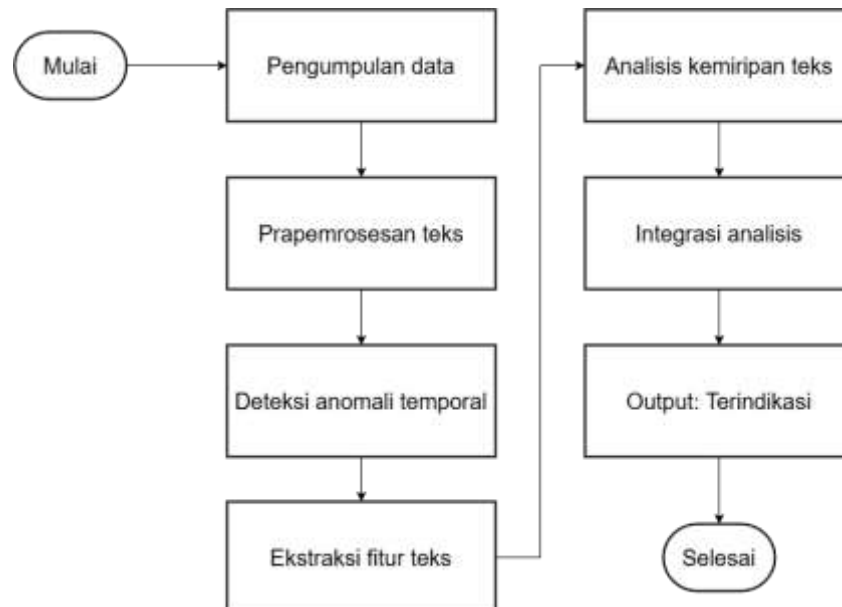
Instrumen Penelitian

Instrumen penelitian yang digunakan berupa perangkat lunak dan pustaka pemrograman Python, yaitu:

1. TensorFlow dan Keras untuk membangun dan melatih model LSTM.
2. Transformers (Hugging Face) untuk implementasi model IndoBERT.
3. Scikit-learn untuk perhitungan *cosine similarity* dan evaluasi model.

4. Pandas untuk pengolahan dan manipulasi data.
5. NLTK untuk proses *preprocessing* teks.
6. Matplotlib untuk visualisasi data dan hasil deteksi anomali.
7. Google Colab/Jupyter Notebook sebagai lingkungan pengembangan dan eksekusi program.

Teknik Analisis Data



Gambar 1. Teknik Analisis Data

Teknik analisis data dalam penelitian ini terdiri dari beberapa tahapan sebagai berikut:

1. *Preprocessing Teks*, meliputi *cleaning data*, *case folding*, normalisasi kata, *tokenizing*, dan *stemming*.
2. Analisis Temporal, yaitu mengubah data timestamp menjadi deret waktu (*time-series*) untuk mengetahui pola distribusi komentar.
3. Pelatihan Model LSTM, untuk mempelajari pola temporal normal dan menghasilkan prediksi jumlah komentar.
4. Deteksi Anomali Temporal, dilakukan menggunakan nilai *residual* dan *threshold anomaly* yang dihitung dari distribusi *error* prediksi model.
5. Analisis Kemiripan Teks, menggunakan IndoBERT untuk menghasilkan representasi teks dan *Cosine Similarity* untuk mengukur tingkat kemiripan antar komentar.
6. Evaluasi Model, menggunakan metrik Mean Absolute Error (MAE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE).
7. Interpretasi Hasil, dengan menggabungkan hasil deteksi anomali temporal dan analisis kemiripan teks untuk mengidentifikasi indikasi aktivitas komentar yang terkoordinasi.

Hasil dan Pembahasan

1. Deskripsi Dataset

Penelitian ini menggunakan data komentar produk parfum dari platform Instagram yang terdiri dari teks komentar dan informasi waktu (*timestamp*). Data dikumpulkan menggunakan teknik *scraping* berbasis Python pada unggahan produk parfum yang bersifat publik. Dataset

yang diperoleh kemudian digunakan sebagai data utama dalam proses analisis temporal dan analisis teks.

2. Pengumpulan Data

Proses pengumpulan data dilakukan menggunakan teknik *scraping* untuk mengambil komentar dan *timestamp* secara otomatis dari Instagram. Data yang berhasil dikumpulkan kemudian disimpan dalam bentuk terstruktur dan digunakan pada tahap preprocessing serta analisis selanjutnya.

Tabel 1. Contoh data hasil Scraping

<i>Timestamp</i>	<i>Ulasan</i>
2026-02-27 08:30:28+0000	Wow penasaran banget sama caramel di middle notes nya 😊
2026-02-27 08:24:46+0000	penasaran sama twist leather, fruity, sweet, fresh, & floral nya 😊
2026-02-27 11:37:13+0000	Kebayang wanginya bakal Segar & juicy di awal, manis elegan di tengah, lalu menjadi woody hangat yang maskulin. Cocok buat lebaran nanti nih

3. Preprocessing Data

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data sebelum dianalisis lebih lanjut. Proses yang dilakukan meliputi:

1. *Cleaning data* untuk menghapus URL, emoji, simbol, dan karakter yang tidak diperlukan.
2. *Case folding* untuk menyeragamkan huruf menjadi *lowercase*.
3. Normalisasi kata untuk mengubah kata tidak baku menjadi kata baku.
4. *Tokenizing* untuk memecah teks menjadi token kata.
5. *Stemming* menggunakan Sastrawi untuk mengubah kata menjadi bentuk dasar.

Hasil preprocessing menunjukkan bahwa data menjadi lebih bersih, konsisten, dan siap digunakan pada tahap analisis temporal serta analisis kemiripan teks.

Tabel 2. Data siap pakai

<i>Timestamp</i>	<i>Ulasan</i>
2026-02-27 08:30:28+0000	wow penasaran banget sama caramel di middle notes
2026-02-27 08:24:46+0000	penasaran sama twist leather fruity sweet fresh floral
2026-02-27 11:37:13+0000	kebayang wangi bakal segar juicy di awal manis elegan di tengah lalu menjadi woody hangat yang maskulin cocok buat lebaran nanti

4. Pelatihan Model LSTM

Pelatihan model *Long Short-Term Memory* (LSTM) dilakukan menggunakan data *time-series* berdasarkan distribusi waktu komentar untuk mempelajari pola aktivitas normal pada data. Model LSTM dipilih karena mampu mempelajari ketergantungan temporal jangka pendek maupun jangka panjang pada data berurutan, sehingga efektif digunakan untuk memodelkan pola distribusi komentar berdasarkan waktu. Data yang telah diproses kemudian dibagi menjadi data latih dan data uji sebelum proses pelatihan dilakukan.

Proses pelatihan model dipantau melalui kurva *training loss* dan *validation loss* selama 100 *epoch*. Grafik pada Gambar 2 menunjukkan bahwa *training loss* (biru) terus menurun dan stabil mendekati 0,0008, sementara *validation loss* (oranye) juga konvergen mendekati

0,00174. Kedua kurva tidak menunjukkan perbedaan yang signifikan, sehingga dapat disimpulkan bahwa model tidak mengalami *overfitting* dan memiliki kemampuan generalisasi yang baik terhadap data baru.



Gambar 2. Kurva *Training Loss* dan *Validation Loss* Model LSTM

Berdasarkan hasil pelatihan, model memperoleh nilai *Mean Absolute Error* (MAE) sebesar 0,3318, *Root Mean Square Error* (RMSE) sebesar 1,9483, dan *Mean Absolute Percentage Error* (MAPE) sebesar 16,37%. Nilai-nilai tersebut menunjukkan bahwa model mampu mengikuti pola temporal komentar dengan tingkat kesalahan prediksi yang masih dapat diterima. Selain itu, diperoleh nilai *threshold residual* sebesar 4,0511 yang digunakan sebagai batas untuk menentukan aktivitas anomali. Nilai residual yang melebihi batas tersebut diidentifikasi sebagai aktivitas komentar yang menyimpang dari pola normal. Rekapitulasi metrik evaluasi model ditampilkan pada Tabel 3.

Tabel 3. Hasil Evaluasi Model LSTM

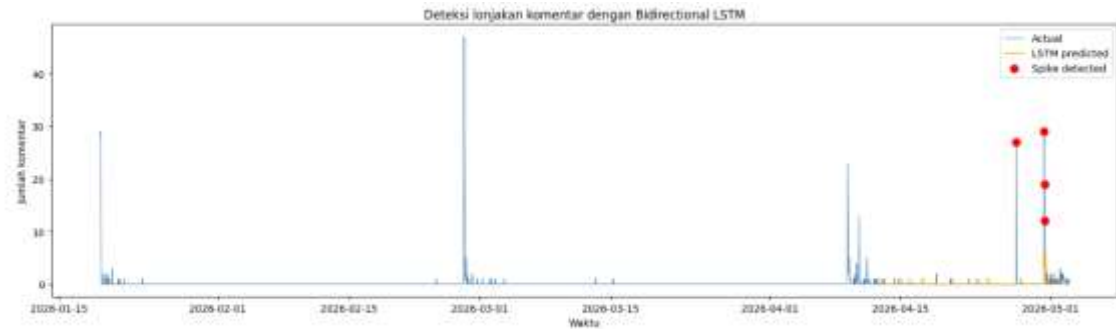
Metrik	Nilai
<i>Mean Absolute Error</i> (MAE)	0,3318
<i>Root Mean Square Error</i> (RMSE)	1,9483
<i>Mean Absolute Percentage Error</i> (MAPE)	16,37%
<i>Threshold Residual</i>	4,0511

Nilai MAPE sebesar 16,37% tergolong dalam kategori *Good Forecast* berdasarkan skala interpretasi, di mana nilai MAPE di bawah 20% dianggap sebagai prediksi yang baik. Hal ini menunjukkan bahwa model LSTM cukup andal dalam mempelajari pola distribusi temporal komentar Instagram pada data produk parfum yang dianalisis.

Berdasarkan *threshold residual* tersebut, ditemukan sejumlah jam anomali yang ditandai oleh lonjakan jumlah komentar secara tidak wajar. Hasil deteksi anomali temporal ditampilkan pada Tabel 4 dan Gambar 3 berikut.

Tabel 4. Hasil Deteksi Anomali Temporal

No	Waktu Anomali	Jumlah Komentar
1	2026-04-27 09:00	27 komentar
2	2026-04-30 08:00	29 komentar
3	2026-04-30 09:00	19 komentar
4	2026-04-30 10:00	12 komentar



Gambar 3. Deteksi Anomali Komentar Berbasis LSTM

Hasil deteksi menunjukkan bahwa terdapat empat periode waktu anomali, yaitu pada tanggal 27 April dan 30 April 2026. Lonjakan komentar terbesar terjadi pada pukul 08.00 tanggal 30 April 2026 dengan jumlah 29 komentar. Selain itu, terdeteksi lonjakan sebanyak 27 komentar pada pukul 09.00 tanggal 27 April 2026, serta 19 komentar dan 12 komentar masing-masing pada pukul 09.00 dan 10.00 tanggal 30 April 2026. Pola lonjakan yang terkonsentrasi pada rentang waktu yang berdekatan ini mengindikasikan adanya aktivitas komentar yang tidak mengikuti pola normal, sehingga perlu dianalisis lebih lanjut melalui analisis kemiripan teks untuk mengidentifikasi kemungkinan adanya aktivitas komentar terkoordinasi.

5. Analisis Kemiripan Teks Menggunakan *Cosine Similarity*

Analisis kemiripan teks dilakukan untuk mengidentifikasi indikasi aktivitas komentar terkoordinasi pada jam yang sebelumnya terdeteksi sebagai anomali temporal oleh model LSTM. Pada tahap ini digunakan model IndoBERT (*indobenchmark/indobert-base-pl*) untuk mengubah komentar menjadi representasi vektor (*text embedding*) sehingga hubungan semantik antar komentar dapat dianalisis dengan lebih akurat.

Berdasarkan hasil deteksi anomali temporal, diperoleh sebanyak 87 komentar yang kemudian dianalisis menggunakan model IndoBERT. Selanjutnya dilakukan perhitungan *cosine similarity* untuk mengukur tingkat kemiripan antar komentar berdasarkan konteks dan struktur teks yang dimiliki.

Hasil perhitungan menunjukkan bahwa matriks *cosine similarity* memiliki nilai minimum sebesar 0,1424, nilai maksimum sebesar 1,0000, dan rata-rata *similarity* sebesar 0,5378. Nilai *similarity* yang mendekati 1 menunjukkan bahwa komentar memiliki struktur dan makna yang sangat mirip, sedangkan nilai yang mendekati 0 menunjukkan komentar yang berbeda secara semantik.

Setelah proses perhitungan *cosine similarity* dilakukan, diperoleh beberapa kelompok komentar yang memiliki tingkat kemiripan teks yang tinggi. Kemiripan tersebut menunjukkan adanya pola komentar berulang dengan konteks dan isi yang serupa pada waktu terjadinya anomali temporal. Hasil analisis menunjukkan bahwa sebagian komentar memiliki nilai *cosine similarity* mendekati 1, yang menandakan bahwa komentar-komentar tersebut sangat mirip secara semantik meskipun ditulis oleh akun yang berbeda.

Tabel 5. Hasil Analisis *Cosine Similarity* pada Periode Anomali

Jumlah Komentar	Nilai <i>Cosine Similarity</i>	Komentar
14 komentar	1,0000	ketahanannya kurang banget, baru 1 jam sudah hilang
13 komentar	0,9912	kurang worth it untuk harganya

37 komentar	0,8618	inception
2 komentar	0,7576	percuma di tampil in terus tapi ga ready

Berdasarkan Tabel 5, ditemukan empat kelompok komentar dengan tingkat kemiripan tinggi. Sebanyak 14 komentar memiliki nilai *cosine similarity* 1,0000 dengan konten yang identik, mengindikasikan bahwa komentar tersebut diunggah secara massal atau disalin dari satu sumber yang sama. Sebanyak 13 komentar lainnya memiliki similarity 0,9912 dengan pola negatif berulang terkait nilai produk. Selain itu, ditemukan 37 komentar dengan similarity 0,8618 yang didominasi kata "inception", yang kemungkinan merupakan komentar tanpa konteks produk yang relevan. Temuan ini secara keseluruhan mengindikasikan adanya aktivitas komentar terkoordinasi (*coordinated inauthentic behavior*) pada produk parfum yang dianalisis.

Dari perspektif bisnis, pola komentar terkoordinasi seperti ini dapat memberikan dampak negatif berupa penurunan kepercayaan konsumen jika ulasan negatif terkoordinasi tersebut tidak terdeteksi. Sebaliknya, deteksi dini terhadap aktivitas semacam ini dapat membantu pemilik merek maupun *platform* untuk mengambil tindakan moderasi secara tepat waktu, melindungi reputasi produk, serta menjaga integritas ulasan yang terlihat oleh calon konsumen.

Hasil penelitian ini sejalan dengan berbagai penelitian sebelumnya yang menunjukkan bahwa metode berbasis *deep learning* mampu mempelajari pola temporal dan mendeteksi anomali pada data deret waktu secara efektif. Penelitian terdahulu umumnya berfokus pada penggunaan LSTM, Transformer, maupun pendekatan berbasis *deep learning* lainnya untuk mendeteksi penyimpangan pola aktivitas berdasarkan dimensi waktu. Namun, sebagian besar penelitian tersebut hanya menitikberatkan pada deteksi anomali temporal tanpa melakukan analisis terhadap kemiripan isi komentar yang muncul pada waktu anomali.

Selain itu, terdapat penelitian yang menggunakan pendekatan berbasis *Graph Neural Network* untuk mendeteksi aktivitas tidak autentik dan *fake review* dengan memanfaatkan hubungan antar pengguna, produk, dan transaksi. Meskipun memberikan hasil yang baik, pendekatan tersebut memerlukan data relasi yang kompleks dan tidak selalu tersedia pada platform media sosial. Berbeda dengan penelitian sebelumnya, penelitian ini menggabungkan deteksi anomali temporal menggunakan LSTM dan analisis kemiripan teks menggunakan *cosine similarity* berbasis IndoBERT sehingga mampu mengidentifikasi lonjakan aktivitas komentar sekaligus menemukan pola komentar berulang yang mengindikasikan aktivitas terkoordinasi tanpa memerlukan pemodelan jaringan antar pengguna. Pendekatan ini memberikan alternatif yang lebih sederhana dan mudah diterapkan pada data komentar media sosial yang hanya memiliki informasi teks dan waktu unggahan.

Kesimpulan

Berdasarkan hasil penelitian, metode *Long Short-Term Memory* (LSTM) mampu mempelajari pola temporal aktivitas komentar pada unggahan produk parfum di Instagram dengan baik. Hal ini ditunjukkan oleh nilai evaluasi model yang terdiri atas *Mean Absolute Error* (MAE) sebesar 0,3318, *Root Mean Square Error* (RMSE) sebesar 1,9483, dan *Mean Absolute Percentage Error* (MAPE) sebesar 16,37%. Model berhasil mendeteksi empat periode waktu anomali berdasarkan nilai *threshold residual*, sedangkan analisis kemiripan teks menggunakan IndoBERT dan *cosine similarity* menunjukkan adanya kelompok komentar

dengan tingkat kemiripan semantik yang tinggi. Hasil tersebut mengindikasikan bahwa kombinasi kedua metode mampu mengidentifikasi lonjakan aktivitas komentar sekaligus memberikan indikasi adanya aktivitas komentar terkoordinasi pada media sosial Instagram.

Penelitian ini memberikan kontribusi dalam pengembangan metode deteksi aktivitas komentar tidak autentik melalui integrasi analisis temporal dan analisis kemiripan teks dalam satu kerangka kerja. Secara praktis, pendekatan yang diusulkan berpotensi mendukung pelaku bisnis, pengelola media sosial, dan pengembang sistem moderasi dalam mendeteksi aktivitas komentar yang mencurigakan serta menjaga kredibilitas interaksi pengguna. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan beragam, melibatkan berbagai platform media sosial, serta mengembangkan sistem deteksi otomatis secara *real-time* untuk meningkatkan akurasi dan efektivitas identifikasi aktivitas komentar terkoordinasi.

Daftar Pustaka

- Abidin, M. I., & Pamungkas, E. W. (2025). Analisis Sentimen Terhadap Timnas Indonesia Di Piala Asia 2023 Dengan Model Transformer Berbahasa Indonesia. *Rabit: Jurnal Teknologi Dan Sistem Informasi Univrab*, 10(2). <https://doi.org/10.36341/rabit.v10i2.6142>
- Akhtar, M. M., Masood, R., Ikram, M., & Kanhere, S. S. (2024). SoK: False Information, Bots and Malicious Campaigns: Demystifying Elements of Social Media Manipulations. *ACM AsiaCCS 2024 - Proceedings of the 19th ACM Asia Conference on Computer and Communications Security*. <https://doi.org/10.1145/3634737.3644998>
- Alfiaturrekhamania, A., & Fakhri, E. A. (2025). The Role of Social Media in Building Brand Image: The Influence of Storytelling Marketing and User-Generated Content on Hmns Perfume. *Jurnal Manajemen Pemasaran Dan Perilaku Konsumen*, 4(4).
- Becker, L. (2025). Anomaly Detection in E-Commerce Platforms via Graph Neural Networks. *Asia Pacific Economic and Management Review*, 2(2). <https://doi.org/10.62177/apemr.v2i2.208>
- Celik, M., Dokuz, A. S., Ecemis, A., & Erdogmus, E. (2025). Discovering temporal, spatial, and contextual anomalous social activities from streaming social media datasets. *Engineering Science and Technology, an International Journal*, 64. <https://doi.org/10.1016/j.jestch.2025.102006>
- Han, M., Zhang, X., Yuan, X., Jiang, J., Yun, W., & Gao, C. (2021). A survey on the techniques, applications, and performance of short text semantic similarity. *Concurrency and Computation: Practice and Experience*, 33(5). <https://doi.org/10.1002/cpe.5971>
- Iskandar, D., & Kurniawati, A. (2025). Analisis Perbandingan Teknik Word2vec dan Doc2vec dalam Mengukur Kemiripan Dokumen Menggunakan Cosine Similarity. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 12(1). <https://doi.org/10.25126/jtiik.2025129143>
- Ismagilova, E., Rana, N. P., Slade, E. L., & Dwivedi, Y. K. (2021). A meta-analysis of the factors affecting eWOM providing behaviour. *European Journal of Marketing*, 55(4). <https://doi.org/10.1108/EJM-07-2018-0472>
- Jiang, K., Huang, Z., Zhou, X., Tong, C., Zhu, M., & Wang, H. (2023). Deep belief improved bidirectional LSTM for multivariate time series forecasting. *Mathematical Biosciences and Engineering*, 20(9). <https://doi.org/10.3934/mbe.2023739>
- Lee, S., Shin, D., Kwon, K. H., Han, S. P., & Lee, S. K. (2024). Disinformation Spillover: Uncovering The Ripple Effect Of Bot-Assisted Fake Social Engagement On Public Attention. *MIS Quarterly: Management Information Systems*, 48(3). <https://doi.org/10.25300/MISQ/2023/17195>

- Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, 99. <https://doi.org/10.1016/j.procir.2021.03.088>
- Maulidya Prastita Syah, Ajeng Puspa Wardani, Mohammad Idhom, & Trimono. (2025). Perbandingan Representasi Teks Tf-Idf Dan Bert Terhadap Akurasi Cosine Similarity Dalam Penilaian Otomatis Jawaban Berbasis Teks. *Data Sciences Indonesia (DSI)*, 5(1). <https://doi.org/10.47709/dsi.v5i1.6021>
- Ngo, T. T. A., Bui, C. T., Chau, H. K. L., & Tran, N. P. N. (2024). Electronic word-of-mouth (eWOM) on social networking sites (SNS): Roles of information credibility in shaping online purchase intention. *Heliyon*, 10(11). <https://doi.org/10.1016/j.heliyon.2024.e32168>
- Pacheco, D., Hui, P.-M., Torres-Lugo, C., Truong, B. T., Flammini, A., & Menczer, F. (2021). Uncovering Coordinated Networks on Social Media: Methods and Case Studies. *Proceedings of the International AAAI Conference on Web and Social Media*, 15. <https://doi.org/10.1609/icwsm.v15i1.18075>
- Permadi, R. N. (2022). Pemanfaatan Media Sosial Sebagai Platform Utama Pemasaran Produk UMKM. *Avant Garde*, 10(1). <https://doi.org/10.36080/ag.v10i1.1695>
- Rachmad, Y. E. (2024). *From emotional triggers to social media algorithms*. Yoesoep Edhie Rachmad Publication.
- Silalahi, S., Ahmad, T., & Studiawan, H. (2023). Transformer-based Sentiment Analysis for Anomaly Detection on Drone Forensic Timeline. *ISDFS 2023 - 11th International Symposium on Digital Forensics and Security*. <https://doi.org/10.1109/ISDFS58141.2023.10131749>
- Song, Z., Fan, X., Wang, Y., Yan, F., Wu, Z., & Kang, Z. (2026). When comments aren't what they seem: The social media comment toxicity detector for understanding contextual comments. *Expert Systems with Applications*, 305. <https://doi.org/10.1016/j.eswa.2025.130902>
- Usmani, U. A., Abdul Aziz, I., Jaafar, J., & Watada, J. (2024). Deep Learning for Anomaly Detection in Time-Series Data: An Analysis of Techniques, Review of Applications, and Guidelines for Future Research. *IEEE Access*, 12. <https://doi.org/10.1109/ACCESS.2024.3495819>
- Wani, M. A., ElAffendi, M., & Shakil, K. A. (2024). AI-Generated Spam Review Detection Framework with Deep Learning Algorithms and Natural Language Processing. *Computers*, 13(10). <https://doi.org/10.3390/computers13100264>
- Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C., & Salehi, M. (2024). Deep Learning for Time Series Anomaly Detection: A Survey. *ACM Computing Surveys*, 57(1). <https://doi.org/10.1145/3691338>