

TESIS

**EVALUASI HASIL ANALISIS SENTIMENT SERTA DETEKSI
POLA *ECHO CHAMBER* DAN *FILTER BUBBLE* PADA ULASAN
PRODUK *MARKETPLACE***



Oleh:

Muhimmatul Mahmuda

24206051014

**STATE ISLAMIC UNIVERSITY
PROGRAM STUDI INFORMATIKA
PROGRAM MAGISTER FAKULTAS SAINS DAN TEKNOLOGI
UIN SUNAN KALIJAGA
YOGYAKARTA**

2026

PENGESAHAN TUGAS AKHIR



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1927/Un.02/DST/PP.00.9/08/2026

Tugas Akhir dengan judul : Evaluasi Hasil Analisis Sentimen serta Deteksi Pola Echo Chamber dan Filter Bubble pada Ulasan Produk Marketplace

yang dipersiapkan dan disusun oleh:

Nama : MUHIMMATUL MAHMUDA, S. Kom
Nomor Induk Mahasiswa : 24206051014
Telah diujikan pada : Kamis, 13 Agustus 2026
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Valid ID: 6a86844d78ac8

Ketua Sidang

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.
SIGNED



Valid ID: 6a856542792d1

Penguji I

Dr. Agung Fatwanto, S.Si., M.Kom.
SIGNED



Valid ID: 6a85b8ba01013

Penguji II

Prof. Dr. Ir. Shofwatul 'Uyun, S.T., M.Kom.,
IPM., ASEAN Eng.
SIGNED



Valid ID: 6a8782cebc01

Yogyakarta, 13 Agustus 2026
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

SURAT PERNYATAAN KEASLIAN

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : Muhimmatul Mahmuda
NIM : 24206051014
Program Studi : Informatika (S2)
Fakultas : Sains dan Teknologi

Menyatakan dengan sesungguhnya bahwa laporan tesis saya yang berjudul **“Evaluasi Hasil Analisis Sentiment Serta Deteksi Pola *Echo chamber* Dan *Filter bubble* Pada Ulasan Produk Marketplace”** merupakan hasil penelitian dan karya tulis saya pribadi. Kecuali, pada bagian-bagian tertentu yang dirujuk sumber referensinya.

Yogyakarta, 29 Juli 2026

Yang Menyatakan,



Muhimmatul Mahmuda

NIM. 24206051014

STATE ISLAMIC UNIVERSITY
SUNAN KAHMAGA
YOGYAKARTA

SURAT PERNYATAAN BEBAS PLAGIASI

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : Muhimmatul Mahmuda
NIM : 24206051014
Program Studi : Informatika (S2)
Fakultas : Sains dan Teknologi

Menyatakan dengan sesungguhnya bahwa laporan tesis saya yang berjudul **“Evaluasi Hasil Analisis Sentiment Serta Deteksi Pola Echo chamber Dan Filter bubble Pada Ulasan Produk Marketplace”** bebas dari plagiasi yang artinya peneliti menulis sendiri karya ini dan tidak menggunakan karya orang lain tanpa menyebutkan sumber referensinya. Jika dikemudian hari ditemukan bukti bahwa saya telah melanggar pernyataan ini, maka saya siap dikenai sanksi berdasarkan aturan yang berlaku di Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.

Yogyakarta, 29 Juli 2026

Yang Menyatakan,



Muhimmatul Mahmuda
NIM. 24206051014

NOTA DINAS PEMBIMBING

NOTA DINAS PEMBIMBING

Kepada Yth.,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga
Yogyakarta

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul:

**EVALUASI HASIL ANALISIS SENTIMENT SERTA DETEKSI POLA
ECHO CHAMBER DAN FILTER BUBBLE PADA ULASAN PRODUK
MARKETPLACE**

Yang ditulis oleh:

Nama : Muhimmatul Mahmuda
NIM : 24206051014
Jenjang : Magister
Program Studi : Informatika

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta untuk diujikan dalam rangka memperoleh gelar Magister Informatika.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Yogyakarta, 29 Juli 2026
Pembimbing,

Ir. Maria Ulfah Siregar, S.Kom, MIT., Ph.D.
NIP. 19780106 200212 2 001

ABSTRAK

Penelitian ini mengevaluasi performa klasifikasi sentimen serta mendeteksi pola *echo chamber* dan *filter bubble* pada ulasan produk *marketplace* di Indonesia. Latar belakang riset ini didasarkan pada melimpahnya ulasan produk sebagai aset informasi yang belum banyak dieksplorasi implikasi kolektifnya terhadap isolasi informasi pengguna. Menggunakan dataset sebanyak 76.256 ulasan produk kecantikan, penelitian ini bertujuan membandingkan efektivitas tiga model klasifikasi dan memetakan dinamika opini dalam ekosistem digital. Pendekatan yang digunakan mengintegrasikan teori analisis sentimen dengan analisis jaringan kemiripan teks melalui metode *Support Vector Machine* (SVM), *Logistic Regression*, dan IndoBERT. Deteksi pola isolasi informasi dilakukan menggunakan algoritma Louvain untuk identifikasi komunitas dan enam metrik keragaman kosakata. Hasil eksperimen menunjukkan bahwa IndoBERT merupakan model paling optimal dengan akurasi 81,17% dan F1-Score 0,8123. Analisis jaringan mengungkap adanya fenomena *echo chamber* dan *filter bubble* pada tingkat sedang dengan skor modularitas 0,2978 dan terbentuknya 8 komunitas signifikan. Temuan ini mengindikasikan adanya awal fragmentasi opini di mana pengguna cenderung berinteraksi dalam kluster sentimen yang homogen namun belum sepenuhnya terisolasi. Penelitian ini berkontribusi dalam memperluas literatur isolasi informasi pada domain *e-commerce* Indonesia serta menyediakan kerangka kerja praktis bagi pengelola platform untuk memonitor kesehatan interaksi ulasan pengguna.

Kata Kunci: Analisis Sentimen, *Echo chamber*, *Filter bubble*, *Marketplace*, IndoBERT, Jaringan Kemiripan Teks.

ABSTRACT

This study evaluates sentiment classification performance and detects *echo chamber* and *filter bubble* patterns in Indonesian *marketplace* product reviews. The research is motivated by the abundance of product reviews as valuable information assets whose collective implications on user information isolation remain largely unexplored. Using a dataset of 76,256 beauty product reviews, this study aims to compare the effectiveness of three classification models and map opinion dynamics within the digital ecosystem. The approach integrates sentiment analysis theory with social network analysis using *Support Vector Machine* (SVM), *Logistic Regression*, and IndoBERT. Information isolation patterns are detected using the Louvain *Community Detection* algorithm and six *Vocabulary diversity* metrics. Experimental results show that IndoBERT is the most optimal model with an *accuracy* of 81.17% and an *F1-Score* of 0.8123. Network analysis reveals moderate *echo chamber* and *filter bubble* phenomena with a *Modularity Score* of 0.2978 and the formation of 8 significant communities. These findings indicate early stages of opinion fragmentation where users tend to interact within homogeneous sentiment clusters but are not yet fully isolated. This research contributes to expanding the literature on information isolation in the Indonesian *e-commerce* domain and provides a practical framework for platform managers to monitor the health of user review interactions.

Keywords: Sentiment Analysis, *Echo chamber*, *Filter bubble*, *Marketplace*, IndoBERT, *Text similarity network*

MOTTO

(Q.S Al-Kahfi: 39)

“Jangan menyerah, sebab keajaiban sering datang ketika kita hampir putus
asa”



PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT atas segala rahmat, nikmat, dan kekuatan yang telah diberikan, karya tesis ini saya persembahkan kepada:

1. Kedua orang tua saya tercinta, Ayah Mawardi dan Ibu Rabiatal Adawiyah, yang senantiasa memberikan doa, kasih sayang, dukungan, serta pengorbanan yang tidak pernah ternilai sepanjang perjalanan hidup dan pendidikan saya. Terima kasih atas setiap doa yang menjadi kekuatan, setiap nasihat yang menjadi pegangan, dan setiap pengorbanan yang menjadi jalan bagi saya untuk sampai pada tahap ini.
2. Kepada Nenek tercinta, yang selalu memberikan doa, kasih sayang, nasihat, serta dukungan yang tulus. Terima kasih atas perhatian, kehangatan, dan doa-doa baik yang senantiasa mengiringi setiap langkah saya hingga sampai pada tahap ini.
3. Kepada keluarga besar saya yang selalu memberikan semangat, doa, dan dukungan dalam setiap proses yang saya lalui.
4. Kepada dosen pembimbing, dosen penguji, serta seluruh pihak akademik yang telah memberikan ilmu, arahan, dan bimbingan selama proses penyusunan tesis ini.
5. Muhlisha, S. Pd, Lailatul Mukarromah, S.H dan Marwan Hamdany yang tak pernah berhenti memberikan suport dan semangat yang luar biasa untuk penulis.
6. Kepada rekan-rekan dan sahabat yang selalu memberikan bantuan, motivasi, serta kebersamaan dalam menyelesaikan studi ini.
7. Untuk seseorang yang spesial yang selalu memberikan semangat, dukungan serta menjadi tempat suka dan duka dalam penyusunan

Thesis ini. Atas kesabaran, pengertian dan kehadiranmu yang telah menjadi penyemangat di setiap langkah, terimakasih telah menjadi bagian dalam perjalanan ini.

8. Dan yang terakhir kepada diri sendiri. Terimakasih sudah bertahan sejauh ini, selalu siap menyelesaikan penelitian ini walaupun sambil menangis dan jalan-jalan. Terimakasih sudah berjuang sampai dititik ini. Apapun kekurangan dan kelebihanmu mari kita rayakan diri sendiri.

Semoga karya ini dapat memberikan manfaat bagi pembaca dan menjadi bagian kecil dari kontribusi dalam pengembangan ilmu pengetahuan.



KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT atas rahmat, hidayah, dan karunia-Nya, serta shalawat dan salam bagi Baginda Nabi Muhammad SAW, sehingga penulis dapat menyelesaikan penyusunan tesis yang berjudul “Evaluasi Hasil Analisis Sentiment Serta Deteksi Pola *Echo chamber* Dan *Filter bubble* Pada Ulasan Produk *Marketplace*” ini dengan baik. Penyusunan tesis ini dimaksudkan untuk memenuhi salah satu syarat dalam memperoleh gelar Magister Komputer (M. Kom) pada Program Studi Magister Informatika, Fakultas Sains dan Teknologi. Penulis menyadari bahwa keberhasilan dalam penyusunan tesis ini tidak lepas dari bimbingan dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ribuan kata terima kasih kepada:

1. Dekan Fakultas Sains dan Teknologi, Prof. Dr. Dra. Hj. Khurul Wardati, M.Si. atas fasilitas dan dukungan kebijakan akademik yang diberikan selama proses perkuliahan dan penelitian.
2. Ketua Program Studi Magister Informatika, Dr. Ir. Sumarsono, S.T., M. Kom. yang telah memberikan arahan serta kemudahan dalam penyelesaian studi.
3. Dosen Pembimbing Tesis, Ir. Maria Ulfah Siregar, S. Kom., MIT., Ph.D. yang dengan penuh kesabaran telah meluangkan waktu, memberikan bimbingan, masukan, kritik dan saran yang sangat berharga demi kesempurnaan tesis ini.
4. Seluruh dosen Program Studi Magister Informatika UIN Sunan Kalijaga yang telah memberikan ilmu, wawasan, dan pengalaman akademik selama masa perkuliahan.
5. Kepada Lembaga Pengelola Dana Pendidikan (LPDP) dan Program Beasiswa Santri Berprestasi (PBSB) yang telah memberikan beasiswa,

dukungan serta kepercayaan kepada penulis selama menempuh pendidikan. Bantuan tersebut sangat menjadi motivasi dan kesempatan yang sangat berharga bagi penulis untuk menyelesaikan Studi Magister Informatika dan penyusunan tesis ini.

6. Rekan-rekan mahasiswa Magister Informatika serta semua pihak yang telah membantu, baik secara langsung maupun tidak langsung, dalam proses penyelesaian tesis ini.

Semoga segala kebaikan dan pertolongan semuanya mendapat berkah dari Allah Swt. dan akhirnya Penulis menyadari bahwa Thesis ini masih jauh dari kata sempurna, karena keterbatasan ilmu yang Penulis miliki. Untuk itu Penulis dengan kerendahan hati mengharapkan saran dan kritik yang sifatnya membangun dari semua pihak demi membangun laporan penelitian ini.

Yogyakarta, 28 Juli 2026

Yang Menyatakan,



Muhimmatul Mahmuda

NIM. 24206051014

DAFTAR ISI

TESIS	i
PENGESAHAN TUGAS AKHIR.....	ii
SURAT PERNYATAAN KEASLIAN.....	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
NOTA DINAS PEMBIMBING.....	v
ABSTRAK	vi
ABSTRACT	vii
MOTTO	viii
PERSEMBAHAN	ix
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR TABEL.....	xviii
DAFTAR GAMBAR.....	xix
DAFTAR LAMPIRAN.....	xx
DAFTAR SINGKATAN.....	xxi
BAB I PENDAHULUAN	1
A. Latar Belakang.....	1
B. Rumusan Masalah	5
C. Batasan Penelitian	6
D. Tujuan Penelitian.....	8
E. Manfaat Penelitian.....	8

BAB II KAJIAN PUSTAKA DAN LANDASAN TEORI.....	10
A. Kajian Pustaka	10
B. Landasan Teori.....	28
1. <i>Machine learning</i>	28
2. <i>Marketplace</i>	28
3. <i>Natural Language Processing (NLP)</i>	29
4. <i>Text Preprocessing</i>	29
5. <i>Analisis Sentimen</i>	30
6. <i>Ketidakseimbangan Data (Imbalanced Data)</i>	31
7. <i>SMOTE (Synthetic Minority Oversampling Technique)</i>	32
8. <i>GridSearchCV (Grid Search with Cross-Validation)</i>	33
9. <i>TF-IDF (Term Frequency - Inverse Document Frequency)</i>	34
10. <i>Cosine similarity</i>	35
11. <i>Support Vector Machine (SVM)</i>	36
12. <i>Logistic Regression (LR)</i>	37
13. <i>IndoBERT</i>	37
14. <i>Confusion Matrix</i>	38
15. <i>Evaluation Metrics</i>	39
a) <i>Accuraccy</i>	40
b) <i>Precision</i>	40
c) <i>Recall</i>	41
d) <i>F1-Score</i>	41

16.	Jaringan Kemiripan Teks (<i>Text similarity network</i>).....	42
17.	<i>Community Detection</i>	43
18.	<i>Modularity</i>	45
19.	<i>Homophily</i>	46
20.	<i>Isolasi Edge</i>	47
21.	<i>Echo chamber</i>	48
22.	<i>Filter bubble</i>	49
a.	<i>Vocabulary diversity</i>	49
b.	<i>Keyword Dominance</i>	50
c.	<i>Jaccard similarity</i>	51
d.	<i>Shannon entropy</i>	52
e.	<i>Rasio Kata Unik</i>	53
BAB III METODE PENELITIAN		55
A.	Alat dan Bahan	55
B.	Data Penelitian.....	55
C.	Tahapan Penelitian.....	57
1.	Start.....	59
2.	Dataset	59
3.	Preprocessing.....	59
4.	Pembagian Data.....	60
5.	Ekstraksi Fitur TF-IDF	61
6.	Confusion Matrix.....	64

7.	Evaluasi Perbandingan	64
8.	Pilih Model Terbaik	65
9.	Klasifikasi Sentimen Menggunakan Model Terbaik	67
10.	Pembentukan Jaringan.....	67
11.	Deteksi Komunitas	68
12.	Echo chamber	68
13.	Filter bubble.....	69
14.	Evaluasi	72
15.	End.....	72
BAB IV HASIL DAN PEMBAHASAN.....		73
A.	Hasil Penelitian.....	73
1.	Distribusi Data Dan Hasil <i>Preprocessing</i>	73
2.	Hasil Klasifikasi <i>Sentimen</i>	74
3.	Perbandingan <i>Performa</i> Model	79
4.	Pemilihan <i>Model</i> Terbaik.....	81
5.	Klasifikasi <i>Seluruh</i> Data dengan Model Terbaik.....	82
6.	<i>Pembentukan</i> Jaringan Kemiripan Teks	82
7.	Deteksi <i>Komunitas</i>	82
8.	Hasil Analisis <i>Echo chamber</i>	86
9.	Hasil Analisis <i>Filter bubble</i>	89
B.	Pembahasan	93
C.	Analisis Keterbatasan dan Error Model.....	98

1. Karakteristik Data yang Sulit Diklasifikasikan	98
2. Faktor Pembatas Akurasi	99
3. Keterbatasan Threshold Klasifikasi	100
BAB V PENUTUP	102
A. Kesimpulan	102
B. Saran	103
DAFTAR PUSTAKA	105
PERSETUJUAN TIM PENGUJI UJIAN TESIS	120
CURRICULUM VITAE	121



DAFTAR TABEL

Tabel 2. 1 Research Gap Penelitian	22
Tabel 2. 2 Confusion Matrix.....	39
Tabel 3. 1 Jumlah Kategori Rating	57
Tabel 4. 1 Distribusi Data dan Pembagian Dataset	73
Tabel 4. 2 Pembagian Data Training	74
Tabel 4. 3 Pembagian Data Validation dan Data Testing.....	74
Tabel 4. 4 Hasil Support Vector Machine dan Hasil Classification.....	75
Tabel 4. 5 Hasil Logistic Regression dan Hasil Classification.....	76
Tabel 4. 6 Distrubusi Hasil IndoBERT	78
Tabel 4. 7 Perbandingan Hasil Model Klasifikasi Sentimen	79
Tabel 4. 8 Jumlah komunitas pada Jaringan Kemiripan Teks	83
Tabel 4. 9 Homophily Score	87
Tabel 4. 10 Komparasi Deteksi Komunitas Louvain vs METIS	88
Tabel 4. 11 Komparasi Deteksi Filter Bubble TF-IDF VS Word Embedding.....	91

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

Gambar 2. 1 Diagram Fishbone.....	20
Gambar 3. 1 Dataset Publik Ulasan Produk Kecantikan	56
Gambar 3. 2 Dataset Preprocessing Ulasan Produk Kecantikan	57
Gambar 3. 3 Flowchart Penelitian	58
Gambar 4. 1 Confusion matrix Support Vector Machine	75
Gambar 4. 2 Confusion Matrix Logistic Regression.....	77
Gambar 4. 3 Confusion Matrix IndoBERT	78
Gambar 4. 4 Perbandingan Performa Model Klasifikasi Sentimen	80
Gambar 4. 5 Network graph optimized	85
Gambar 4. 6 Matriks kuantitatif echo chamber	88
Gambar 4. 7 Analisis Filter bubble - Karakteristik Bahasa per Sentimen (Vocabulary diversity, Keyword Dominance, Similarity, dan Entropy)....	90

DAFTAR LAMPIRAN

Lampiran 1. Persetujuan Tim Penguji Tesis	120
Lampiran 2. Curriculum Vitae	121



DAFTAR SINGKATAN

SVM	: <i>Support Vector Machine</i>
LR	: <i>Logistic Regression</i>
TF-IDF	: <i>Term Frequency-Inverse Document Frequency</i>
IndoBERT	: <i>Indonesian Bidirectional Encoder Representations from Transformers</i>
BERT	: <i>Bidirectional Encoder Representations from Transformers</i>
NLP	: <i>Natural Language Processing</i>
SMOTE	: <i>Synthetic Minority Over-sampling Technique</i>
GMV	: <i>Gross Merchandise Value</i>
AUC	: <i>Area Under the Curve</i>
TTR	: <i>Type-Token Ratio</i>
LLM	: <i>Large Language Model</i>



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

BAB I PENDAHULUAN

A. Latar Belakang

Pertumbuhan pesat ekonomi digital di Indonesia telah mendorong sektor *marketplace* menjadi salah satu pilar utama perdagangan nasional. Berdasarkan laporan e-Conomy SEA 2024 yang diterbitkan oleh Google, Temasek, dan Bain & Company, ekonomi digital Indonesia diproyeksikan mencapai GMV (*Gross Merchandise Value*) sebesar USD 90 miliar pada tahun 2024 dengan pertumbuhan 13% secara tahunan, menjadikannya yang terbesar di Asia Tenggara (Development Bank, n.d.). Sektor *e-commerce* menjadi kontributor utama dengan nilai GMV mencapai USD 65 miliar, didorong oleh platform seperti Shopee, Tokopedia, dan Lazada yang mencatat jutaan ulasan produk setiap harinya. Ulasan produk ini merupakan aset informasi yang sangat berharga, tidak hanya bagi calon pembeli dalam menentukan keputusan pembelian, tetapi juga bagi penjual dan pengembang produk untuk mengevaluasi kualitas, mengidentifikasi area perbaikan, serta mempertahankan loyalitas pelanggan (Quan et al., 2023). Namun, volume ulasan yang sangat besar dan terus bertambah menyulitkan analisis manual, sehingga diperlukan pendekatan komputasi untuk mengekstrak wawasan secara otomatis.

Analisis sentimen telah menjadi metode utama untuk mengklasifikasikan opini pengguna ke *dalam* kategori positif atau negatif. Dalam konteks ulasan produk *marketplace* berbahasa Indonesia, beberapa pendekatan telah digunakan secara luas. Pendekatan *Machine learning* klasik seperti *Support Vector Machine* (SVM) dan *Logistic Regression* (LR) dengan ekstraksi fitur TF-IDF masih menjadi baseline yang kompetitif karena efisiensi komputasinya, terutama pada dataset berdimensi tinggi

(Alaiya & Agusniar, 2025; Hamdi et al., 2026). Di sisi lain, pendekatan deep learning berbasis Transformer, khususnya IndoBERT yang dilatih secara spesifik pada korpus bahasa Indonesia, telah menunjukkan performa yang unggul dalam memahami konteks dan nuansa linguistik teks informal Indonesia, mencapai akurasi hingga 93% pada data ulasan *marketplace* (Aras et al., 2024; Koto et al., n.d.). Kombinasi ketiga model ini memungkinkan perbandingan performa yang komprehensif antara pendekatan *Machine learning* klasik dan deep learning dalam konteks analisis sentimen ulasan produk *marketplace* berbahasa Indonesia.

Namun, sebagian besar riset analisis sentimen di Indonesia masih berhenti pada tahap perbandingan akurasi antar model tanpa mengeksplorasi implikasi yang lebih dalam dari distribusi sentimen tersebut terhadap perilaku pengguna secara kolektif (Harsanto & Sudarmilah, 2025). Padahal, pengguna dengan sentimen serupa cenderung menggunakan kosakata, topik, dan cara ekspresi yang mirip, yang berpotensi membentuk kelompok-kelompok yang terisolasi secara informasional atau yang dikenal sebagai *echo chamber* (Ge et al., 2020). *Echo chamber* merujuk pada kondisi ketika individu hanya terpapar informasi yang memperkuat pandangan mereka sendiri akibat interaksi yang homogen secara sosial (Hartmann et al., 2025), sedangkan *filter bubble* menggambarkan isolasi informasi yang terjadi ketika konten yang tidak sesuai preferensi pengguna tersaring secara otomatis oleh algoritma platform (Terren & Borge, 2021).

Studi terbaru menunjukkan bahwa fenomena ini tidak hanya terjadi di platform media sosial, tetapi juga mulai terdeteksi di sistem rekomendasi *marketplace*. Riset pada platform Taobao di Alibaba menemukan bukti adanya kecenderungan *echo chamber* pada perilaku klik pengguna di e-commerce, meskipun efeknya lebih rendah pada perilaku pembelian (Ge et

al., 2020). Lebih lanjut, penelitian tentang information cocoon di lingkungan *e-commerce* mengungkapkan bahwa rekomendasi berbasis kecerdasan buatan memperkuat hubungan antara preferensi individu dengan pembentukan gelembung informasi, yang pada gilirannya mengumpulkan konsensus dalam kelompok namun memperlebar perbedaan antar kelompok (Chen et al., 2025).

Meskipun demikian, masih terdapat kesenjangan penelitian yang *signifikan*. Pertama, sebagian besar studi *echo chamber* dan *filter bubble* berfokus pada konteks politik, berita, dan media sosial di negara-negara Barat. Sebuah tinjauan sistematis terhadap 129 studi mengungkapkan bahwa basis bukti menderita bias geografis yang kuat terhadap konteks Barat, sementara pendekatan lintas budaya dan studi di luar *Global North* masih sangat kurang (Hartmann et al., 2025). Kedua, studi yang mengintegrasikan analisis sentimen dengan *deteksi echo chamber* dan *filter bubble* dalam satu kerangka kerja yang utuh pada data ulasan produk *marketplace* belum banyak dilakukan. Tinjauan sistematis terhadap 28 studi *deteksi echo chamber* menunjukkan bahwa sebagian besar metode berbasis topologi jaringan tanpa mengintegrasikan analisis semantik konten secara memadai (Mahmoudi et al., 2024). Baru-baru ini, sebuah studi berhasil mengintegrasikan *aspect-based sentiment analysis* dengan metrik konsensus untuk mendeteksi *echo chamber*, namun masih terbatas pada data media sosial dan belum diterapkan pada konteks ulasan produk *marketplace* (Amendola et al., 2024). Riset yang ada cenderung memisahkan kedua domain ini, yaitu analisis sentimen berhenti pada evaluasi model, sedangkan analisis *echo chamber* berdiri sendiri tanpa memanfaatkan prediksi sentimen sebagai dasar pembentukan jaringan.

Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan *mengusulkan* sebuah pipeline yang mengintegrasikan analisis sentimen menggunakan tiga pendekatan, SVM dengan SMOTE untuk mengatasi ketidakseimbangan data (Ramadhanti & Handayani, n.d.), *Logistic Regression* yang dioptimasi dengan *GridSearchCV* (Yasyfi Imran et al., 2025), dan IndoBERT dengan *weighted loss function* (Br Sembiring et al., 2026), serta deteksi pola *echo chamber* dan *filter bubble* melalui jaringan kemiripan teks pada dataset ulasan produk *marketplace* yang terdiri dari 76.256 ulasan. Untuk memperkuat validitas akademik dari temuan *echo chamber* dan *filter bubble* yang dihasilkan, hasil algoritma Louvain divalidasi silang dengan algoritma METIS mengikuti pendekatan (Villa et al., 2021), sementara hasil analisis *filter bubble* berbasis TF-IDF dibandingkan dengan representasi word embedding, sehingga pola yang teridentifikasi dapat dipastikan tidak bergantung pada satu metode atau satu representasi fitur teks saja, mengingat literatur terdahulu menunjukkan bahwa metrik dan metode deteksi *echo chamber* dan *filter bubble* belum sepenuhnya memadai dan belum mencapai konsensus universal (Mahmoudi et al., 2024). Dengan menghubungkan hasil prediksi sentimen ke dalam pembentukan jaringan pengguna berdasarkan kemiripan kosakata menggunakan TF-IDF dan *cosine similarity*, penelitian ini tidak hanya mengevaluasi kinerja model dalam mengklasifikasikan sentimen, tetapi juga mengungkap bagaimana polaritas opini berkorelasi dengan struktur sosial dan pola isolasi informasi di antara pengguna.

Lebih jauh, penelitian ini bertujuan memberikan manfaat praktis bagi para pelaku bisnis dan pengembang produk. Dengan memahami bagaimana opini positif dan negatif membentuk kluster-kluster yang terisolasi, di mana pengguna dengan sentimen tertentu cenderung hanya menggunakan

kosakata dan perspektif yang sama (Mahmoudi et al., 2024), penjual dapat merancang strategi yang lebih efektif untuk mempertahankan loyalitas pelanggan. Identifikasi pengguna yang terperangkap dalam *echo chamber* negatif memungkinkan intervensi yang ditargetkan, seperti kampanye edukasi produk atau peningkatan kualitas layanan (Gupta & Bansal, 2024). Sementara itu, pengguna dalam *echo chamber* positif dapat dijadikan brand advocate untuk memperkuat reputasi produk, karena personalisasi yang tepat dapat memperdalam ikatan emosional konsumen dengan brand (Obiegbu & Larsen, 2025). Deteksi *filter bubble* juga memberikan wawasan tentang kesenjangan persepsi antara kelompok pengguna yang puas dan tidak puas (Xu et al., 2026), sehingga perbaikan produk dapat difokuskan pada aspek yang benar-benar dikeluhkan oleh segmen pengguna tertentu, bukan hanya berdasarkan suara mayoritas (Modgil et al., 2024). Diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan analisis perilaku pengguna *marketplace* berbasis NLP dan analisis jaringan teks.

B. Rumusan Masalah

Berdasarkan *latar* belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa model SVM dengan SMOTE dan *GridSearchCV*, *Logistic Regression* dengan *GridSearchCV*, serta IndoBERT dengan *weighted loss function* dalam mengklasifikasikan sentimen ulasan produk *marketplace* berbahasa Indonesia secara binary?
2. Bagaimana pola *echo chamber* dan *filter bubble* terbentuk pada jaringan pengguna berdasarkan kemiripan teks ulasan produk?
3. Bagaimana hubungan antara hasil analisis sentimen dengan struktur *echo chamber* serta *filter bubble* yang terdeteksi dalam jaringan ulasan produk *marketplace*?

C. Batasan Penelitian

Agar penelitian *lebih* terarah dan tidak menyimpang dari tujuan yang telah ditentukan, maka batasan penelitian dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan *berupa* ulasan produk marketplace kecantikan berbahasa Indonesia yang diperoleh dari sumber dataset publik yaitu HuggingFace. Penelitian ini tidak menggunakan data primer hasil scraping mandiri maupun dataset ulasan dari domain produk lain (misalnya elektronik, fashion, atau makanan).
2. Dataset penelitian berasal dari platform marketplace Shopee dan Tokopedia. Penelitian ini tidak mencakup platform marketplace lain seperti Lazada, Blibli, atau TikTok Shop.
3. Penelitian hanya berfokus pada analisis sentimen dengan dua kategori, yaitu positif dan negatif. Penelitian ini tidak menyertakan kategori sentimen netral maupun klasifikasi multi-kelas lainnya.
4. Metode machine learning yang digunakan dalam penelitian ini adalah Support Vector Machine (SVM), Logistic Regression (LR), dan IndoBERT. Penelitian ini tidak menggunakan algoritma machine learning lain seperti Naive Bayes, Random Forest, atau K-Nearest Neighbors, maupun algoritma deep learning lain seperti LSTM, CNN, atau varian BERT multibahasa (mBERT, XLM-RoBERTa).
5. Model analisis sentimen yang digunakan terbatas pada tiga pendekatan: SVM dengan SMOTE dan GridSearchCV, Logistic Regression dengan SMOTE dan GridSearchCV, serta IndoBERT dengan weighted loss function. Penelitian ini tidak menguji kombinasi teknik penyeimbangan kelas lain seperti ADASYN, random undersampling, maupun focal loss.

6. Representasi fitur teks untuk model SVM dan Logistic Regression menggunakan TF-IDF, sedangkan IndoBERT menggunakan tokenisasi berbasis transformer (WordPiece). Penelitian ini tidak menguji representasi fitur lain seperti Word2Vec, GloVe, atau FastText untuk model machine learning.
7. Ekstraksi fitur untuk model SVM dan LR menggunakan TF-IDF dengan 5.000 fitur dan n-gram range (1,2). Penelitian ini tidak menguji ukuran fitur maupun rentang n-gram lain di luar konfigurasi tersebut.
8. Jaringan kemiripan teks dibangun menggunakan cosine similarity dengan threshold adaptif berdasarkan persentil ke-75, dan jumlah node dibatasi pada 1.500 sampel pengguna yang dipilih secara proporsional berdasarkan sentimen. Penelitian ini tidak membangun jaringan kemiripan teks pada skala populasi penuh maupun dengan metrik kemiripan lain seperti Euclidean distance atau Pearson correlation.
9. Deteksi echo chamber menggunakan algoritma Louvain (NetworkX) dengan metrik modularitas, homofili, dan isolasi tepi, serta dibandingkan dengan algoritma METIS sebagai validasi silang. Penelitian ini tidak menggunakan algoritma deteksi komunitas lain seperti Girvan-Newman atau Label Propagation.
10. Analisis filter bubble dilakukan berdasarkan vocabulary diversity, keyword dominance, cosine similarity, jaccard similarity, rasio kata unik, dan Shannon entropy pada kelompok sentimen, dengan perbandingan representasi TF-IDF dan word embedding. Penelitian ini tidak menggunakan metrik filter bubble lain di luar keenam metrik tersebut.
11. Penelitian ini tidak membahas sistem rekomendasi marketplace maupun analisis perilaku pengguna secara real-time.

D. Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengevaluasi dan membandingkan performa tiga model analisis sentimen SVM dengan SMOTE dan *GridSearchCV*, *Logistic Regression* dengan *GridSearchCV*, serta IndoBERT dengan *weighted loss function* dalam mengklasifikasikan sentimen ulasan produk *marketplace* berbahasa Indonesia secara binary.
2. Mendeteksi dan menganalisis pola *echo chamber* dan *filter bubble* yang terbentuk dalam jaringan ulasan produk berdasarkan kemiripan teks.
3. Mengintegrasikan hasil analisis sentimen dengan deteksi *echo chamber* serta *filter bubble* untuk memberikan gambaran komprehensif tentang dinamika opini pengguna dalam ekosistem ulasan produk *marketplace*.

E. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis sebagai berikut:

1. Bagi penjual dan pengembang produk: Memberikan wawasan tentang segmentasi opini pengguna yang terperangkap dalam *echo chamber*, sehingga strategi pemasaran dan peningkatan produk dapat lebih ditargetkan. Pengguna dalam *echo chamber* negatif dapat diidentifikasi untuk diberikan edukasi atau solusi khusus, sementara konsumen positif dapat dimanfaatkan sebagai brand advocate untuk memperkuat reputasi produk.
2. Bagi platform *marketplace*: Menyediakan kerangka kerja untuk memonitor kesehatan ekosistem ulasan, mendeteksi polarisasi opini yang ekstrem antar kelompok pengguna, dan merancang intervensi algoritmik yang dapat mengurangi *filter bubble* tanpa mengurangi kualitas rekomendasi.

3. Bagi akademisi: Memberikan kontribusi pada literatur *echo chamber* dan *filter bubble* di luar konteks media sosial dan politik yaitu pada domain *marketplace* di Indonesia yang masih jarang diteliti serta menyediakan pipeline yang dapat direproduksi untuk penelitian serupa di masa depan.
4. Bagi konsumen: Meningkatkan kesadaran akan adanya *filter bubble* dalam ulasan produk sehingga konsumen dapat lebih kritis dalam mengevaluasi informasi sebelum membuat keputusan pembelian.

BAB V

PENUTUP

A. Kesimpulan

Penelitian ini bertujuan mengevaluasi performa klasifikasi sentimen serta mengidentifikasi pola *echo chamber* dan *filter bubble* pada ulasan produk *marketplace* berbasis jaringan kemiripan teks. Dengan menggunakan 76.256 ulasan produk kecantikan, penelitian ini menunjukkan bahwa pendekatan berbasis Transformer mampu memberikan representasi semantik yang lebih kuat dibandingkan model berbasis fitur tradisional, sehingga hasil klasifikasi lebih stabil dan representatif untuk data ulasan berbahasa Indonesia yang cenderung informal.

Dari sisi performa model, IndoBERT dengan *weighted loss function* terbukti paling optimal dengan *accuracy* 0.8117 dan *F1-Score* 0.8123, mengungguli *Logistic Regression* (*F1-Score* 0.7892) dan SVM (*F1-Score* 0.7883). Keunggulan ini menunjukkan bahwa mekanisme self-attention pada arsitektur BERT mampu *menangkap* konteks dua arah secara lebih baik, sehingga prediksi sentimen menjadi lebih akurat meskipun distribusi kelas tidak seimbang.

Sementara itu, analisis jaringan memperlihatkan adanya fenomena *echo chamber* pada *tingkat* sedang, dengan *Modularity* 0.2978, 8 komunitas signifikan, *Homophily* negatif 61.5%, *Homophily* positif 57.7%, dan isolasi *edge* 40.8%. Temuan ini menandakan bahwa polarisasi berdasarkan sentimen sudah cukup terbentuk, namun belum mengarah pada isolasi informasi yang sepenuhnya tertutup.

Pada aspek *filter bubble*, data menunjukkan adanya kecenderungan penguatan kosakata spesifik pada masing-masing kelompok sentimen, dengan *cosine similarity* 0.818, *Jaccard similarity* 0.328, *Shannon entropy*

negatif 10.21 bits, dan *Shannon entropy* positif 10.15 bits. Kondisi ini memperlihatkan bahwa meskipun masih terdapat kesamaan kosakata antarkelompok, masing-masing sentimen cenderung membentuk pola bahasa yang lebih terkonsentrasi pada narasi emosional atau estetik produk, sehingga memperkuat terbentuknya *filter bubble*.

Kedua temuan tersebut divalidasi lebih lanjut melalui komparasi metode tambahan. Deteksi echo chamber menggunakan algoritma METIS menghasilkan Modularity Score sebesar 0,2715 dan Coverage 37,1%, sedangkan pengukuran filter bubble menggunakan pendekatan word embedding menghasilkan Cosine Similarity 0,920 dan Jaccard Similarity 0,075. Kedua hasil komparasi ini secara konsisten mengonfirmasi kategori sedang pada echo chamber dan kuat pada filter bubble yang diperoleh melalui algoritma Louvain dan representasi TF-IDF, sehingga memperkuat validitas akademik dari kedua temuan utama penelitian ini.

B. Saran

Hasil penelitian ini memberikan beberapa implikasi dan rekomendasi bagi pengelola platform, pengembang model, maupun peneliti selanjutnya. Pertama, pengelola platform *marketplace* perlu membangun mekanisme moderasi yang lebih cerdas, misalnya dengan mendeteksi lebih dini komunitas yang mulai terbentuk dari sentimen yang homogen sehingga dapat diberikan intervensi berupa diversifikasi ulasan yang ditampilkan kepada pengguna.

Kedua, peneliti pengembangan model *Machine learning* dapat menggabungkan fitur teks dengan fitur struktural jaringan, misalnya melalui graph neural networks, agar klasifikasi sentimen tidak hanya bergantung pada pola kata tetapi juga mempertimbangkan hubungan antar pengguna dalam jaringan. Pendekatan ini berpotensi meningkatkan akurasi

sekaligus memberikan pemahaman yang lebih kaya mengenai konteks sosial di balik sentimen.

Ketiga, penelitian lanjutan sebaiknya diperluas ke kategori produk lain, seperti elektronik atau fashion, guna melihat apakah pola homofili negatif dan pembentukan *filter bubble* bersifat umum atau justru spesifik pada produk kecantikan. Perluasan ini *penting* agar temuan penelitian tidak terbatas pada satu domain, melainkan dapat memberikan gambaran yang lebih komprehensif mengenai dinamika opini di *marketplace* Indonesia.

Keempat, mengingat komparasi metode tambahan pada penelitian ini (algoritma METIS untuk echo chamber dan word embedding untuk filter bubble) masih dijalankan pada skala sampel terbatas, penelitian lanjutan disarankan untuk mereplikasi kedua metode pembandingan tersebut pada skala dataset penuh (76.256 ulasan) guna menguji konsistensi temuan pada cakupan data yang lebih besar. Selain itu, disarankan pula untuk menguji sensitivitas threshold yang digunakan dalam penelitian ini, seperti ambang Modularity Score ($Q > 0,3$), Homophily (60%), dan dominansi komunitas (70%), mengingat literatur belum menetapkan konsensus universal atas nilai-nilai tersebut, sehingga keandalan kategorisasi echo chamber dan filter bubble dapat diuji lebih lanjut pada berbagai skenario ambang batas.

DAFTAR PUSTAKA

- Adya, T. V. (n.d.). *Impact Of E-Commerce and Internet Users on Indonesia's Economic Growth*.
- Alaiya, A., & Agusniar, C. (2025a). Sentiment Analysis of E-Commerce Product Reviews on Tokopedia Using Support Vector Machine. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Number 5). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Alaiya, A., & Agusniar, C. (2025b). Sentiment Analysis of E-Commerce Product Reviews on Tokopedia Using Support Vector Machine. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Number 5). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Alatawi, F., Sheth, P., & Liu, H. (2023). *Quantifying the Echo Chamber Effect: An Embedding Distance-based Approach*. <https://doi.org/https://doi.org/10.48550/arXiv.2307.04668>
- Alnuaimi, A. F. A. H., & Albaldawi, T. H. K. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, 97. <https://doi.org/10.1051/bioconf/20249700133>
- Amendola, M., Cavaliere, D., De Maio, C., Fenza, G., & Loia, V. (2024). Towards echo chamber assessment by employing aspect-based sentiment analysis and GDM consensus metrics. *Online Social Networks and Media*, 39–40. <https://doi.org/10.1016/j.osnem.2024.100276>
- Andriyani, D., Faqih, A., & Permana, S. E. (2025). *Journal of Artificial Intelligence and Engineering Applications The Effect of SMOTE Application on Support Vector Machine Performance in Sentiment Classification on Imbalanced Datasets* (Vol. 4, Number 2). <https://doi.org/10.59934/jaiea.v4i2.742>

- Anuar, S. H. H., Abas, Z. A., Yunos, N. M., Mukhtar, M. F., Setiadi, T., & Shibghatullah, A. S. (2024). Identifying Communities with Modularity Metric Using Louvain and Leiden Algorithms. *Pertanika Journal of Science and Technology*, 32(3), 1285–1300. <https://doi.org/10.47836/pjst.32.3.16>
- Aras, S., Yusuf, M., Ruimassa, R. Y., Wambrauw, E. A. B., & Pala'langan, E. B. (2024). Sentiment Analysis on Shopee Product Reviews Using IndoBERT. *Journal of Information Systems and Informatics*, 6(3), 1616–1627. <https://doi.org/10.51519/journalisi.v6i3.814>
- Aufar, A. F., Mochamad Alfian Rosid, Eviyanti, A., & Astutik, I. R. I. (2023). Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews. *JICTE (Journal of Information and Computer Technology Education)*, 7(2), 42–50. <https://doi.org/10.21070/jicte.v7i2.1650>
- Azmina, F. N., Liebenlito, M., & Zulkifli, D. U. (2025). Community Detection, Topic, and Sentiment Analysis of the Palestine-Israel Issue on Social Media X. *Jurnal Matematika, Statistika Dan Komputasi*, 21(2), 502–519. <https://doi.org/10.20956/j.v21i2.41308>
- Azzahra, F., Faqih, A., Pratama, D., & Kurnia, D. A. (2026). *Journal of Artificial Intelligence and Engineering Applications Comparison of TF-IDF and Word2Vec Feature Representations for Emotion Classification of Tokopedia E-Commerce Review Using LinearSVC* (Vol. 5, Number 2). <https://ioinformatic.org/>
- Bellina, A., Sardo, D. R. Lo, Brugnoli, E., Saracco, F., Gravino, P., Loreto, V., & Di Bona, G. (2025). *Language bubbles in online social networks*. <http://arxiv.org/abs/2507.13068>

- Bestgen, Y. (n.d.). *Measuring Lexical Diversity in Texts Measuring Lexical Diversity in Texts: The Twofold Length Problem*.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10). <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Blondel, V., Guillaume, J. L., & Lambiotte, R. (2024). Fast unfolding of communities in large networks: 15 years later. In *Journal of Statistical Mechanics: Theory and Experiment* (Vol. 2024, Number 10). Institute of Physics. <https://doi.org/10.1088/1742-5468/ad6139>
- Bojić, L., Ilić, V., Prodanović, V., & Vuković, V. (2025). An Agent-Based Simulation of Politicized Topics Using Large Language Models: Algorithmic Personalization and Polarization on Social Media. *Chinese Political Science Review*. <https://doi.org/10.1007/s41111-025-00326-x>
- Br Sembiring, A. B. S., Robet, R., & Hoki, L. (2026). Comparison of IndoBERT and SVM Performance in Sentiment Analysis of Digital Education Platforms. *Sinkron*, 10(1), 64–74. <https://doi.org/10.33395/sinkron.v10i1.15472>
- Budianoor, R., Saputro, S. W., Abadi, F., Nugroho, R. A., & Farmadi, A. (2026). Quantifying the Impact of Text Preprocessing on IndoBERT Fine-Tuning for Indonesian Informal Culinary Sentiment Analysis. *Journal of Computing Theories and Applications*, 3(4), 564–581. <https://doi.org/10.62411/jcta.15980>
- Cahya, L. D., Luthfiarta, A., Krisna, J. I. T., Winarno, S., & Nugraha, A. (2024). Improving Multi-label Classification Performance on Imbalanced Datasets Through SMOTE Technique and Data

- Augmentation Using IndoBERT Model. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(3), 290–298. <https://doi.org/10.25077/teknosi.v9i3.2023.290-298>
- Cartus, A. R., Samuels, E. A., Cerdá, M., & Marshall, B. D. L. (2023). Outcome class imbalance and rare events: An underappreciated complication for overdose risk prediction modeling. *Addiction*, 118(6), 1167–1176. <https://doi.org/10.1111/add.16133>
- Cau, E., Morini, V., & Rossetti, G. (2024). Trends and topics: Characterizing echo chambers' topological stability and in-group attitudes. *PLOS Complex Systems*, 1(2), e0000008. <https://doi.org/10.1371/journal.pcsy.0000008>
- Chamidy, T. (2025). Application of SMOTE in Sentiment Analysis of MyXL User Reviews on Google Play Store. In *Jurnal Informatika Sunan Kalijaga* (Vol. 10, Number 1).
- Chen, S., Qiu, H., & He, W. (2025). The information cocoon paradox: fostering unity or fueling divergence? *Humanities and Social Sciences Communications*, 12(1). <https://doi.org/10.1057/s41599-025-05169-0>
- Civico, M. (2025). Linguistic statistical universals: comparing computer- and human-generated texts. *International Journal of Digital Humanities*, 7(1), 1–37. <https://doi.org/10.1007/s42803-025-00096-7>
- Cui, J., Wang, Z., Ho, S. B., & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, 56(8), 8469–8510. <https://doi.org/10.1007/s10462-022-10386-z>
- De Francisci Morales, G., Monti, C., & Starnini, M. (2021). No echo in the chambers of political interactions on Reddit. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-81531-x>

- De Nicola, G., Tuekam Mambou, V. H., & Kauermann, G. (2023). COVID-19 and social media: Beyond polarization. *PNAS Nexus*, 2(8). <https://doi.org/10.1093/pnasnexus/pgad246>
- Development Bank, A. (n.d.). *E-commerce Evolution in Asia and the Pacific*. *E-Commerce Evolution in Asia and The Pacific Opportunities and Challenges*. <https://doi.org/http://dx.doi.org/10.22617/TCS230473-2>
- Fajar Nawulansih, D., Ceisa Santi, N., & Aristia Sa'ida, I. (2025a). Analisis Sentimen Ulasan Aplikasi DANA di Google Play Store: Penerapan Support Vector Machine dan Synthetic Minority Over-sampling Technique. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(9), 2660–2671. <https://doi.org/10.52436/1.jpti.1053>
- Fajar Nawulansih, D., Ceisa Santi, N., & Aristia Sa'ida, I. (2025b). Analisis Sentimen Ulasan Aplikasi DANA di Google Play Store: Penerapan Support Vector Machine dan Synthetic Minority Over-sampling Technique. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(9), 2660–2671. <https://doi.org/10.52436/1.jpti.1053>
- Farhadpour, S., Warner, T. A., & Maxwell, A. E. (2024). Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices. *Remote Sensing*, 16(3). <https://doi.org/10.3390/rs16030533>
- Feng, B., Cheng, F., Liu, Y., Chang, X., Wang, X., & Jin, D. (2024). Community Detection on Social Networks with Sentimental Interaction. *International Journal on Semantic Web and Information Systems*, 20(1). <https://doi.org/10.4018/IJSWIS.341232>
- Fredrick, D. R., & Craven, L. (2025). Lexical diversity, syntactic complexity, and readability: a corpus-based analysis of ChatGPT and

- L2 student essays. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1616935>
- Gao, Y., Liu, F., & Gao, L. (2023). Echo chamber effects on short video platforms. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-33370-1>
- Ge, Y., Zhao, S., Zhou, H., Pei, C., Sun, F., Ou, W., & Zhang, Y. (2020). Understanding Echo Chambers in E-commerce Recommender Systems. *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2261–2270. <https://doi.org/10.1145/3397271.3401431>
- Guo, Y., Shang, G., & Clavel, C. (2025). Benchmarking Linguistic Diversity of Large Language Models. *Transactions of the Association for Computational Linguistics*, 13, 1507–1526. <https://doi.org/10.1162/tacl.a.47>
- Gupta, T., & Bansal, S. (2024). *Beyond Echo Chambers: Unraveling the Impact of Social Media Algorithms on Consumer Behavior and Exploring Pathways to a Diverse Digital Discourse*. 7(1), 15–37. www.carijournals.org
- Hadiwijaya, M. A., & Sutoyo, R. (n.d.). *Sentiment Analysis on Tokopedia Product Reviews using Natural Language Processing 1 st*. Retrieved <https://scikit-learn.org/>
- Hamdi, R. H., Hutabarat, I. M., Sinaga, H. G., Muthoharoh, L., Satria, A., & Manullang, M. C. T. (2026). *Benchmarking Logistic Regression, SVM, and LightGBM Against BiLSTM with Attention for Sentiment Analysis on Indonesian Product Reviews*. <http://arxiv.org/abs/2604.25452>

- Harsanto, M., & Sudarmilah, E. (2025). Tinjauan Literatur Analisis Sentimen Produk E-Commerce: Dataset, Pendekatan, Metode, Dan Performa. *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 10(3), 2290–2303. <https://doi.org/10.29100/jipi.v10i3.8217>
- Hartmann, D., Wang, S. M., Pohlmann, L., & Berendt, B. (2025a). A systematic review of echo chamber research: comparative analysis of conceptualizations, operationalizations, and varying outcomes. *Journal of Computational Social Science*, 8(2). <https://doi.org/10.1007/s42001-025-00381-z>
- Hartmann, D., Wang, S. M., Pohlmann, L., & Berendt, B. (2025b). *A Systematic Review of Echo Chamber Research: Comparative Analysis of Conceptualizations, Operationalizations, and Varying Outcomes*. <http://arxiv.org/abs/2407.06631>
- Hasan Fahreza, M., Sipayung, N. Z., Santoso, I. W., Susilowati, S., & Kunci, K. (n.d.). *Komparasi SVM-LR Berbasis Indobert dan SMOTE-TOMEK pada Deteksi Buzzer*.
- Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. *IEEE Access*, 10, 19083–19095. <https://doi.org/10.1109/ACCESS.2022.3151048>
- Imron, S., Setiawan, E. I., Santoso, J., & Purnomo, M. H. (2023). Aspect Based Sentiment Analysis Marketplace Product Reviews Using BERT, LSTM, and CNN. *Jurnal RESTI*, 7(3), 586–591. <https://doi.org/10.29207/resti.v7i3.4751>
- Interian, R., Marzo, R. G., Mendoza, I., & Ribeiro, C. C. (2023). *Network polarization, filter bubbles, and echo chambers: An annotated review of measures and reduction methods*. <https://doi.org/10.1111/itor.13224>

- Jazuli, A., Widowati, & Kusumaningrum, R. (2025). Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback. *Applied Sciences (Switzerland)*, 15(1). <https://doi.org/10.3390/app15010172>
- Kamal, H. (2024). Advanced Hybrid Transformer-CNN Deep Learning Model for Effective Intrusion Detection Systems with Class Imbalance Mitigation Using Resampling Techniques. *Future Internet*, 16(12). <https://doi.org/10.3390/fi16120481>
- Kim, J., Park, S., & Carriquiry, A. (2024). A deep learning approach for the comparison of handwritten documents using latent feature vectors. *Statistical Analysis and Data Mining*, 17(1). <https://doi.org/10.1002/sam.11660>
- Kohavi, R. (1995). *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. <http://robotics.stanford.edu/~ronnyk>
- Kono, M. F., Fajri, I. N., & Pristyanto, Y. (2025). Public Sentiment Analysis on Corruption Issues in Indonesia Using IndoBERT Fine-Tuning, Logistic Regression, and Linear SVM. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Number 5). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (n.d.). *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*. Online. Retrieved <https://huggingface.co/>
- Kratzke, N. (2023). How to Find Orchestrated Trolls? A Case Study on Identifying Polarized Twitter Echo Chambers. *Computers*, 12(3). <https://doi.org/10.3390/computers12030057>
- Li, J., Lai, S., Shuai, Z., Tan, Y., Jia, Y., Yu, M., Song, Z., Peng, X., Xu, Z., Ni, Y., Qiu, H., Yang, J., Liu, Y., & Lu, Y. (2024). *A Comprehensive*

Review of Community Detection in Graphs.

<http://arxiv.org/abs/2309.11798>

- Lubis, A. R., Lase, Y. Y., Rahman, D. A., & Witarsyah, D. (2023). Improving Spell Checker Performance for Bahasa Indonesia Using Text Preprocessing Techniques with Deep Learning Models. *Ingenierie Des Systemes d'Information*, 28(5), 1335–1342. <https://doi.org/10.18280/isi.280522>
- Mahmoudi, A., Jemielniak, D., & Ciechanowski, L. (2024a). Echo Chambers in Online Social Networks: A Systematic Literature Review. *IEEE Access*, 12, 9594–9620. <https://doi.org/10.1109/ACCESS.2024.3353054>
- Mahmoudi, A., Jemielniak, D., & Ciechanowski, L. (2024b). Echo Chambers in Online Social Networks: A Systematic Literature Review. *IEEE Access*, 12, 9594–9620. <https://doi.org/10.1109/ACCESS.2024.3353054>
- Min, Y., Jiang, T., Jin, C., Li, Q., & Jin, X. (2019). Endogenetic structure of filter bubble in social networks. *Royal Society Open Science*, 6(11). <https://doi.org/10.1098/rsos.190868>
- Misriati, T., & Aryanti, R. (2024). Optimalisasi Random Forest dan Support Vector Machine dengan Hyperparameter GridSearchCV untuk Analisis Sentimen Ulasan PrimaKu. *Journal of Information System Research (JOSH)*, 5(4), 1333–1341. <https://doi.org/10.47065/josh.v5i4.5347>
- Modgil, S., Singh, R. K., Gupta, S., & Dennehy, D. (2024). A Confirmation Bias View on Social Media Induced Polarisation During Covid-19. *Information Systems Frontiers*, 26(2), 417–441. <https://doi.org/10.1007/s10796-021-10222-9>

- Mohammed, M. T., & Rashid, O. F. (2023). Document retrieval using term frequency inverse sentence frequency weighting scheme. *Indonesian Journal of Electrical Engineering and Computer Science*, 31(3), 1478–1485. <https://doi.org/10.11591/ijeecs.v31.i3.pp1478-1485>
- Motohashi, K., Koshiba, H., & Ikeuchi, K. (2024). Measuring science and innovation linkage using text mining of research papers and patent information. *Scientometrics*, 129(4), 2159–2179. <https://doi.org/10.1007/s11192-024-04949-w>
- Mujahid, M., Kina, E. R. O. L., Rustam, F., Villar, M. G., Alvarado, E. S., De La Torre Diez, I., & Ashraf, I. (2024). Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-024-00943-4>
- Muñoz-Ortiz, A., Gómez-Rodríguez, C., & Vilares, D. (2024). Contrasting Linguistic Patterns in Human and LLM-Generated News Text. *Artificial Intelligence Review*, 57(9). <https://doi.org/10.1007/s10462-024-10903-2>
- Nabilah, A., Sari, N. I., Afrina, M., & Ibrahim, A. (2025). The Sentiment Analysis of Indonesian Startup Application Reviews Using TF-IDF+SVM and FastText: A Comparative Study. In *Journal of Information Technology and Computer Science* (Vol. 10, Number 3). www.jitecs.ub.ac.id
- Nakajima, H., & Sasaki, M. (2023). Text Classification Based on the Heterogeneous Graph Considering the Relationships between Documents. *Big Data and Cognitive Computing*, 7(4). <https://doi.org/10.3390/bdcc7040181>

- Newman, M. E. J. (2006). *Modularity and community structure in networks*. www.pnas.org/cgi/doi/10.1073/pnas.0601602103
- Obiegbu, C. J., & Larsen, G. (2025). Algorithmic personalization and brand loyalty: An experiential perspective. *Marketing Theory*, 25(2), 199–219. <https://doi.org/10.1177/14705931241230041>
- Ortu, M., Romano, M., & Carta, A. (2024). Semi-supervised topic representation through sentiment analysis and semantic networks. *Big Data Research*, 37. <https://doi.org/10.1016/j.bdr.2024.100474>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2018). *Scikit-learn: Machine Learning in Python*. <http://arxiv.org/abs/1201.0490>
- Purba, M., Ermatita, E., Abdiansah, A., Noprisson, H., Ayumi, V., Salamah, U., Setiawan, H., & Yadi, Y. (n.d.). Effect of Random Splitting and Cross Validation for Indonesian Opinion Mining using Machine Learning Approach. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 13, Number 9). Retrieved www.ijacsa.thesai.org
- Purwanto, D. D. (2026). Empirical Evaluation of IndoBERT and LSTM for Sentiment Analysis of Tourism Reviews: A Data-Driven Study on Kenjeran Park. *Jurnal Teknik Informatika (Jutif)*, 7(1), 463–474. <https://doi.org/10.52436/1.jutif.2026.7.1.4901>
- Quan, H., Li, S., Zeng, C., Wei, H., & Hu, J. (2023). *Big Data and AI-driven Product Design: A Survey*. <https://doi.org/10.21203/rs.3.rs-2899904/v1>

- Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-56706-x>
- Ramadhanti, N. D., & Handayani, R. N. (n.d.). *JIP (Jurnal Informatika Polinema)* Analisis Sentimen Ulasan Pengguna Aplikasi Shopee di Google Play Store Menggunakan *Support Vector Machine* Dengan Teknik SMOTE.
- Riyanto, S., Sitanggang, I. S., Djatna, T., & Atikah, T. D. (n.d.). Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification. In *IJACSA International Journal of Advanced Computer Science and Applications* (Vol. 14, Number 6). Retrieved <http://gcancer.org/pdr>
- Rofi'i, M., Vikri, M. J., & Rohmah, R. (2026). Implementasi SMOTE dan GrideSearchCV untuk Klasifikasi Sentimen Imbalanced pada Isu Reshuffle Kabinet. *VOCATECH: Vocational Education and Technology Journal*, 8(1), 140–154. <https://doi.org/10.38038/vocatech.v8i1.305>
- Rosillo-Rodes, P., Miguel, M. S., & Sanchez, D. (2025). *Entropy and type-token ratio in gigaword corpora*. <https://doi.org/10.1103/rxxz-lk3n>
- Santani Mulyono, H., & Saprudin, U. (2025). Efektivitas Logistic Regression dalam Analisis Sentimen Berbahasa Indonesia pada Komentar YouTube tentang Isu Ketenagakerjaan. In *Jurnal Indonesia : Manajemen Informatika dan Komunikasi (JIMIK)* (Vol. 6, Number 3). <https://journal.stmiki.ac.id>
- Sarker, I. H. (2021). *Machine Learning: Algorithms, Real-World Applications and Research Directions*. <https://doi.org/10.20944/preprints202103.0216.v1>

- Setiawan, A., Abidin, Z., & Imamudin, M. (2025). Impact of Preprocessing on Indonesian Extractive Summarization Using LexRank, TextRank, DivRank, and Cosine Similarity. *G-Tech: Jurnal Teknologi Terapan*, 9(4), 2311–2321. <https://doi.org/10.70609/g-tech.v9i4.8306>
- Shabrina Assyifa, D., & Luthfiarta, A. (2024). SMOTE-Tomek Re-sampling Based on Random Forest Method to Overcome Unbalanced Data for Multi-class Classification. *Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 9(2), 151–160. <https://doi.org/10.25139/inform.v9i2.8410>
- Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, 2, 100003. <https://doi.org/10.1016/j.nlp.2022.100003>
- Song, X., Guo, S., & Gao, Y. (2024). Personality traits and their influence on Echo chamber formation in social media: a comparative study of Twitter and Weibo. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1323117>
- Sudaryanto, S., Hanim, A., Rosediana Dewi, I., Kartikasari, A. D., & Rusdiyanto, R. (2025). The mediating effect of customer trust of E-WOM and online customer reviews impacting purchase decision of household electronic products at a marketplace: evidence from Indonesia. *Cogent Business and Management*, 12(1). <https://doi.org/10.1080/23311975.2025.2503093>
- Sujon, K. M., Hassan, R., Choi, K., & Samad, M. A. (2025). Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models.

- Journal of Big Data*, 12(1). <https://doi.org/10.1186/s40537-025-01313-4>
- Terren, L., & Borge, R. (2021). Echo Chambers on Social Media: A Systematic Review of the Literature. *Review of Communication Research*, 9, 1–39. <https://doi.org/10.12840/ISSN.2255-4165.028>
- Tucudean, G., Bucos, M., Dragulescu, B., & Căleanu, C. D. (2024). Natural language processing with transformers: a review. In *PeerJ Computer Science* (Vol. 10). PeerJ Inc. <https://doi.org/10.7717/PEERJ-CS.2222>
- Ulya, D. M., Juhari, J., Yuliana, R. E., & Jamhuri, M. (2025). Reliable and Efficient Sentiment Analysis on IMDb with Logistic Regression. *CAUCHY: Jurnal Matematika Murni Dan Aplikasi*, 10(2), 821–834. <https://doi.org/10.18860/cauchy.v10i2.33809>
- Villa, G., Pasi, G., & Viviani, M. (2021). Echo chamber detection and analysis: A topology- and content-based approach in the COVID-19 scenario. *Social Network Analysis and Mining*, 11(1). <https://doi.org/10.1007/s13278-021-00779-3>
- Wahyu Widyananda, Maskur, & Ahmad Fauzi. (2025). Sentiment analysis of Indonesian e-commerce product reviews using machine learning classifiers. *International Journal of Science and Research Archive*, 16(2), 445–450. <https://doi.org/10.30574/ijrsra.2025.16.2.2345>
- Williams, H. T. P., McMurray, J. R., Kurz, T., & Hugo Lambert, F. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change*, 32, 126–138. <https://doi.org/10.1016/j.gloenvcha.2015.03.006>
- Xu, J., Liu, C., & Lu, L. (2026). Emotionally Structured Interaction Networks and Consumer Perception of New Energy Vehicle Technology: A Behavioral Network Analysis of Online Brand

Communities. *Behavioral Sciences*, 16(1), 112.
<https://doi.org/10.3390/bs16010112>

Yasyfi Imran, A., Rudi Sanjaya, M., Putra, B. W., Ekoputra, G., Cahyadi, H., Sriwijaya, U., Informasi, S., Palembang -Prabumulih, J., 32, K. M., Ogan Ilir, K., & Selatan, S. (2025). *Optimization Of Sentiment Analysis on Tokopedia User Reviews Using Gridsearchcv and SMOTE with Machine Learning Algorithms* Optimasi Analisis Sentimen Ulasan Tokopedia Dengan Gridsearchcv, Smote Dan Algoritma Machine Learning. 10(3).

Zaidi, S. F. A., Woo, H., & Lee, C. G. (2022). A Graph Convolution Network-Based Bug Triage System to Learn Heterogeneous Graph Representation of Bug Reports. *IEEE Access*, 10, 20677–20689.
<https://doi.org/10.1109/ACCESS.2022.3153075>

