



**PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING
DALAM ANALISIS SENTIMEN BERBASIS TOPIK TERHADAP
TWEET SERIAL BLACK MIRROR**

**TUGAS AKHIR
SKRIPSI**

**AKASYAH SAVINA NOORLY
41522010073**

MERCU BUANA

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2025**



**PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING
DALAM ANALISIS SENTIMEN BERBASIS TOPIK TERHADAP
TWEET SERIAL BLACK MIRROR**

**TUGAS AKHIR
SKRIPSI**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

**AKASYAH SAVINA NOORLY
41522010073**

MERCU BUANA

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2025**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Akasyah Savina Noorly
NIM : 41522010073
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“Perbandingan Kinerja Algoritma Machine Learning dalam Analisis Sentimen Berbasis Topik terhadap Tweet Serial *Black Mirror*” adalah hasil karya saya sendiri, tidak mengandung unsur plagiarisme, pelanggaran hak cipta, atau konten ilegal dalam bentuk apapun dan tidak melanggar hukum atau hak pihak manapun.

Apabila di kemudian hari ditemukan pelanggaran terhadap pernyataan ini, saya bersedia menanggung seluruh konsekuensi hukum dan membebaskan Universitas Mercu Buana dari segala bentuk tuntutan hukum dan saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Demikian pernyataan ini saya buat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

UNIVERSITAS
MERCU BUANA

Jakarta, 20 Januari 2026



Akasyah Savina Noorly

HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN



072.423.4.03.01

PERNYATAAN SIMILARITY CHECK **/SIMILARITY CHECK STATEMENT**

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
/The undersigned, hereby declare that the scientific work written by

Nama /Name : Akasyah Savina Noorly
NIM /Student ID Number : 41522010073
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

/The title:

**“PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING DALAM
ANALISIS SENTIMEN BERBASIS TOPIK TERHADAP TWEET TENTANG
SERIAL BLACK MIRROR”**

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

/Similarity checks have been carried out with the Turnitin system on the date:

31 Desember 2025

dengan nilai persentase sebesar :

/with a percentage value of:

13%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di Fakultas Ilmu Komputer
Universitas Mercu Buana. */declared to meet standards in accordance with applicable
regulations at the Faculty of Computer Science, Universitas Mercu Buana.*

File hasil cek similarity turnitin:

/Turnitin similarity report file

https://drive.google.com/file/d/1AYLbWZb7E9NzVPK0dAVNK9lfcFx2a4/view?usp=drive_link



Jakarta, 31 Desember 2025
Admin Turnitin Fasilkom UMB

Agung Prawoto, S.Kom., B.Sc
NIK : 322970503

Fakultas Ilmu Komputer
KAMPUS MENARA BHAKTI
Jl. Raya Meruya Selatan No. 1 Kembangan, Jakarta Barat 11650
Telp. 021-5840816 (Hunting), Psw : 5700 Fax. 021-5840813
<http://www.mercubuana.ac.id>, e-mail : fasilkom@mercubuana.ac.id

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Akasyah Savina Noorly
NIM : 41522010073
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir : Perbandingan Kinerja Algoritma Machine Learning
dalam Analisis Sentimen Berbasis Topik Terhadap
Tweet Serial *Black Mirror*

Telah berhasil dipertahankan pada sidang tanggal 20 Januari 2026 dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing



Fauzi Nur Iman S.Kom, M.Kom

NIDN/NUPTK: 0318088903

Jakarta, 20 Januari 2026

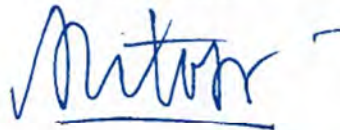
Mengetahui,

Dekan Fakultas Ilmu Komputer



Dr. Bambang Jokonowo, S.Si., MTI
NIDN/NUPTK: 0320037002

Ketua Program Studi
Teknik Informatika



Dr. Hadi Santoso, S.Kom., M.Kom
NIDN/NUPTK: 0225067701

KATA PENGANTAR

Puji syukur dipanjatkan kepada Tuhan Yang Maha Esa atas segala kebaikannya, sehingga Tugas Akhir ini dapat terselesaikan. Penulisan Tugas Akhir ini dilakukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Teknik Informatika pada Fakultas Ilmu Komputer Universitas Mercu Buana.

Penulis menyadari bahwa keberhasilan penyusunan Tugas Akhir ini merupakan hasil dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., MTI selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana.
4. Bapak Fauzi Nur Iman S.Kom., M.Kom. selaku dosen pembimbing yang telah memberikan arahan, motivasi, serta meluangkan waktu dan pemikiran dalam proses penyusunan Tugas Akhir ini.
5. Mamah dan Papah, atas segala bentuk kasih sayang, dukungan, dan doa yang selalu mengiringi di sepanjang hidup, khususnya selama menempuh masa studi di Universitas Mercu Buana.
6. Yanto 'si bocil', si pelipur lara berbulu yang selalu berhasil menjadi hiburan sejenak di sela-sela rutinitas menyusun tugas akhir.
7. *And for every individual who's left a drag path etched into mine.*

Akhir kata, penulis berharap segala bantuan dan dukungan yang telah diberikan dapat berbalas dengan kebaikan. Semoga Tugas Akhir ini dapat memberikan manfaat dan kontribusi bagi pengembangan ilmu komputer di masa mendatang.

Jakarta, 20 Januari 2026



Akasyah Savina Noorly

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Akasyah Savina Noorly
NIM : 41522010073
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Laporan Skripsi : Perbandingan Kinerja Algoritma Machine Learning dalam Analisis Sentimen Berbasis Topik terhadap Tweet Serial Black Mirror

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif** (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 20 Januari 2026

Yang menyatakan.

UNIVERSITAS
MERCU BUANA



Akasyah Savina Noorly

PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING DALAM ANALISIS SENTIMEN BERBASIS TOPIK TERHADAP TWEET SERIAL *BLACK MIRROR*

ABSTRAK

Media sosial X (Twitter) telah menjadi wadah utama bagi masyarakat untuk mengekspresikan opini terhadap fenomena budaya populer, termasuk serial televisi *Black Mirror*. Namun, analisis sentimen konvensional sering kali gagal menangkap konteks spesifik dari topik yang dibicarakan. Penelitian ini bertujuan untuk melakukan analisis sentimen berbasis topik guna memetakan opini publik secara lebih mendalam dan membandingkan kinerja tiga algoritma machine learning: Logistic Regression, Support Vector Machine (SVM), dan Extreme Gradient Boosting (XGBoost). Metode penelitian meliputi preprocessing data, pemodelan topik menggunakan Latent Dirichlet Allocation (LDA), pelabelan sentimen otomatis menggunakan RoBERTa dengan penyesuaian aturan *genre*, serta klasifikasi sentimen. Dataset yang digunakan terdiri dari 6.497 tweet berbahasa Inggris. Hasil pemodelan topik berhasil mengidentifikasi lima topik diskusi utama (Topik 0 hingga Topik 4). Distribusi sentimen menunjukkan dominasi respons Netral (40,6%), diikuti Negatif (32%) yang dominan pada Topik 0, dan Positif (27,5%) yang dominan pada Topik 1. Dalam komparasi kinerja algoritma, XGBoost terbukti menjadi model terbaik dengan akurasi mencapai 95%, mengungguli Logistic Regression (93%) dan SVM (92%). Penelitian ini menyimpulkan bahwa penggabungan LDA untuk pemodelan topik dan XGBoost dengan fitur embedding kontekstual merupakan pendekatan yang sangat efektif untuk menganalisis opini publik pada data media sosial yang kompleks.

Kata kunci: Analisis Sentimen, Topic Modeling, LDA, XGBoost, *Black Mirror*, RoBERTa.

PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING DALAM ANALISIS SENTIMEN BERBASIS TOPIK TERHADAP TWEET SERIAL *BLACK MIRROR*

ABSTRACT

Social media platform X (Twitter) has become a primary venue for public expression regarding popular culture phenomena, including the television series Black Mirror. However, conventional sentiment analysis often fails to capture the specific context of the discussion topics. This study aims to conduct topic-based sentiment analysis to map public opinion more deeply and compare the performance of three machine learning algorithms: Logistic Regression, Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). The research method involves data preprocessing, topic modeling using Latent Dirichlet Allocation (LDA), automated sentiment labeling using RoBERTa with genre-rule adjustments, and sentiment classification. The dataset consists of 6,497 English tweets. The topic modeling results successfully identified five main discussion topics (Topic 0 to Topic 4). Sentiment distribution shows a dominance of Neutral responses (40.6%), followed by Negative (32%) which is dominant in Topic 0, and Positive (27.5%) which is dominant in Topic 1. In the algorithm performance comparison, XGBoost proved to be the best model with an accuracy of 95%, outperforming Logistic Regression (93%) and SVM (92%). This study concludes that combining LDA for topic modeling and XGBoost with contextual embedding features is a highly effective approach for analyzing public opinion on complex social media data.

Keywords: *Sentiment Analysis, Topic Modeling, LDA, XGBoost, Black Mirror, RoBERTa.*

DAFTAR ISI

HALAMAN SAMPUL.....	0
HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Penelitian Terdahulu.....	7
2.2 Teori Pendukung	27
2.2.1 Analisis Sentimen.....	27
2.2.2 Natural Language Processing (NLP).....	27
2.2.3 RoBERTa (Robustly Optimized BERT Pretraining Approach) .	28
2.2.4 Machine Learning dalam Analisis Sentimen	29

2.2.5	Analisis Sentimen Berbasis Topik	32
2.2.6	Media Sosial X (Twitter) sebagai Sumber Data.....	33
2.2.7	Evaluasi Kinerja Algoritma.....	34
BAB III	METODE PENELITIAN	35
3.1	Jenis Penelitian	35
3.2	Tahapan Penelitian	36
3.2.1	Identifikasi Permasalahan	37
3.2.2	Studi Literatur	37
3.2.3	Pengumpulan Data	37
3.2.4	Text Preprocessing	37
3.2.5	Topic Modeling (LDA)	38
3.2.6	Pelabelan Sentimen (RoBERTa).....	38
3.2.7	Feature Extraction dan Split Data	39
3.2.8	Klasifikasi Sentimen	39
3.2.9	Evaluasi Model.....	39
BAB IV	PEMBAHASAN	40
4.1	Deskripsi dan Statistik Dataset.....	40
4.2	Implementasi Text Preprocessing.....	41
4.2.1	Text Cleaning	41
4.2.2	Penghapusan Duplikasi	41
4.2.3	Contraction Expansion	42
4.2.4	Stopword Removal.....	43
4.2.5	Tokenisasi dan Lemmatization (spaCy).....	43
4.2.6	Bigram.....	44
4.2.7	Validasi Data Akhir (Final Cleaning)	44
4.3	Analisis Pemodelan Topik (LDA).....	45
4.3.1	Penentuan Jumlah Topik Optimal (Coherence Score)	46
4.3.2	Tuning Hyperparameter	47
4.3.3	Hasil Topic Modeling.....	48
4.3.4	Distribusi Topik Dominan.....	51
4.4	Sentiment Labeling (RoBERTA)	52
4.4.1	Genre-Specific Rules.....	52

4.4.2	Hasil Distribusi Sentimen Global.....	53
4.5	Analisis Sentimen Berbasis Topik	54
4.5.1	Distribusi Sentimen per Topik	54
4.5.2	Interpretasi Hasil Analisis	55
4.6	Feature Extraction	55
4.6.1	Label Encoding	56
4.6.2	RoBERTa Embeddings (Mean Pooling)	56
4.7	Evaluasi Kinerja Algoritma Machine Learning	56
4.7.1	Stratified Split	57
4.7.2	Perbandingan Performa Model (Global)	57
4.7.3	Confusion Matrix	59
4.7.4	Evaluasi Akurasi Berbasis Topik	61
BAB V	KESIMPULAN DAN SARAN.....	64
5.1	Kesimpulan.....	64
5.2	Saran.....	65
DAFTAR PUSTAKA.....		67
LAMPIRAN.....		73


 UNIVERSITAS
 MERCU BUANA

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	7
Tabel 4. 1 Preview Data Hasil <i>Crawling</i>	40
Tabel 4. 2 Nilai Hasil <i>Coherence Score</i>	46
Tabel 4. 3 Tahapan Perhitungan <i>Coherence Score</i>	47
Tabel 4. 4 Topik Dominan	50
Tabel 4. 5 Distribusi Sentimen Global	53



DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	36
Gambar 4. 1 Hasil <i>Text Cleaning</i>	41
Gambar 4. 2 Penghapusan Duplikasi	42
Gambar 4. 3 Daftar pemetaan kata untuk proses <i>contraction expansion</i>	42
Gambar 4. 4 Hasil <i>Contraction Expansion</i>	42
Gambar 4. 5 List <i>Stopwords</i> Khusus Topik.....	43
Gambar 4. 6 Hasil <i>Stopword Removal</i> dan Tokenisasi.....	43
Gambar 4. 7 Hasil Pembuatan <i>Bigram</i>	44
Gambar 4. 8 Hasil Drop Row Kosong	44
Gambar 4. 9 Hasil Akhir <i>Text Preprocessing</i>	45
Gambar 4. 10 Implementasi <i>Dictionary</i> dan <i>Filter Extremes</i>	46
Gambar 4. 11 Hasil <i>Coherence Score</i>	46
Gambar 4. 12 Hasil <i>Tuning Hyperparameter</i>	48
Gambar 4. 13 Wordcloud Topik 0	48
Gambar 4. 14 Wordcloud Topik 1	49
Gambar 4. 15 Wordcloud Topik 2	49
Gambar 4. 16 Wordcloud Topik 3	49
Gambar 4. 17 Wordcloud Topik 4	50
Gambar 4. 18 <i>Intertopic Distance Map</i>	51
Gambar 4. 19 Hasil Distribusi Topik Dominan	52
Gambar 4. 20 Input Data untuk Sentiment Labeling	52
Gambar 4. 21 Implementasi <i>Genre Praise Words</i> dan <i>Explicit Hate Words</i>	53
Gambar 4. 22 Bar Chart Distribusi Sentimen Global	53
Gambar 4. 23 Tabel Distribusi Sentimen per Topik	54
Gambar 4. 24 Stacked Bar Chart Sentimen per Topik.....	55
Gambar 4. 25 <i>Extract Embeddings</i>	56
Gambar 4. 26 Hasil Split Data	57
Gambar 4. 27 <i>Classification Report</i> Algoritma Logistic Regression	57
Gambar 4. 28 <i>Classification Report</i> Algoritma SVM.....	58
Gambar 4. 29 <i>Classification Report</i> Algoritma XGBoost	58
Gambar 4. 30 Grafik Perbandingan Metrik Evaluasi.....	59

Gambar 4. 31 <i>Confusion Matrix</i> Algoritma Logistic Regression	59
Gambar 4. 32 <i>Confusion Matrix</i> Algoritma SVM.....	60
Gambar 4. 33 <i>Confusion Matrix</i> Algoritma XGBoost.....	60
Gambar 4. 34 Tabel Pivot Akurasi Sentimen per Topik.....	61
Gambar 4. 35 <i>Graph</i> Persentase Sentimen dengan Logistic Regression.....	62
Gambar 4. 36 <i>Graph</i> Persentase Sentimen dengan SVM.....	62
Gambar 4. 37 <i>Graph</i> Persentase Sentimen dengan XGBoost	63



DAFTAR LAMPIRAN

Lampiran 1. Kartu Asistensi	73
Lampiran 2. Curriculum Vitae	74
Lampiran 3. Surat Pernyataan.....	75
Lampiran 4. Surat Pengalihan Hak Cipta.....	76
Lampiran 5. Keterangan BNSP.....	77
Lampiran 6. Form Revisi Dosen Penguji Pertama.....	78
Lampiran 7. Form Revisi Dosen Penguji Kedua	79

