



**PERBANDINGAN METODE PELABELAN OTOMATIS
TERHADAP KINERJA SUPPORT VECTOR MACHINE
DALAM ANALISIS SENTIMEN KOMENTAR POLITIK DI
MEDIA SOSIAL X**

**TUGAS AKHIR
SKRIPSI**

MAHESA HARYO PAMUNGKAS
41522010017

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**



**PERBANDINGAN METODE PELABELAN OTOMATIS
TERHADAP KINERJA SUPPORT VECTOR MACHINE
DALAM ANALISIS SENTIMEN KOMENTAR POLITIK DI
MEDIA SOSIAL X**

**TUGAS AKHIR
SKRIPSI**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

**UNIVERSITAS
MAHESA HARYO PAMUNGKAS
41522010017
MERCU BUANA**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Mahesa Haryo Pamungkas

NIM : 41522010017

Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“Perbandingan Metode Pelabelan Otomatis terhadap Kinerja Support Vector Machine dalam Analisis Sentimen Komentar Politik di Media Sosial X” 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, 30 Juli 2026



Mahesa Haryo Pamungkas

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 : Mahesa Haryo Pamungkas
NIM /Student ID Number : 41522010017
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

/The title:

“Perbandingan Metode Pelabelan Otomatis terhadap Kinerja Support Vector Machine dalam Analisis Sentimen Komentar Politik di Media Sosial X”

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

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

22 Juli 2026

dengan nilai persentase sebesar :

/with a percentage value of:

22%

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/19lBF8-kaKgEZVWRkTzDxRI-xaxyrcBvQ/view?usp=drive_link



Jakarta, 22 Juli 2026
Admin Turnitin Fasilkom UMB



Agung Prawoto

Agung Prawoto, S.Kom., B.Sc

NIK : 322970503

HALAMAN PENGESAHAN

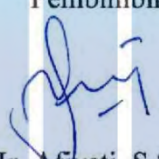
Tugas Akhir ini diajukan oleh:

Nama : MAHESA HARYO PAMUNGKAS
NIM : 41522010017
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir : Perbandingan Metode Pelabelan Otomatis terhadap Kinerja Support Vector Machine dalam Analisis Sentimen Komentar Politik di Media Sosial X

Telah berhasil dipertahankan pada sidang tanggal 30 Juli 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


Dr. Ir. Afyati, S.Si., M.T.

NIDN/NUPTK: 0316106908

UNIVERSITAS

Jakarta, 30 Juli 2026

Mengetahui,

MERCU BUANA

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 saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Tugas Akhir ini. Penulisan Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Teknik Informatika pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Tugas Akhir ini, sangatlah sulit bagi saya untuk menyelesaikan Tugas Akhir ini. Oleh karena itu, saya 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. Ibu Dr. Ir. Afiyati, S.Si., M.T. selaku dosen pembimbing tugas akhir yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan tugas akhir ini terjadwal dengan baik.
5. Kepada Ayah dan Bunda yang selalu mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercu Buana. Tanpa doa, kerja keras, dan pengorbanan beliau, saya tidak akan mampu menyelesaikan pendidikan sarjana ini.
6. Kepada seorang perempuan istimewa yang selalu mengisi hari-hari saya dengan kebahagiaan. Terima kasih atas doa, dukungan, perhatian, dan semangat yang selalu diberikan, terutama di saat saya merasa lelah dan terpuruk. Kebaikan hati, paras yang cantik, serta senyum manisnya menjadi sumber semangat bagi saya untuk terus melangkah. Terima kasih telah menjadi pendengar yang baik dan selalu percaya pada kemampuan saya. Semoga segala kebaikanmu dibalas dengan kebahagiaan, serta semoga kita dapat terus saling mendukung dalam setiap langkah menuju masa depan.
7. Kepada teman-teman Emplang–Emplang yang selalu memberikan semangat, dukungan, dan kebersamaan selama proses penyusunan tugas akhir ini. Terima kasih atas canda tawa, motivasi, serta kebersamaan yang

telah menjadi pelepas penat di tengah kesibukan dan tekanan dalam menyelesaikan penelitian ini. Semoga pertemanan yang telah terjalin dapat terus terjaga dan membawa kebaikan bagi kita semua.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Tugas Akhir ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 30 Juli 2025



Mahesa Haryo Pamungkas



HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB

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

Nama : MAHESA HARYO PAMUNGKAS
NIM : 41522010017
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Laporan Skripsi : Perbandingan Metode Pelabelan Otomatis terhadap Kinerja Support Vector Machine dalam Analisis Sentimen Komentar Politik di Media Sosial X

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.

UNIVERSITAS
MERCU BUANA

Jakarta, 30 Juli 2026

Yang menyatakan,



Mahesa Haryo Pamungkas

**PERBANDINGAN METODE PELABELAN OTOMATIS TERHADAP
KINERJA SUPPORT VECTOR MACHINE DALAM ANALISIS
SENTIMEN KOMENTAR POLITIK DI MEDIA SOSIAL X**

MAHESA HARYO PAMUNGKAS

ABSTRAK

Media sosial X menjadi salah satu platform utama bagi masyarakat untuk menyampaikan opini terhadap tokoh politik sehingga menghasilkan data komentar dalam jumlah besar. Kondisi tersebut menyebabkan pelabelan manual menjadi kurang efisien dan mendorong penggunaan metode pelabelan otomatis. Penelitian ini bertujuan membandingkan performa tiga metode pelabelan otomatis, yaitu InSet Lexicon, IndoBERT, dan RoBERTa, sebagai sumber label pelatihan Support Vector Machine (SVM) pada analisis sentimen komentar di akun X Joko Widodo. Data diperoleh melalui proses *crawling* menggunakan GraphQL API sebanyak 62.614 komentar dan menghasilkan 48.546 data setelah tahap *preprocessing*. Sebanyak 8.790 komentar digunakan sebagai *ground truth* yang dianotasi oleh tiga annotator dan divalidasi menggunakan Fleiss' Kappa, dengan nilai 0,7182 (*Substantial Agreement*). Evaluasi dilakukan menggunakan Accuracy, Macro F1-score, Weighted F1-score, serta McNemar Test, kemudian dilengkapi dengan eksperimen RandomOverSampler untuk menganalisis pengaruh ketidakseimbangan kelas terhadap performa model. Hasil penelitian menunjukkan bahwa SVM yang dilatih menggunakan label RoBERTa memberikan performa terbaik pada dataset asli dengan Accuracy 68,98%, Macro F1-score 55,31%, dan Weighted F1-score 69,61%. Hasil McNemar Test menunjukkan seluruh perbedaan performa antar model signifikan secara statistik ($p\text{-value} < 0,05$). Sementara itu, eksperimen RandomOverSampler menunjukkan bahwa distribusi kelas memengaruhi performa SVM, di mana penyeimbangan data mengubah nilai metrik evaluasi pada setiap metode pelabelan. Temuan ini menunjukkan bahwa metode pelabelan otomatis dan distribusi kelas berperan penting dalam membentuk performa klasifikasi sentimen menggunakan SVM.

Kata kunci: Analisis Sentimen, Pelabelan Otomatis, Support Vector Machine, Transformer-based Labeling, InSet Lexicon.

COMPARISON OF AUTOMATIC LABELING METHODS ON THE PERFORMANCE OF SUPPORT VECTOR MACHINE FOR SENTIMENT ANALYSIS OF POLITICAL COMMENTS ON SOCIAL MEDIA X

MAHESA HARYO PAMUNGKAS

ABSTRACT

X is one of the primary social media platforms where users express opinions about political figures, generating a large volume of textual data. This condition makes manual annotation inefficient and encourages the use of automatic labeling methods. This study compares the performance of three automatic labeling approaches, namely InSet Lexicon, IndoBERT, and RoBERTa, as training label sources for a Support Vector Machine (SVM) model in sentiment analysis of comments posted on Joko Widodo's X account. A total of 62,614 comments were collected using the GraphQL API, resulting in 48,546 preprocessed comments. Among them, 8,790 comments were manually annotated by three annotators and used as the ground truth. The reliability of the manual annotations was assessed using Fleiss' Kappa, yielding a value of 0.7182, which indicates Substantial Agreement. Model performance was evaluated using Accuracy, Macro F1-score, Weighted F1-score, and McNemar's Test. An additional experiment employing RandomOverSampler was conducted to examine the effect of class imbalance on SVM performance. The results show that the SVM trained with RoBERTa-generated labels achieved the best performance on the original dataset, with an Accuracy of 68.98%, Macro F1-score of 55.31%, and Weighted F1-score of 69.61%. Furthermore, McNemar's Test indicated statistically significant differences among the three labeling methods (p -value < 0.05). The RandomOverSampler experiment also demonstrated that class distribution influences SVM performance by affecting the evaluation metrics of each labeling method. These findings suggest that both automatic labeling strategies and class distribution play important roles in sentiment classification using SVM.

Keywords : Sentiment Analysis, Automatic Labeling, Support Vector Machine, Transformer-based Labeling, InSet Lexicon.

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	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.4.1 Manfaat Teoritis	4
1.4.2 Manfaat Praktis	4
1.5 Batasan Masalah	4
BAB II TINJAUAN PUSTAKA	6
2.1 Penelitian Terdahulu	6
2.2 Teori Pendukung	14
2.2.1 Media Sosial X	14
2.2.2 Analisis Sentimen	14
2.2.3 Natural Language Processing (NLP)	15
2.2.4 TF – IDF	16
2.2.5 Support Vector Machine (SVM)	18
2.2.6 IndoBERT	19
2.2.7 RoBERTa	20
2.2.8 InSet Lexicon	22

BAB III METODE PENELITIAN	24
3.1 Jenis Penelitian.....	24
3.2 Tahapan Penelitian.....	25
3.3 Desain Penelitian	30
BAB IV PEMBAHASAN	32
4.1 Dataset.....	32
4.2 Hasil Preprocessing.....	33
4.3 Hasil Pelabelan Data	34
4.3.1 Pelabelan Manual	34
4.3.2 Pelabelan Menggunakan InSet Lexicon	37
4.3.3 Pelabelan Menggunakan IndoBERT	39
4.3.4 Pelabelan Menggunakan RoBERTa.....	41
4.4 Hasil Klasifikasi SVM	43
4.4.1 SVM Dengan Label InSet Lexicon	44
4.4.2 SVM Dengan Label IndoBERT	46
4.4.3 SVM Dengan Label RoBERTa	48
4.4.4 Perbandingan Kinerja SVM	51
4.5 Perbandingan Performa.....	52
4.5.1 Perbandingan Hasil Evaluasi.....	52
4.5.2 Evaluasi Per-Class.....	53
4.5.3 Uji Signifikansi Menggunakan McNemar	54
4.6 Analisis Pengaruh <i>Imbalance</i> Kelas terhadap Performa SVM	55
4.6.1 Eksperimen Menggunakan <i>RandomOverSampler</i>	56
4.6.2 Perbandingan Dataset Asli dan <i>Balanced</i> Dataset	58
4.6.3 Analisis Pengaruh Distribusi Kelas terhadap Metrik Evaluasi.....	59
4.7 Analisis Kesalahan.....	60
BAB V KESIMPULAN DAN SARAN	62
5.1 Kesimpulan	62
5.2 Saran	63
DAFTAR PUSTAKA.....	65
LAMPIRAN.....	69

DAFTAR TABEL

Tabel 4. 1 Dataset Hasil Crawling	32
Tabel 4. 2 Kolom <i>Name</i> dan Kolom <i>Text</i>	32
Tabel 4. 3 Distribusi Dataset.....	33
Tabel 4. 4 Hasil <i>Preprocessing</i>	34
Tabel 4. 5 Hasil Pengukuran <i>Fleiss' Kappa</i>	35
Tabel 4. 6 Distribusi Pelabelan Manual	36
Tabel 4. 7 Proses Pelabelan Inset Lexicon.....	37
Tabel 4. 8 Distribusi Pelabelan Inset Lexicon	37
Tabel 4. 9 Hasil Pelabelan IndoBERT	39
Tabel 4. 10 Distirbusi Pelabelan IndoBERT.....	40
Tabel 4. 11 Hasil Pelabelan RoBERTa.....	42
Tabel 4. 12 Distribusi Pelabelan RoBERTa.....	42
Tabel 4. 13 Evaluasi Label Inset Lexicon.....	44
Tabel 4. 14 Evaluasi Label Indobert	46
Tabel 4. 15 Evaluasi Label RoBERTA.....	49
Tabel 4. 16 Perbandingan Hasil Evaluasi Label	53
Tabel 4. 17 Perbandingan Evaluasi Per-Class	53
Tabel 4. 18 Perbandingan Uji McNemar	55
Tabel 4. 19 Distribusi Data Sebelum dan Sesudah <i>RandomOverSampler</i>	57
Tabel 4. 20 Perbandingan Hasil Evaluasi Dataset Asli dan <i>Balanced</i> Dataset.....	58

DAFTAR GAMBAR

Gambar 3. 1 Flowchart Penelitian.....	30
Gambar 4. 1 Proses Pelabelan Manual.....	35
Gambar 4. 2 Visualisasi Distribusi Pelabelan Inset Lexicon	38
Gambar 4. 3 Visualisasi Distribusi Pelabelan IndoBERT	41
Gambar 4. 4 Visualisasi Distribusi Pelabelan RoBERTa	43
Gambar 4. 5 Confusion Matrix SVM dengan Label Inset Lexicon	45
Gambar 4. 6 Confusion Matrix SVM dengan Label Indobert	47
Gambar 4. 7 Confusion Matrix SVM dengan Label RoBERTa	50



DAFTAR LAMPIRAN

Lampiran 1. Kartu Asistensi	69
Lampiran 2. Curriculum Vitae	70
Lampiran 3. Surat Pernyataan & Pengalihan HAKI.....	71
Lampiran 4. Bukti BNSP	73

