

**ANALISIS SENTIMEN OPINI PUBLIK TERHADAP
KEBOCORAN DATA DI INDONESIA PADA MEDIA SOSIAL X
MENGUNAKAN SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

Yoga Prayudha

21.11.4433

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS SENTIMEN OPINI PUBLIK TERHADAP
KEBOCORAN DATA DI INDONESIA PADA MEDIA SOSIAL X
MENGUNAKAN SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
Yoga Prayudha
21.11.4433

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN OPINI PUBLIK TERHADAP
KEBOCORAN DATA DI INDONESIA PADA MEDIA SOSIAL X
MENGUNAKAN SUPPORT VECTOR MACHINE (SVM)**

yang disusun dan diajukan oleh

Yoga Prayudha

21.11.4433

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Agustus 2025

Dosen Pembimbing,


Robert Marco, S.T., M.T., Ph.D
NIK. 190302228

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN OPINI PUBLIK TERHADAP
KEBOCORAN DATA DI INDONESIA PADA MEDIA SOSIAL X
MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)

yang disusun dan diajukan oleh

Yoga Prayudha

21.11.4433

Telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Agustus 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ika Nur Fairi, S.Kom., M.Kom.
NIK. 190302268

Dwi Nurani, S.Kom., M.Kom.
NIK. 190302236

Robert Marco, S.T., M.T., Ph.D.
NIK. 190302228



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Yoga Prayudha
NIM : 21.11.4433

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Opini Publik Terhadap Kebocoran Data di Indonesia Pada Media Sosial X Menggunakan Support Vector Machine (SVM)

Dosen Pembimbing : Robert Marco, S.T., M.T., Ph.D

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 26 Agustus 2025

Yang Menyatakan,



Yoga Prayudha

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur, saya ingin mengucapkan terimakasih terimah kasih yang sebesar-besarnya kepada kedua orang tua saya atas segala doa, kasih sayang, semangat, serta dukungan moral dan finansial yang tak pernah putus selama saya menempuh pendidikan ini. Setiap pengorbanan dan ketulusan yang di berikan telah menjadi sumber kekuatan yang mendorong saya untuk terus berjuang hingga akhirnya dapat mencapai titik pencapaian ini. Tanpa kehadiran, doa, dan dukungan mereka, keberhasilan ini tidak akan pernah terwujud.

Ucapan terima kasih yang tulus saya sampaikan kepada seluruh dosen dan staf Universitas Amikom Yogyakarta atas ilmu, bimbingan, dan fasilitas yang telah mendukung proses belajar saya. Dedikasi dan perhatian yang diberikan menjadi bekal berharga dalam perjalanan akademik saya hingga tugas akhir ini dapat terselesaikan.

Tak lupa, saya mengucapkan terimakasih kepada teman-teman seperjuangan atas kebersamaan, dukungan, bantuan, dan semangat yang di berikan sepanjang perjalanan perkuliahan hingga penyusunan tugas akhir ini. Kehadiran kalian menjadikan perjalanan ini lebih bermakna dan penuh kenangan.

Terakhir, saya juga menyampaikan penghargaan dan terima kasih kepada semua pihak yang tidak dapat saya sebutkan satu per satu, baik secara langsung maupun tidak langsung. Semoga segala kebaikan yang telah kalian berikan dibalas dengan keberkahan dan balasan yang berlipat ganda.

Dengan penuh rasa syukur dan kerendahan hati, karya ini saya persembahkan.

KATA PENGANTAR

Segala puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, sehingga skripsi berjudul "Analisis Sentimen Opini Publik Terhadap Kebocoran Data di Indonesia Pada Media Sosial X Menggunakan Support Vector Machine (SVM)" dapat terselesaikan dengan baik. Saya berharap hasil penelitian ini dapat memberikan manfaat bagi berbagai pihak, khususnya dalam pengembangan ilmu di bidang analisis sentimen dan pemrosesan bahasa alami (NLP).

Penelitian ini disusun sebagai salah satu syarat untuk menyelesaikan program studi Teknik Komputer di Universitas Amikom Yogyakarta. Dalam proses penyusunan skripsi ini, saya menyadari bahwa pencapaian ini tidak akan terwujud tanpa bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat, saya ingin mengucapkan terima kasih kepada:

1. Bapak, Ibu, serta Keluarga Besar saya atas dukungan moral dan kasih sayang yang tak pernah putus tiada hentinya.
2. Bapak Robert Marco, S.T., M.T., Ph.D. yang memberikan banyak arahan, bimbingan, dan motivasi selama proses penyusunan penelitian ini.
3. Teman-teman terdekat yang telah memberikan semangat, bantuan, dan kebersamaan yang berarti selama proses penyusunan skripsi ini.
4. Tak lupa, terima kasih juga untuk diri sendiri yang sudah berjuang sejauh ini. Meski sering merasa lelah dan ingin menyerah, akhirnya tetap memilih untuk melangkah dan menyelesaikan perjalanan ini.

Demikian kata pengantar ini Terima Kasih untuk semuanya.

Yogyakarta, 5 agustus 2025

Penulis

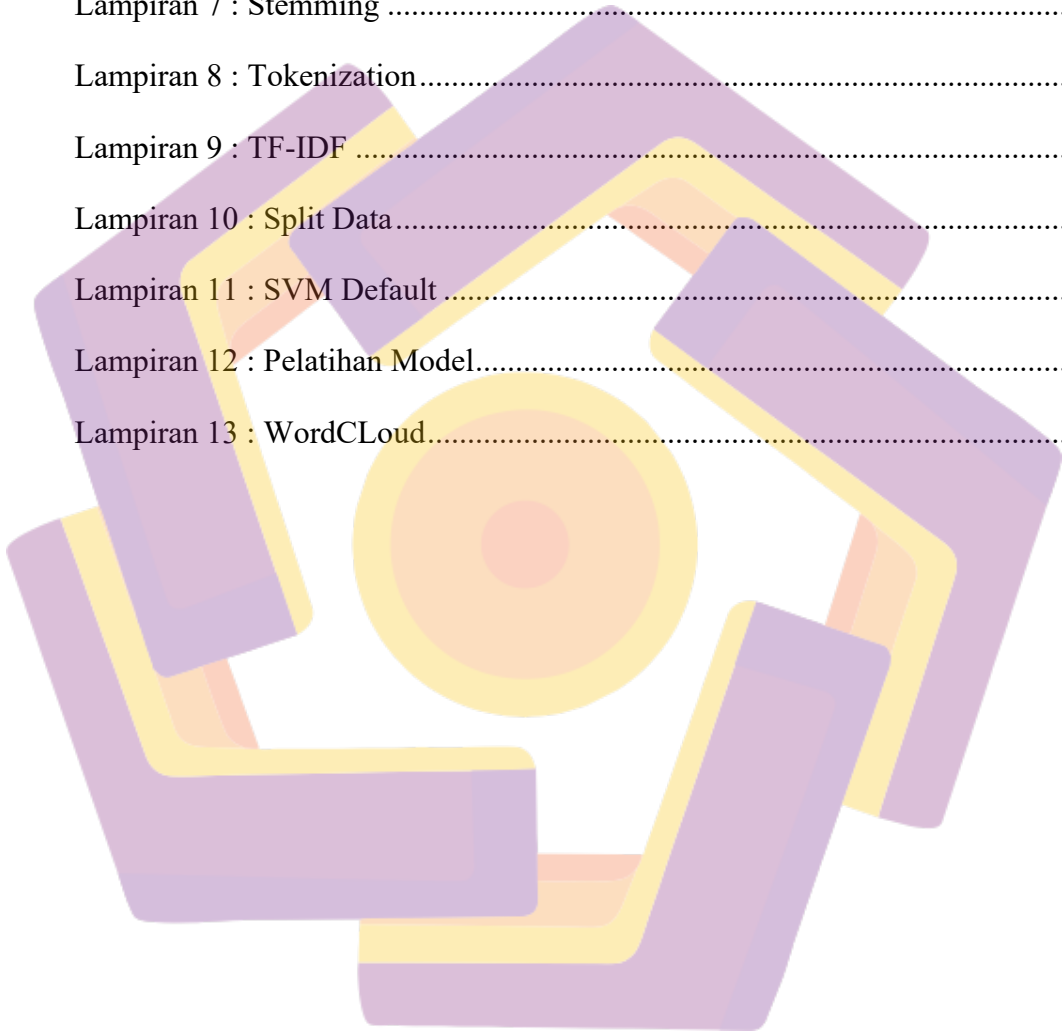
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH.....	xv
INTISARI	xvii
<i>ABSTRACT</i>	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6

2.1	Studi Literatur	6
2.2	Dasar Teori.....	16
2.2.1	Data Mining	16
2.2.2	Natural Language Processing (NLP)	16
2.2.3	Analisis Sentimen	17
2.2.4	X (Twitter)	18
2.2.5	Kebocoran Data di Indonesia	19
2.2.6	Support Vector Machine (SVM).....	19
2.2.7	Tweet Harvest	22
2.2.8	Labeling	22
2.2.9	Text Processing.....	24
2.2.10	TF-IDF	26
2.2.11	GridSearchCV	26
BAB III METODE PENELITIAN		28
3.1	Objek Penelitian.....	28
3.1.1	Objek Penelitian.....	28
3.1.2	Jenis Data	28
3.1.3	Sumber Data.....	28
3.1.4	Teknik Pengambilan Data.....	28
3.2	Alur Penelitian	29
3.2.1	Pengambilan Dataset.....	29
3.2.2	Labeling	30
3.2.3	Preprocessing Data.....	31
3.2.4	Cleaning	31
3.2.5	Case Folding	32
3.2.6	Stopword Removal.....	33
3.2.7	Stemming	33
3.2.8	Tokenization	34
3.2.9	TF-IDF	35
3.2.10	Split Data	36

3.2.11	SVM.....	36
3.2.12	GridSearchCV	37
3.2.13	Evaluasi Performa	37
3.3	Alat dan Bahan.....	38
3.3.1	Perangkat Keras	39
3.3.2	Perangkat Lunak	39
3.3.3	Tools dan Library	39
BAB IV HASIL DAN PEMBAHASAN		40
4.1	Pengambilan Dataset.....	40
4.2	Labeling	40
4.3	Preprocessing Data.....	41
4.3.1	Cleaning	41
4.3.2	Case-Folding	42
4.3.3	<i>Stopword Removal</i>	43
4.3.4	Stemming	43
4.3.5	Tokenization	44
4.4	TF-IDF	44
4.5	Split Data	45
4.6	SVM.....	45
4.7	Evaluasi Performa	45
BAB V PENUTUP		53
5.1	Kesimpulan	53
5.2	Saran	54
REFERENSI		55
Lampiran 1 : Scraping Dataset.....		58
Lampiran 2 : Combined Dataset		61

Lampiran 3 : Labeling.....	61
Lampiran 4 : Cleaning	62
Lampiran 5 : Case Folding.....	63
Lampiran 6 : Stopword Removal.....	63
Lampiran 7 : Stemming	64
Lampiran 8 : Tokenization.....	64
Lampiran 9 : TF-IDF	65
Lampiran 10 : Split Data.....	65
Lampiran 11 : SVM Default	66
Lampiran 12 : Pelatihan Model.....	67
Lampiran 13 : WordCLoud.....	68

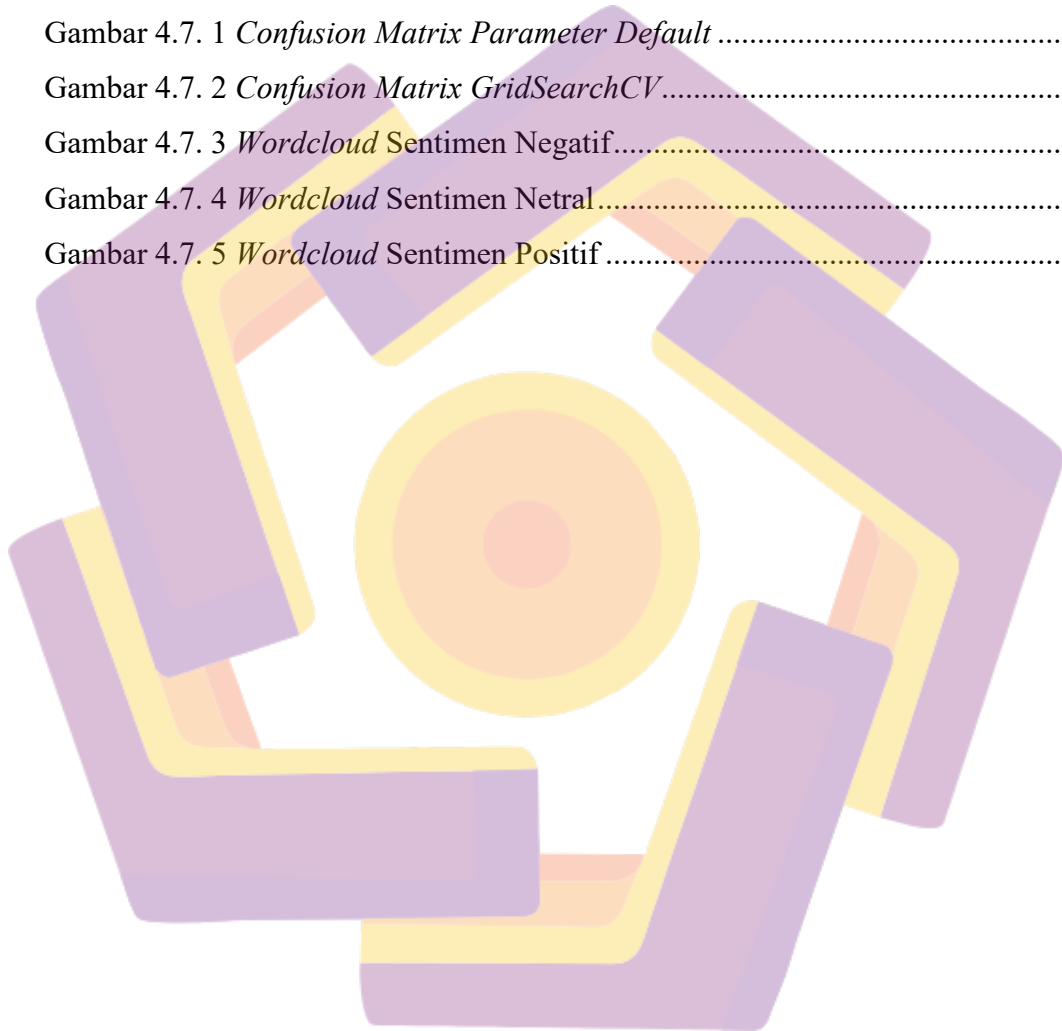


DAFTAR TABEL

Tabel 2.1 1 Keaslian Penelitian	11
Tabel 3.2.1. 1 Contoh Data	30
Tabel 3.2.1. 2 Data Setelah <i>Labeling</i>	31
Tabel 3.2.1. 3 Data Setelah <i>Cleaning</i>	32
Tabel 3.2.1. 4 Data Setelah <i>Case Folding</i>	32
Tabel 3.2.1. 5 Data Setelah <i>Stopword Removal</i>	33
Tabel 3.2.1. 6 Data Setelah <i>Stemming</i>	34
Tabel 3.2.1. 7 Data Setelah <i>Tokenization</i>	34
Tabel 3.2.1. 8 Data Setelah TF-IDF	35
Tabel 3.2.1. 9 Parameter Algoritma SVM	36
Tabel 3.2.1. 10 Parameter Model GridSearchCV	37
Tabel 4.1. 1 Dataset	40
Tabel 4.2. 1 Tahap <i>Labeling</i>	41
Tabel 4.3.1. 1 Tahap <i>Cleaning</i>	42
Tabel 4.3.2. 1 Tahap <i>Case Folding</i>	42
Tabel 4.3.3. 1 Tahap <i>Stopword Removal</i>	43
Tabel 4.3.4. 1 Tahap <i>Stemming</i>	43
Tabel 4.3.5. 1 Tahap <i>Tokenization</i>	44
Tabel 4.7.1. 1 Hasil Pengujian Parameter <i>Default</i>	47
Tabel 4.7.2. 1 Hasil Pengujian GridSearchCV	48
Tabel 4.7.2. 2 Perbandingan Akurasi Algoritma	49

DAFTAR GAMBAR

Gambar 2.2.1. 1 <i>Natural Launguage Processing</i>	17
Gambar 2.2.1. 2 <i>Support Vector Machine (SVM)</i>	20
Gambar 2.2.1. 3 <i>Radial Basis Function (RBF)</i>	21
Gambar 3.2. 1 Alur Penelitian	29
Gambar 3.2. 2 Confusion Matrix	38
Gambar 4.7. 1 <i>Confusion Matrix Parameter Default</i>	46
Gambar 4.7. 2 <i>Confusion Matrix GridSearchCV</i>	48
Gambar 4.7. 3 <i>Wordcloud</i> Sentimen Negatif.....	50
Gambar 4.7. 4 <i>Wordcloud</i> Sentimen Netral.....	51
Gambar 4.7. 5 <i>Wordcloud</i> Sentimen Positif	51

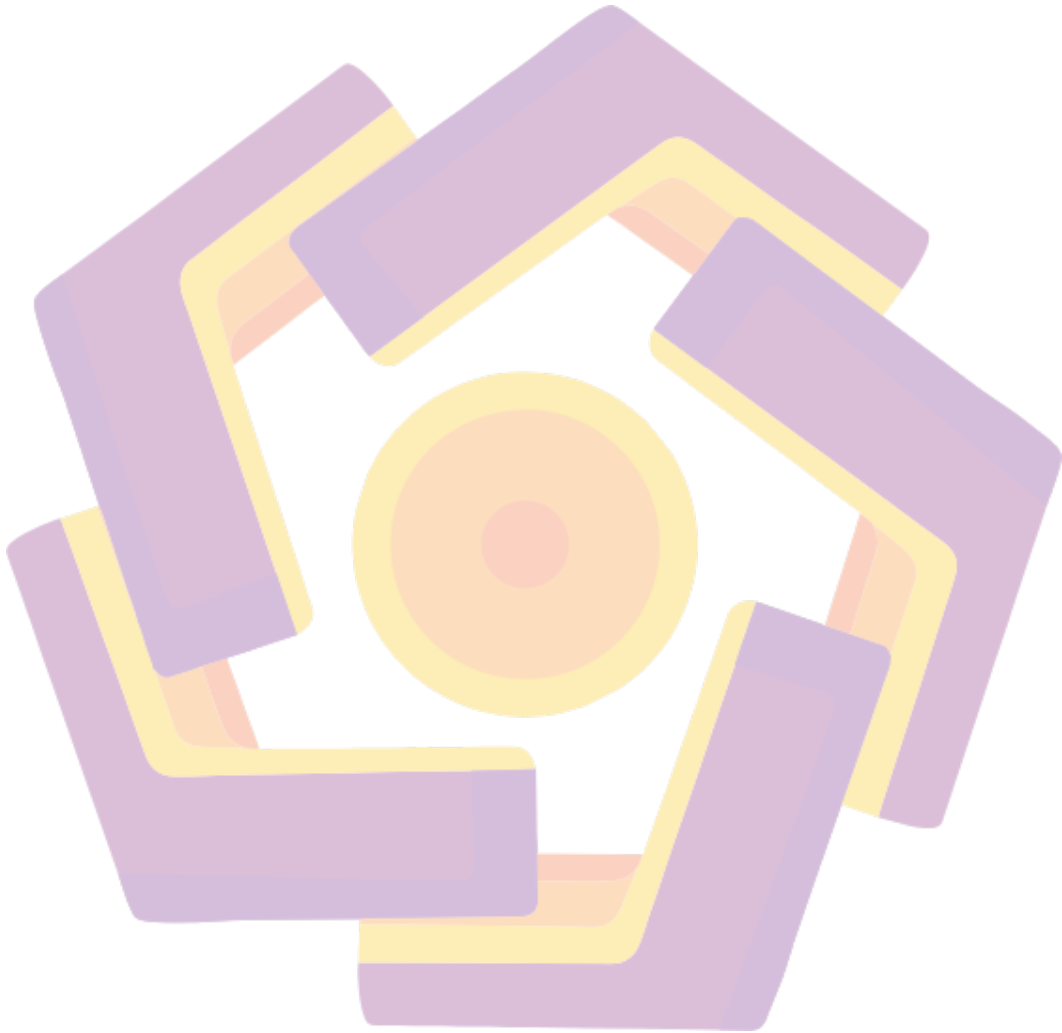


DAFTAR LAMPIRAN

Lampiran 1. 1 Scraping Dataset.....	58
Lampiran 1. 2 Scraping Dataset.....	58
Lampiran 1. 3 Scraping Dataset.....	59
Lampiran 1. 4 Scraping Dataset.....	59
Lampiran 1. 5 Scraping Dataset.....	60
Lampiran 1. 6 Scraping Dataset.....	60
Lampiran 2. 1 Combine Dataset	61
Lampiran 3. 1 Labeling.....	61
Lampiran 3. 2 Labeling.....	62
Lampiran 4. 1 Cleaning.....	62
Lampiran 5. 1 Case Folding.....	63
Lampiran 5. 2 Case Folding.....	63
Lampiran 6. 1 Stopword Removal.....	63
Lampiran 6. 2 Stopword Removal.....	64
Lampiran 7. 1 Stemming.....	64
Lampiran 8. 1 Tokenization.....	64
Lampiran 9. 1 TF-IDF.....	65
Lampiran 10. 1 Split Data.....	65
Lampiran 11. 1 SVM Default	66
Lampiran 11. 2 SVM Default	66
Lampiran 12. 1 Pelatihan Model.....	67
Lampiran 12. 2 Pelatihan Model.....	67
Lampiran 13. 1 WordCloud	68
Lampiran 13. 2 WordCloud	68
Lampiran 13. 3 WordCloud Positif.....	68
Lampiran 13. 4 WordCloud Negatif.....	69
Lampiran 13. 5 WordCloud Netral	69

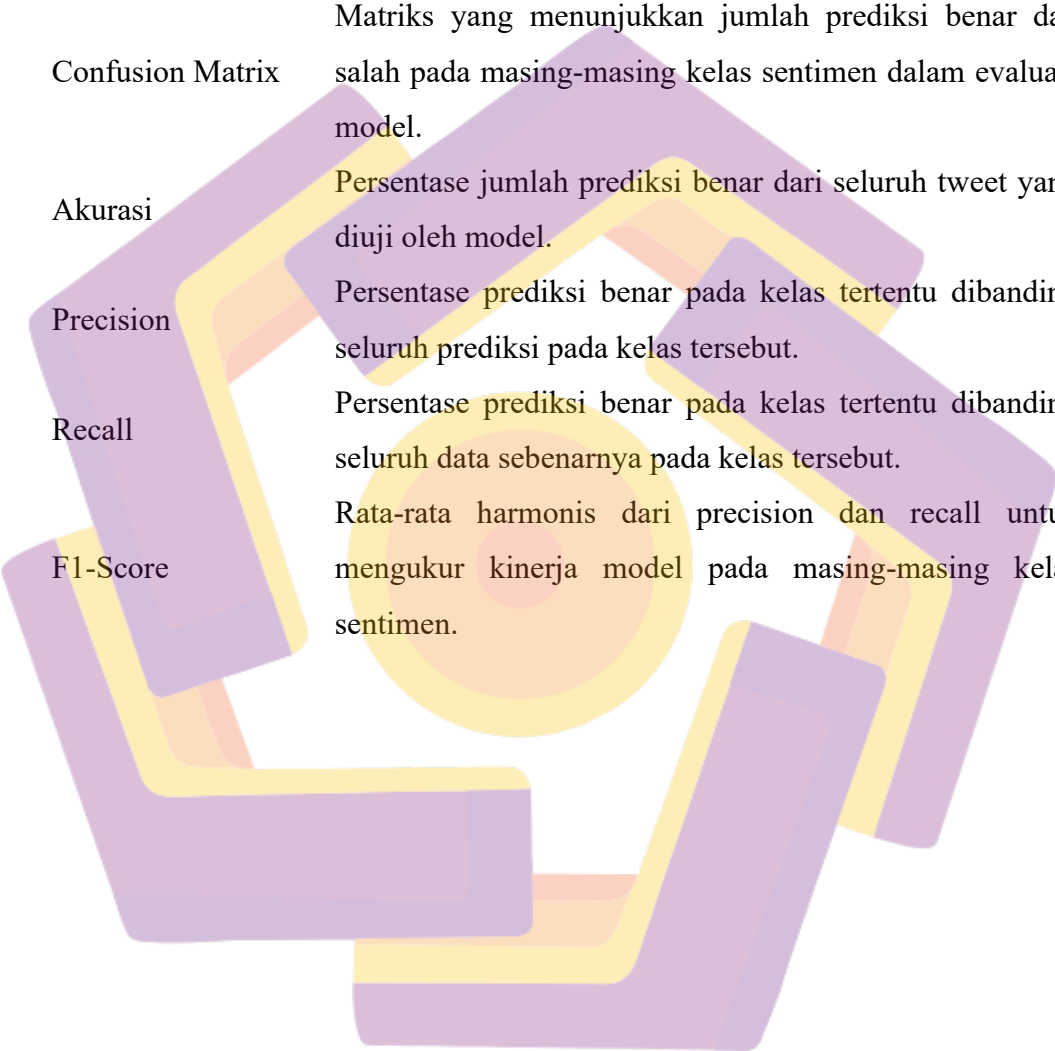
DAFTAR LAMBANG DAN SINGKATAN

TP	True Positive, jumlah prediksi benar pada kelas positif.
FP	False Positive, jumlah prediksi salah pada kelas positif.
TN	True Negative, jumlah prediksi benar pada kelas negatif.
FN	False Negative, jumlah prediksi salah pada kelas negatif.



DAFTAR ISTILAH

Vektor	Representasi numerik dari data teks hasil ekstraksi fitur TF-IDF, digunakan sebagai input pada model SVM.
Hyperplane	Garis atau bidang pemisah yang dibentuk oleh SVM untuk memisahkan kelas sentimen positif, negatif, dan netral.
Margin	Jarak antara hyperplane dan data terdekat (support vector) pada masing-masing kelas sentimen.
Support Vector	Titik data pada batas margin yang digunakan SVM untuk menentukan posisi hyperplane.
Kernel RBF	Fungsi kernel yang digunakan untuk memetakan data teks yang tidak terpisah secara linier menjadi ruang berdimensi lebih tinggi agar dapat dipisahkan oleh hyperplane.
Cleaning	Tahap awal preprocessing yang menghapus elemen tidak relevan pada tweet, seperti tanda baca, angka, username, URL, dan simbol.
Case Folding	Proses mengubah semua huruf dalam tweet menjadi huruf kecil untuk menyamakan format kata.
Stopword Removal	Penghapusan kata-kata umum dalam bahasa Indonesia yang tidak memengaruhi penentuan sentimen, seperti "dan", "yang", "di".
Stemming	Proses mengubah kata berimbuhan dalam tweet menjadi bentuk kata dasar menggunakan algoritma stemming bahasa Indonesia.
Tokenization	Pemisahan tweet menjadi daftar kata (token) untuk memudahkan pembobotan kata.
TF-IDF	Metode pembobotan kata pada tweet berdasarkan frekuensi kata dalam dokumen (tweet) dan kebalikannya di seluruh korpus.



Labeling	Proses manual pemberian label sentimen positif, negatif, atau netral pada setiap tweet sebelum pelatihan model.
GridSearchCV	Metode pencarian kombinasi hyperparameter SVM terbaik dengan menguji berbagai nilai parameter C, gamma, kernel, dan class weight.
Confusion Matrix	Matriks yang menunjukkan jumlah prediksi benar dan salah pada masing-masing kelas sentimen dalam evaluasi model.
Akurasi	Persentase jumlah prediksi benar dari seluruh tweet yang diuji oleh model.
Precision	Persentase prediksi benar pada kelas tertentu dibanding seluruh prediksi pada kelas tersebut.
Recall	Persentase prediksi benar pada kelas tertentu dibanding seluruh data sebenarnya pada kelas tersebut.
F1-Score	Rata-rata harmonis dari precision dan recall untuk mengukur kinerja model pada masing-masing kelas sentimen.

INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen opini publik terhadap kebocoran data di Indonesia yang dibagikan melalui media sosial X (Twitter) menggunakan algoritma *Support Vector Machine* (SVM). Data dikumpulkan melalui metode *web scraping* dan API dengan kata kunci yang relevan, kemudian dilakukan proses labeling menjadi tiga kategori sentimen, yaitu positif, negatif, dan netral. Tahap *preprocessing* meliputi *cleaning*, *case folding*, *stopword removal*, *stemming*, *tokenization*, dan pembobotan TF-IDF. Model SVM diuji menggunakan dua pendekatan, yaitu parameter *default* dan parameter hasil optimasi menggunakan *GridSearchCV*, dengan tujuan menemukan kombinasi *hyperparameter* terbaik. Evaluasi performa dilakukan menggunakan *confusion matrix* untuk menghitung akurasi, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa baik model SVM dengan parameter default maupun yang dioptimasi melalui *GridSearchCV* mencapai akurasi tertinggi sebesar 75%. Perbedaan terlihat pada distribusi klasifikasi: model *default* mampu mengklasifikasikan sebagian sentimen positif, sedangkan model *GridSearchCV* cenderung gagal mendeteksi kelas positif namun memberikan kinerja lebih konsisten pada kelas negatif dan netral. Temuan ini menunjukkan bahwa optimasi *hyperparameter* tidak selalu meningkatkan performa keseluruhan jika distribusi data tidak seimbang. Hasil penelitian ini memberikan gambaran tentang persepsi publik di Indonesia terkait isu kebocoran data, di mana sentimen negatif mendominasi. Informasi ini dapat dimanfaatkan oleh pemerintah, lembaga perlindungan data, dan penyedia layanan digital untuk merumuskan kebijakan keamanan data yang lebih responsif. Selain itu, metode dan hasil penelitian ini dapat menjadi referensi bagi pengembangan sistem pemantauan opini publik berbasis kecerdasan buatan yang mampu bekerja secara otomatis terhadap data media sosial yang bersifat dinamis.

Kata kunci : Analisis Sentimen, Kebocoran Data, Media Sosial X, Support Vector Machine, GridSearchCV

ABSTRACT

This study aims to analyze public opinion sentiment regarding data breaches in Indonesia shared on the X (Twitter) social media platform using the Support Vector Machine (SVM) algorithm. Data were collected through web scraping and API methods with relevant keywords, followed by labeling into three sentiment categories: positive, negative, and neutral. The preprocessing stage included cleaning, case folding, stopword removal, stemming, tokenization, and TF-IDF weighting. The SVM model was tested using two approaches: default parameters and optimized parameters via GridSearchCV, aiming to find the best hyperparameter combination. Performance evaluation was conducted using a confusion matrix to calculate accuracy, precision, recall, and F1-score. The results showed that both the default SVM and the GridSearchCV-optimized SVM achieved the highest accuracy of 75%. However, differences were observed in classification distribution: the default model was able to classify some positive sentiments, while the GridSearchCV model tended to fail in detecting positive classes but delivered more consistent performance for negative and neutral classes. These findings indicate that hyperparameter optimization does not always improve overall performance, especially when the dataset is imbalanced. This research provides an overview of public perception in Indonesia regarding data breach issues, where negative sentiments dominate. Such insights can be utilized by government agencies, data protection institutions, and digital service providers to formulate more responsive data security policies. Furthermore, the methodology and findings of this study can serve as a reference for developing AI-based public opinion monitoring systems capable of automatically processing dynamic social media data.

Keywords: *Sentiment Analysis, Data Breach, X Social Media, Support Vector Machine, GridSearchCV*