

**ANALISIS SENTIMEN PUBLIK TERHADAP KEBOCORAN
DATA KOMINFO PADA TWITTER MENGGUNAKAN
MODEL KLASIFIKASI**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

ECHA AMELYA PUTRI

21.12.2128

Kepada

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

**ANALISIS SENTIMEN PUBLIK TERHADAP
KEBOCORANDATA KOMINFO PADA TWITTER
MENGUNAKAN MODEL KLASIFIKASI**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

ECHA AMELYA PUTRI

21.12.2128

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN PUBLIK TERHADAP KEBOCORAN DATA
KOMINFO PADA TWITTER MENGGUNAKAN MODEL KLASIFIKASI**

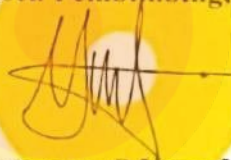
yang disusun dan diajukan oleh

Echa Amelya Putri

21.12.2128

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 17 Februari 2025

Dosen Pembimbing,



Yoga Pristyanto, S.Kom, M.Eng

NIK. 190302412

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN PUBLIK TERHADAP KEBOCORAN DATA
KOMINFO PADA TWITTER MENGGUNAKAN MODEL KLASIFIKI

yang disusun dan diajukan oleh

Echa Amelya Putri

21.12.2128

Telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Februari 2025

Susunan Dewan Penguji

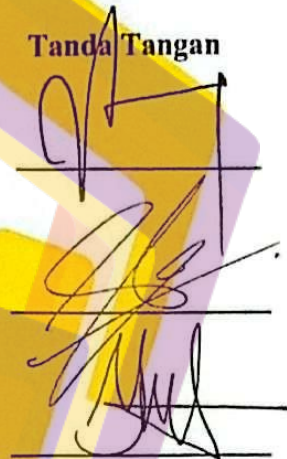
Nama Penguji

Tanda Tangan

Ir. Rum Mohamad Andri K Rasyid, M.Kom
NIK. 190302011

Agung Nugroho, S.Kom., M.Kom
NIK. 190302242

Yoga Pristvanto, S.Kom, M.Eng
NIK. 190302412



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Echa Amelya Putri
NIM : 21.12.2128

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Publik terhadap Kebocoran Data KOMINFO pada Twitter Menggunakan Model Klasifikasi

Dosen Pembimbing : Yoga Pristyanto, S.KOM, M.Eng

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, 17 Februari 2025

Yang Menyatakan,



Echa Amelya Putri

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT yang senantiasa memberikan rahmat, hidayah, dan kekuatan dalam setiap langkah hidup saya, akhirnya saya dapat menyelesaikan skripsi ini. Skripsi ini saya persembahkan kepada:

1. Kedua Orang tua tercinta, yang selalu memberikan doa, dukungan, serta semangat tanpa henti dalam setiap proses saya. Pencapaian ini terwujud atas doa dan perjuangan kalian.
2. Kedua Saudara tersayang saya, yang selalu memberi dukungan dan nasehat dalam menuntut ilmu.
3. Dosen pembimbing saya, yang dengan sabar membimbing serta arahan yang sangat berarti hingga saya bisa menyelesaikan skripsi ini.
4. Teman-teman terbaik saya, yang selalu menemani dan memberikan semangat selama masa perkuliahan.
5. Pasangan tersayang saya, yang selalu mendengarkan keluh kesah dan selalu menemani saya.
6. Diri saya, yang sudah sabar bertahan sejauh ini. Semoga ini menjadi awal dari perjalanan baru yang lebih baik ke depannya.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan Skripsi yang berjudul “Analisis Sentimen Publik terhadap Kebocoran Data Kominfo pada Twitter menggunakan Model Klasifikasi”. Skripsi ini disusun untuk memenuhi salah satu syarat kelulusan Program Sarjana pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selain itu penulis dengan segala kerendahan hati ingin menyampaikan rasa terima kasih kepada semua pihak yang telah berjasa memberikan dukungan dan bantuan untuk menyelesaikan laporan tugas akhir ini. Untuk itu penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta
2. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Program Fakultas Ilmu Komputer
3. Bapak Anggit Dwi Hartanto, M.Kom selaku Ketua Program Studi Sistem Informasi
4. Bapak Yoga Pristyanto, S.Kom, M.Eng selaku dosen pembimbing yang memberikan arahan, saran, dan motivasi terhadap penulis
5. Kedua orang tua tercinta yang memberikan semangat dan doa kepada penulis.
6. Teman-teman yang selalu mendukung dan menyemangati dalam penyelesaian skripsi ini.

Yogyakarta, 17 Februari 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.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	2
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA	4
2.1 Studi Literatur	4
2.2 Dasar Teori	12

2.2.1	Analisis Sentimen	12
2.2.2	Twitter.....	12
2.2.3	Text Mining	12
2.2.4	Klasifikasi	12
2.2.5	Text Preprocessing.....	13
2.2.5	Labeling	14
2.2.6	Pembobotan TF- IDF	14
2.2.7	Split data	14
2.2.6	Naïve Bayes Classifier	14
2.2.7	Random Forest	15
2.2.8	Logistic Regression.....	16
2.2.9	Evaluasi Model	16
BAB III METODE PENELITIAN		19
3.1	Objek Penelitian	19
3.2	Alur Penelitian.....	19
3.2.1	Scraping	21
3.2.2	Preprocessing	21
3.2.3	Labelling	25
3.2.4	TF-IDF	25
3.2.5	Split Data	27
3.2.6	Naïve Bayes	28
3.2.7	Random Forest	28
3.2.8	Logistic Regression.....	28
3.2.9	Evaluasi.....	28
3.3	Alat dan Bahan	29

BAB IV HASIL DAN PEMBAHASAN	30
4.1 Scraping Data	30
4.2 Preprocessing	31
4.2.1 Cleaning	31
4.2.2 Case Folding	33
4.2.3 Tokenization	34
4.2.4 Stopword Removal.....	36
4.2.5 Stemming	37
4.3 Labelling.....	38
4.4 Pembobotan TF-IDF	42
4.5 Split Data.....	45
4.6 Naïve Bayes.....	46
4.7 Random Forest	48
4.8 Logistic Regression	51
4.9 Evaluasi	54
4.9.1 Acuan Evaluasi	54
4.9.2 Matrix Evaluasi.....	54
4.10 Implementasi Model	57
BAB V PENUTUP	65
5.1 Kesimpulan.....	65
5.2 Saran.....	65
REFERENSI	66

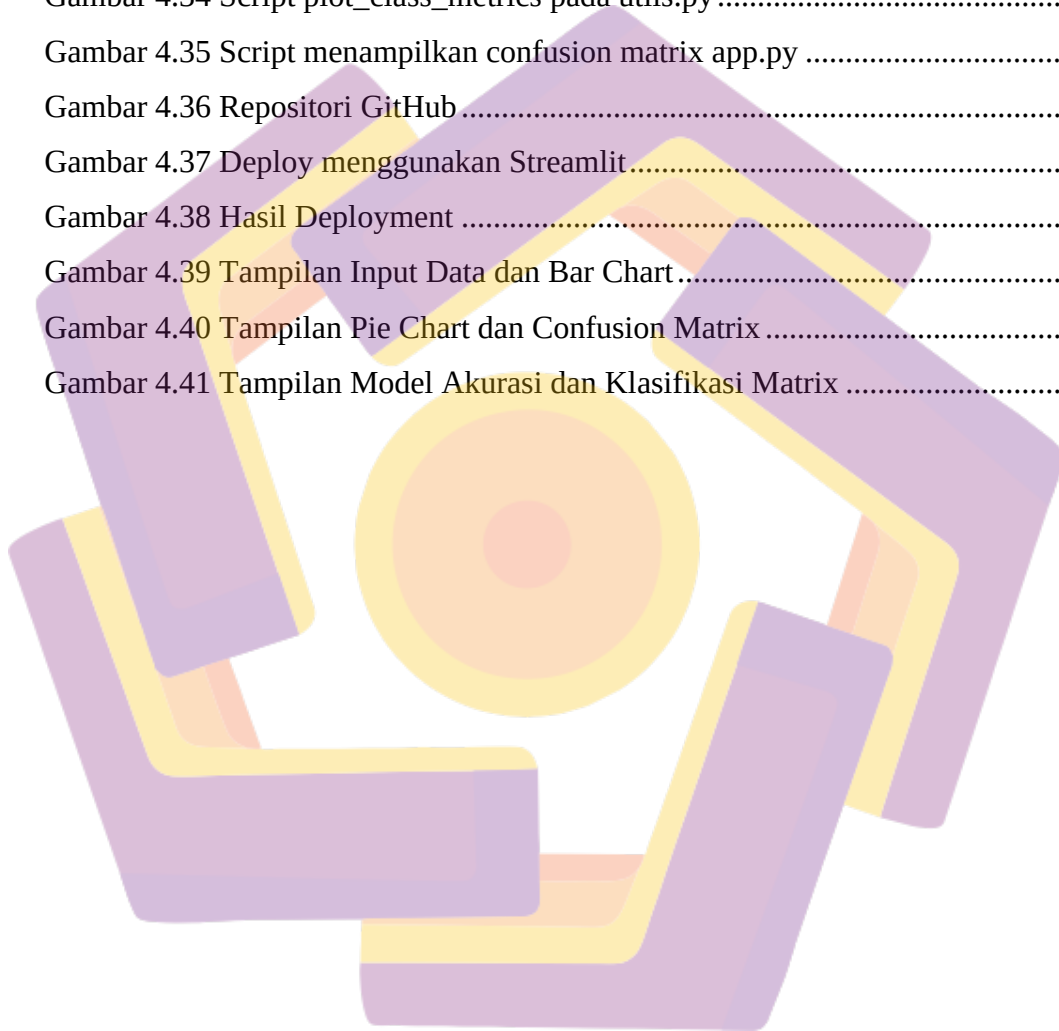
DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	7
Tabel 2.2 Confusion Matrix	16
Tabel 3.1 Contoh Tahap Cleaning	21
Tabel 3.2 Contoh Tabel Case Folding	22
Tabel 3.3 Contoh Tabel Tokenization	23
Tabel 3.4 Contoh Tabel Stopword Removal.....	24
Tabel 3.5 Contoh Tabel Stemming	25
Tabel 3.6 Teks Tweet.....	26
Tabel 3.7 Contoh Tabel TF-IDF	26
Table 4.1 Tahap Cleaning	32
Table 4.2 Tahap Case Folding	33
Table 4.3 Tahap Tokenization	35
Table 4.4 Tahap Stopword Removal.....	36
Table 4.5 Tahap Stemming	37
Table 4.6 Pembobotan TF-IDF	44
Table 4.7 Split Data	45
Table 4.8 Hasil Confusion Matrix Naive Bayes	46
Table 4.9 Hasil Confusion Matrix Random Forest	49
Table 4.10 Confusion Matrix Logistic Regression	52
Table 4.11 Perbandingan Performa Model Klasifikasi.....	55

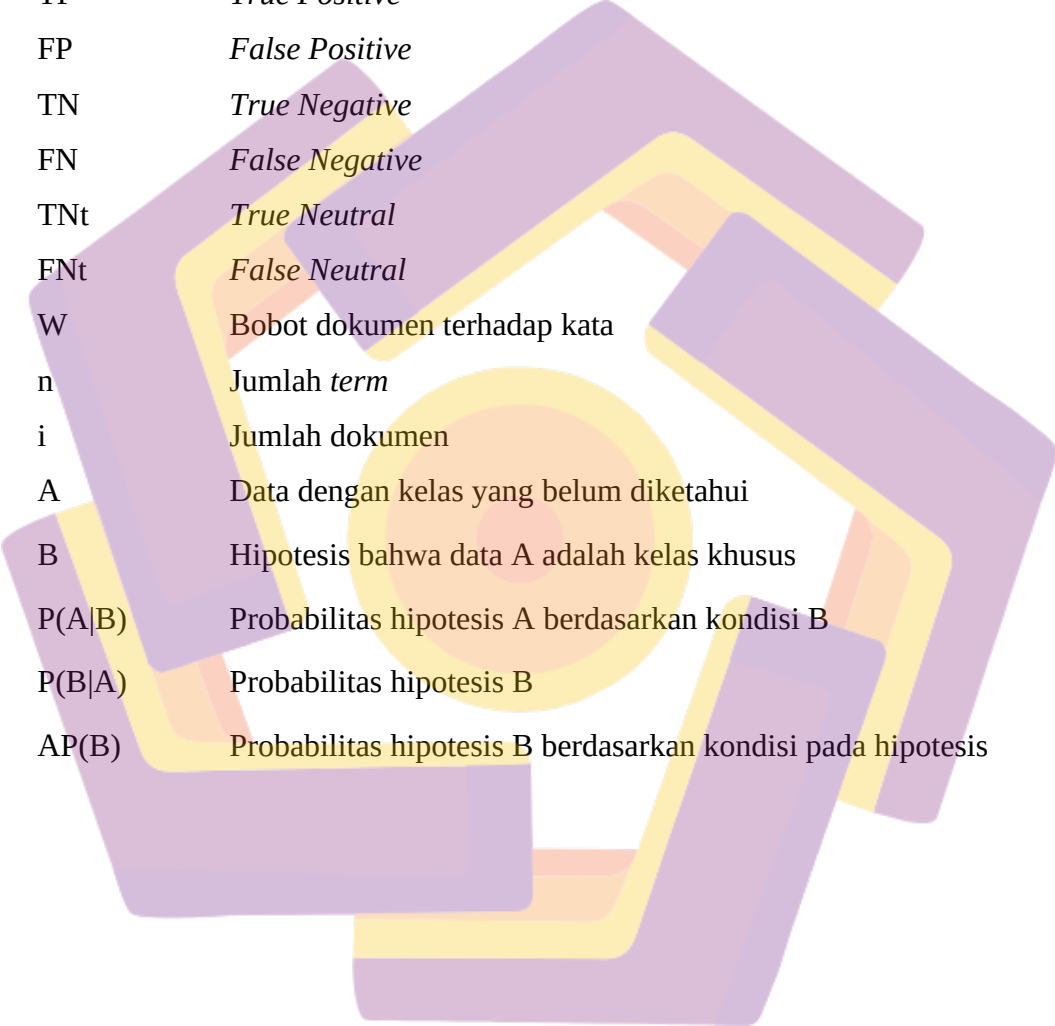
DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	20
Gambar 4.1 Script Tahap Scraping.....	30
Gambar 4.2 Hasil Scraping data Tweet Kebocoran Data Kominfo	31
Gambar 4.3 Script Tahap Cleaning.....	32
Gambar 4.4 Script Tahap Case Folding.....	33
Gambar 4.5 Script Tahap Tokenization	35
Gambar 4.6 Script Tahap Stopword Removal	36
Gambar 4.7 Script Tahap Stemming.....	37
Gambar 4.8 Script Tahap Labelling	39
Gambar 4.9 Hasil Labelling	39
Gambar 4.10 Script Pie Chart Labelling.....	39
Gambar 4.11 Pie Chart HasilLabelling	40
Gambar 4.12 Script Word Cloud	40
Gambar 4.13 Word Cloud Positif	41
Gambar 4.14 Word Cloud Negatif.....	41
Gambar 4.15 Word Cloud Netral.....	41
Gambar 4.16 Script Top 20 Kata Tertinggi	42
Gambar 4.17 Top 20 Kata Tertinggi.....	43
Gambar 4.18 Script Word Cloud 20 Kata Tertinggi.....	43
Gambar 4.19 Word Cloud Top 20 Kata TF- IDF	44
Gambar 4.20 Script Split Data	45
Gambar 4.21 Script Naïve Bayes.....	46
Gambar 4.22 Klasifikasi Naïve Bayes	46
Gambar 4.23 Script Random Forest.....	49
Gambar 4. 24 Klasifikasi Random Forest.....	49
Gambar 4.25 Script Logistic Regression	51
Gambar 4.26 Klasifikasi Logistic Regression.....	52
Gambar 4.27 Evaluasi Model.....	55
Gambar 4.28 Script Perbandingan Akurasi Model	56

Gambar 4.29 Perbandingan Model Akurasi.....	57
Gambar 4.30 Script mengimpor Pustaka utils.py	57
Gambar 4.31 Script mengimpor Pustaka app.py.....	58
Gambar 4.32 Script Fungsi determine pada utils.app	58
Gambar 4.33 Script Determine pada app.py	59
Gambar 4.34 Script plot_class_metrics pada utils.py	59
Gambar 4.35 Script menampilkan confusion matrix app.py	60
Gambar 4.36 Repositori GitHub	60
Gambar 4.37 Deploy menggunakan Streamlit.....	61
Gambar 4.38 Hasil Deployment	61
Gambar 4.39 Tampilan Input Data dan Bar Chart.....	62
Gambar 4.40 Tampilan Pie Chart dan Confusion Matrix	63
Gambar 4.41 Tampilan Model Akurasi dan Klasifikasi Matrix	64



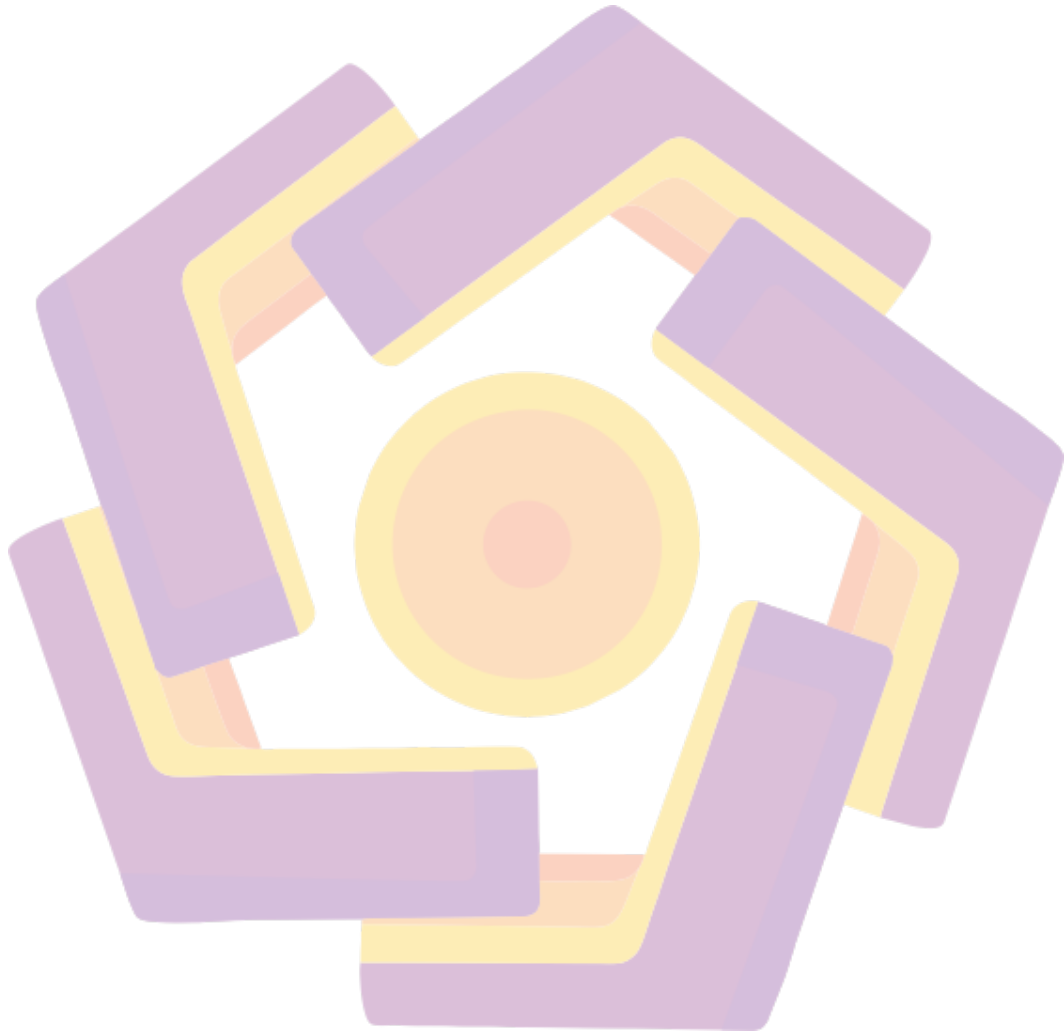
DAFTAR LAMBANG DAN SINGKATAN



NLP	<i>Natural Language Processing</i>
TF- IDF	<i>Term Frequency - Inverse Document Frequency</i>
CSV	<i>Comma-Separated Values</i>
TP	<i>True Positive</i>
FP	<i>False Positive</i>
TN	<i>True Negative</i>
FN	<i>False Negative</i>
TNt	<i>True Neutral</i>
FNt	<i>False Neutral</i>
W	Bobot dokumen terhadap kata
n	Jumlah term
i	Jumlah dokumen
A	Data dengan kelas yang belum diketahui
B	Hipotesis bahwa data A adalah kelas khusus
$P(A B)$	Probabilitas hipotesis A berdasarkan kondisi B
$P(B A)$	Probabilitas hipotesis B
$AP(B)$	Probabilitas hipotesis B berdasarkan kondisi pada hipotesis

DAFTAR ISTILAH

Dataset	Kumpulan data
Library	Perpustakaan
Word Cloud	Visual sekumpulan kata



INTISARI

Kebocoran data Kementrian dan Informatika (Kominfo) menjadi topik perbincangan yang hangat di kalangan masyarakat. Kejadian ini menimbulkan kekhawatiran publik terkait keamanan data pribadi, terutama dengan meningkatnya penggunaan teknologi digital dan layanan online yang mengandalkan penyimpanan data. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kebocoran data Kominfo di Twitter menggunakan tiga model klasifikasi :Naïve Bayes, Random Forest, dan Logistic Regression. Dalam Penelitian ini, tweet yang berkaitan dengan kebocoran data Kominfo dikategorikan ke dalam tiga sentimen utama, yaitu positif, netral, atau negatif. Hasil analisis sentimen menunjukkan bahwa dari ketiga model klasifikasi yang diuji, Logistic Regression memiliki performa terbaik dalam mengklasifikasikan sentimen publik dengan akurasi sebesar 89%. Model ini mencapai presisi sebesar 75% untuk sentimen positif, 44% untuk netral, dan 91% untuk negatif. Recall yang diperoleh adalah 21% untuk positif, 21% untuk netral, dan 100% untuk negatif, dengan F1-score masing-masing sebesar 19% untuk positif dan 29% untuk netral, dan 95% untuk negatif.

Kata kunci: Opini publik, Kebocoran Data, Kominfo, Analisis Sentimen, Twitter, Klasifikasi

ABSTRACT

The Ministry of Information and Communications (MOCI) data leak has become a hot topic of conversation among the public. This incident raises public concerns regarding the security of personal data, especially with the increasing use of digital technology and online services that rely on data storage. This study aims to analyze public sentiment towards the Kominfo data leak on Twitter using three classification models: Naïve Bayes, Random Forest, and Logistic Regression. In this study, tweets related to the Kominfo data leak were categorized into three main sentiments, namely positive, neutral, or negative. The sentiment analysis results show that of the three classification models tested, Logistic Regression has the best performance in classifying public sentiment with an accuracy of 89%. This model achieved a precision of 75% for positive sentiment, 44% for neutral, and 91% for negative. Recall obtained was 21% for positive, 21% for neutral, and 100% for negative, with F1-score of 19% for positive and 29% for neutral, and 95% for negative, respectively.

Keyword: Public opinion, Data leak, MOCI, Sentiment analysis, Twitter, Classification