

**ANALISIS SENTIMEN DATA TWITTER TERHADAP  
KEBOCORAN DATA PADA APLIKASI PEDULILINDUNGI  
MENGUNAKAN ALGORITMA NAÏVE BAYES**

**SKRIPSI**



diajukan oleh

**FAJAR MUHAMMAD SIDIQ**

**18.11.1895**

Kepada

**PROGRAM SARJANA**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2022**

**ANALISIS SENTIMEN DATA TWITTER TERHADAP  
KEBOCORAN DATA PADA APLIKASI PEDULILINDUNGI  
MENGUNAKAN ALGORITMA NAÏVE BAYES**

**SKRIPSI**

untuk memenuhi sebagian persyaratan mencapai gelar Sarjana  
pada Program Studi Informatika



diajukan oleh

**Fajar Muhammad Sidiq**

**18.11.1895**

Kepada

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2022**

# **PERSETUJUAN**

## **SKRIPSI**

### **ANALISIS SENTIMEN DATA TWITTER TERHADAP KEBOCORAN DATA PADA APLIKASI PEDULILINDUNGI MENGGUNAKAN ALGORITMA NAÏVE BAYES**

yang disusun dan diajukan oleh

**Fajar Muhammad Sidiq**

**18.11.1895**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 31 Juli 2022

**Dosen Pembimbing,**

**Yuli Astuti, M.Kom**

**NIK. 190302146**

# PENGESAHAN

## SKRIPSI

### ANALISIS SENTIMEN DATA TWITTER TERHADAP KEBOCORAN DATA PADA APLIKASI PEDULILINDUNGI MENGGUNAKAN ALGORITMA NAÏVE BAYES

yang disusun dan diajukan oleh

**Fajar Muhammad Sidiq**

**18.11.1895**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 19 Agustus 2022

#### Susunan Dewan Penguji

**Nama Penguji**

**Tanda Tangan**

**Rakhma Shafrida Kurnia, S.Kom., M.Kom**  
**NIK. 190302355**

**Donni Prabowo, M.Kom**  
**NIK. 190302253**

**Yuli Astuti, M.Kom**  
**NIK. 190302146**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 19 Agustus 2022

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta, S.Kom., M.Kom.**  
**NIK. 190302096**

## PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fajar Muhammad Sidiq  
NIM : 18.11.1895

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Data Twitter Terhadap Kebocoran Data Pada Aplikasi PeduliLindungi Menggunakan Algoritma Naïve Bayes**

Dosen Pembimbing : Yuli Astuti, M.Kom

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, 19 Agustus 2022

Yang Menyatakan,



Fajar Muhammad Sidiq

## PERSEMBAHAN

Dengan segala puji syukur kepada Allah SWT atas segala rahmat dan karunia-Nya dan atas doa dan dukungan dari orang-orang tercinta, akhirnya skripsi ini dapat selesai dengan baik dan tepat pada waktunya. Oleh karena itu dengan rasa bangga dan bahagia saya menyampaikan ucapan terimakasih sebesar-besarnya kepada:

1. Kedua orang tua saya telah bekerja keras dan selalu memberikan doa serta dukungan selama pengerjaan skripsi ini, dan juga telah memberikan fasilitas selama dalam menjalani perkuliahan hingga mendapat gelar sarjana komputer.
2. Kepada seluruh keluarga besar saya yang berada di Sumatra dan Jawa Tengah terutama kakek dan nenek di Muaradua, serta mbah di Klaten yang telah memberikan doa dan dukungan selama ini dalam menjalani perkuliahan.
3. Ibu Yuli Astuti, M.Kom selaku dosen pembimbing yang penuh kesabaran telah meluangkan waktu, arahan, dan pikirannya untuk memberikan bimbingan skripsi dari awal hingga akhir pengerjaan.
4. Ibu Anna Baita, M.Kom selaku dosen wali saya selama menempuh pendidikan di Universitas AMIKOM Yogyakarta.
5. Seluruh dosen Universitas AMIKOM Yogyakarta yang senantiasa memberikan ilmu, kritik, dan sarannya selama menjalani perkuliahan.
6. Kepada sahabat-sahabat saya Neo Extreme yang telah memberikan dukungan, masukan, dan hiburan selama ini. Serta teman-teman yang tidak bisa saya sebutkan satu per satu, terimakasih sudah memberikan bantuan dan masukan yang sangat berarti dalam pengerjaan skripsi ini.
7. Kepada ibu kost yang telah memberikan tempat tinggal yang murah.

## KATA PENGANTAR

Puji dan syukur yang selalu penulis panjatkan kepada Allah SWT atas segala pertolongan, rahmat, dan hidayah-Nya. Sehingga penulis dapat menyelesaikan penyusunan skripsi ini yang berjudul “Analisis Sentimen Data Twitter Terhadap Kebocoran Data Pada Aplikasi PeduliLindungi Menggunakan Algoritma Naïve Bayes” dengan tepat waktu. Serta shalawat dan salam kepada Nabi Muhammad SAW yang telah menjadi pedoman dan teladan terbaik untuk umatnya.

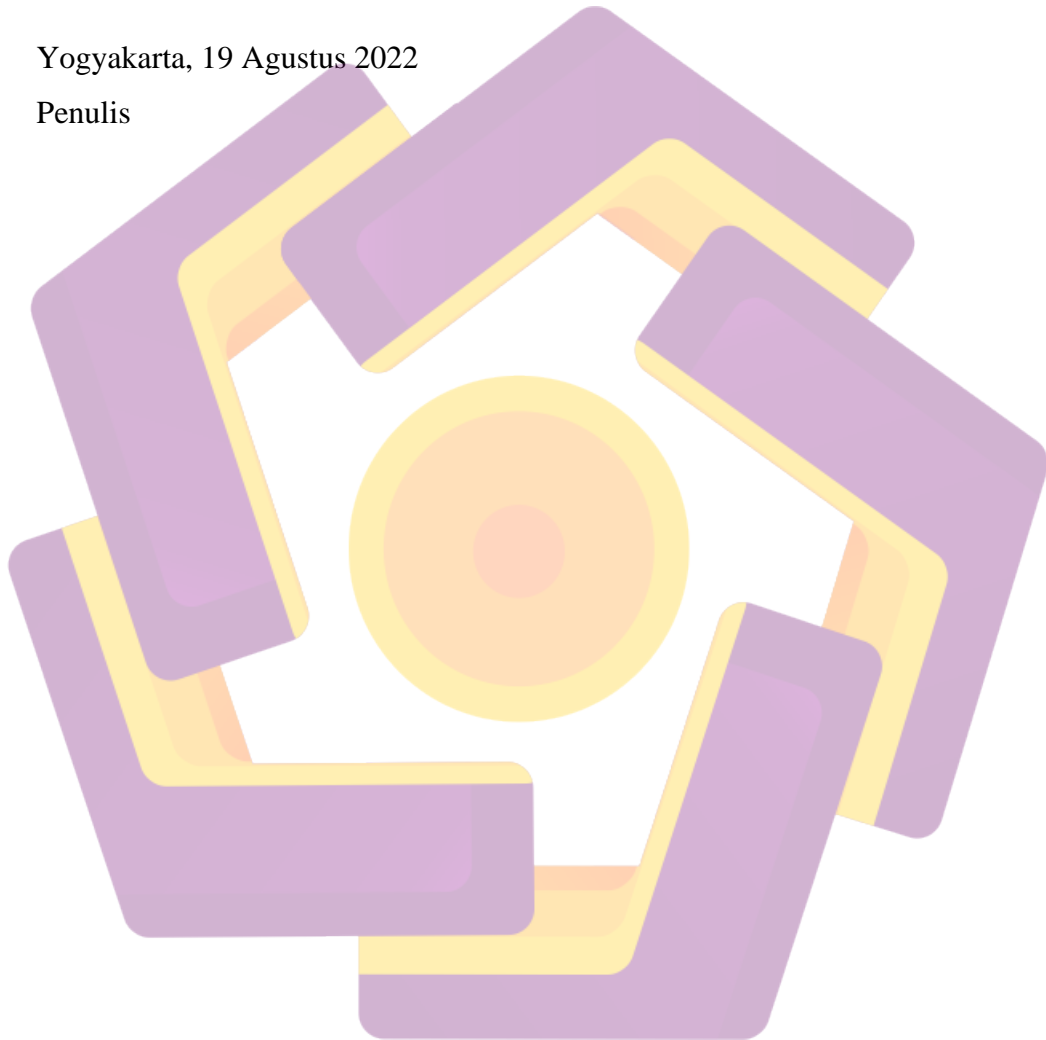
Skripsi ini dibuat guna memenuhi syarat kelulusan jenjang studi Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Penulis menyadari banyak pihak yang memberikan dukungan dan bantuan selama menyelesaikan studi dan skripsi ini. Oleh karena itu, sudah sepantasnya penulis dengan penuh rasa hormat mengucapkan terimakasih dan mendoakan semoga Allah SWT memberikan balasan terbaik kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.ko. selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Yuli Astuti, M.Kom. selaku Dosen Pembimbing yang dengan penuh kesabaran telah memberikan bimbingan, arahan, waktu dan pikirannya.
4. Ibu Anna Baita, M.Kom selaku dosen wali saya selama menempuh pendidikan di Universitas AMIKOM Yogyakarta.
5. Ibu Rakhma Shafrida Kurnia, S.Kom., M.Kom sebagai dosen penguji satu dan bapak Donni Prabowo, M.Kom sebagai dosen penguji kedua.
6. Kedua orang tua yang telah memberikan doa dan dukungan kepada saya sehingga saya dapat menyelesaikan skripsi ini.
7. Semua pihak yang tidak bisa disebutkan satu per satu baik secara langsung maupun tidak langsung yang telah membantu dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, karena keterbatasan ilmu yang penulis miliki. Untuk itu penulis dengan kerendahan hati mengharapkan saran dan kritik yang bersifat membangun sehingga dapat menjadi lebih baik lagi kedepannya. Serta harapan dari penulis, skripsi ini bisa memberi manfaat bagi para pembaca.

Yogyakarta, 19 Agustus 2022

Penulis





## DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	v
HALAMAN PERSEMBAHAN .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN.....	xv
DAFTAR LAMBANG DAN SINGKATAN .....	xvi
DAFTAR ISTILAH .....	xvii
INTISARI.....	xviii
<i>ABSTRACT</i> .....	xix
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang.....	1
1.2 Perumusan masalah.....	3
1.3 Batasan Penelitian .....	3
1.4 Tujuan Penelitian .....	4
1.5 Manfaat Penelitian.....	4
1.6 Metode Penelitian .....	5
1.6.1 <i>Metode Pengumpulan Data</i> .....	5
1.6.2 <i>Metode Study Literature</i> .....	5
1.6.3 <i>Metode Analisis</i> .....	5
1.6.4 <i>Metode Testing</i> .....	5
1.7 Sistematika Penulisan .....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	7
2.1 Studi Literatur .....	7
2.2 Dasar Teori.....	11
2.2.1 <i>Kebocoran Data</i> .....	11
2.2.2 <i>Jenis Kebocoran Data</i> .....	11

2.2.2.1	Disengaja (Intentional Threats).....	11
2.2.2.2	Tidak disengaja (Inadvertent Threats) .....	11
2.2.3	<i>Pengertian Cybercrime</i> .....	12
2.2.3.1	Cyberterrorism.....	12
2.2.3.2	Cyberextortion.....	13
2.2.3.3	Cyberwarfare .....	13
2.2.3.4	Cyberattacks .....	13
2.2.4	<i>Text Mining</i> .....	13
2.2.5	<i>Analisis Sentimen</i> .....	14
2.2.6	<i>Text Preprocessing</i> .....	14
2.2.6.1	Cleaning .....	14
2.2.6.2	Case Folding.....	14
2.2.6.3	Tokenizing .....	15
2.2.6.4	Normalisasi .....	15
2.2.6.5	Stopword Removal .....	15
2.2.6.6	POS-Tagging.....	15
2.2.6.7	Stemming.....	15
2.2.6.8	Lemmatization .....	15
2.2.7	<i>TextBlob</i> .....	15
2.2.8	<i>Algoritma Naïve Bayes Classifier</i> .....	16
2.2.8.1	Jenis Naïve Bayes Classifier.....	17
2.2.9	<i>Count Vectorizer</i> .....	18
2.2.10	<i>Google Colaboratory</i> .....	18
2.2.11	<i>Evaluasi Performance</i> .....	18
<b>BAB III METODE PENELITIAN</b> .....		21
3.1	<b>Alur Kerja Analisis Sentimen</b> .....	21
3.1.1	<i>Study Literature</i> .....	22
3.1.2	<i>Pengumpulan Data</i> .....	22
3.1.3	<i>Preprocessing</i> .....	23
3.1.3.1	Cleaning .....	23
3.1.3.2	Case Folding.....	24
3.1.3.3	Stopword Removal .....	24
3.1.3.4	Tokenizing dan POS-Tagging .....	24
3.1.3.5	Normalisasi .....	25
3.1.4	<i>Labelling</i> .....	26
3.1.5	<i>Klasifikasi Naïve Bayes</i> .....	26
3.1.6	<i>Evaluasi Sentimen</i> .....	27

BAB IV HASIL DAN PEMBAHASAN .....	29
4.1 Labelling .....	29
4.1.1 <i>Pengujian Tahap Pertama Labelling TextBlob</i> .....	29
4.1.2 <i>Pengujian Tahap Kedua Labelling TextBlob</i> .....	30
4.1.3 <i>Pengujian Tahap Ketiga Labelling TextBlob</i> .....	31
4.2 Klasifikasi Naïve Bayes .....	32
4.3 Hasil dan Pembahasan.....	33
4.3.1 <i>Hasil Pengujian Tahap Pertama</i> .....	33
4.3.1.1 <i>Pengujian Kesatu Tahap Pertama</i> .....	33
4.3.1.2 <i>Pengujian Kedua Tahap Pertama</i> .....	35
4.3.1.3 <i>Pengujian Ketiga Tahap Pertama</i> .....	37
4.3.1.4 <i>Pengujian Keempat Tahap Pertama</i> .....	39
4.3.1.5 <i>Pengujian Kelima Tahap Pertama</i> .....	41
4.3.2 <i>Hasil Pengujian Tahap Kedua</i> .....	43
4.3.2.1 <i>Pengujian Kesatu Tahap Kedua</i> .....	43
4.3.2.2 <i>Pengujian Kedua Tahap Kedua</i> .....	45
4.3.2.3 <i>Pengujian Ketiga Tahap Kedua</i> .....	47
4.3.2.4 <i>Pengujian Keempat Tahap Kedua</i> .....	49
4.3.2.5 <i>Pengujian Kelima Tahap Kedua</i> .....	51
4.3.3 <i>Hasil Pengujian Tahap Ketiga</i> .....	53
4.3.2.1 <i>Pengujian Kesatu Tahap Ketiga</i> .....	53
4.3.2.2 <i>Pengujian Kedua Tahap Ketiga</i> .....	55
4.3.2.3 <i>Pengujian Ketiga Tahap Ketiga</i> .....	57
4.3.2.4 <i>Pengujian Keempat Tahap Ketiga</i> .....	59
4.3.2.5 <i>Pengujian Kelima Tahap Ketiga</i> .....	61
4.4 Perbandingan Hasil Pengujian .....	62
BAB V KESIMPULAN DAN SARAN.....	66
5.1 Kesimpulan.....	66
5.2 Saran .....	66
DAFTAR PUSTAKA .....	68
LAMPIRAN .....	1

## DAFTAR TABEL

Tabel 2.1 Perbandingan Tinjauan Pustaka	7
Tabel 2.2 Confusion Matrix	19
Tabel 2.3 Multiclass Confusion Matrix	19
Tabel 3.1 Contoh Tahap Cleaning	24
Tabel 3.2 Contoh Tahap Case Folding	24
Tabel 3.3 Contoh Tahap Stopword Removal	24
Tabel 3.4 Contoh Tahap Tokenizing dan POS-Tag	25
Tabel 3.5 Contoh Normalisasi dengan Lemmatization	25
Tabel 3.6 Contoh Labelling TextBlob	26
Tabel 4.1 Perbandingan Splitting dataset	33
Tabel 4.2 Perbandingan Hasil Pengujian	63

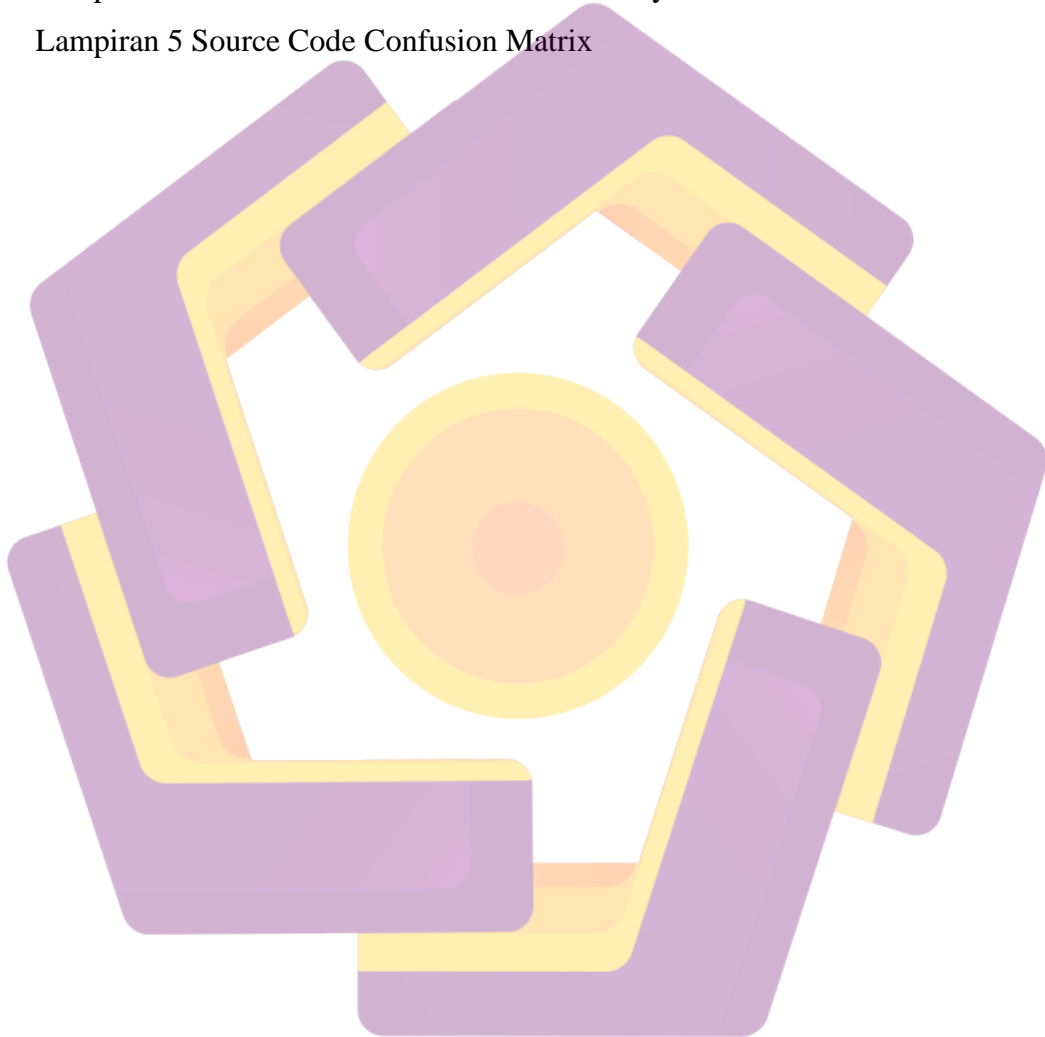
## DAFTAR GAMBAR

Gambar 3.1 Alur Kerja Analisis Sentimen	21
<i>Gambar 3.2 Crawling Data</i>	22
Gambar 3.3 Tahap Preprocessing	23
Gambar 3.4 Tahap Klasifikasi Naive Bayes	27
Gambar 3.5 Contoh Hasil Multiclass Confusion Matrix	28
Gambar 4.1 Source Code TextBlob	29
Gambar 4.2 Hasil Labelling TextBlob Pengujian Tahap Pertama	30
Gambar 4.3 Diagram Pie Labelling TextBlob Pengujian Tahap Pertama	30
Gambar 4.4 Hasil Labelling TextBlob Pengujian Tahap Kedua	30
Gambar 4.5 Diagram Pie Labelling TextBlob Pengujian Tahap Kedua	31
Gambar 4.6 Hasil Labelling TextBlob Pengujian Tahap Ketiga	31
Gambar 4.7 Diagram Pie Labelling TextBlob Pengujian Tahap Ketiga	31
Gambar 4.8 Source Code Count Vectorize	32
Gambar 4.9 Source Code Splitting dataset	32
Gambar 4.10 Source Code Multinomial Naive Bayes	32
Gambar 4.11 Confusion Matrix Pengujian Kesatu Tahap Pertama	33
Gambar 4.12 Hasil Confusion Matrix Pengujian Kesatu Tahap Pertama	34
Gambar 4.13 Confusion Matrix Pengujian Kedua Tahap Pertama	35
Gambar 4.14 Hasil Confusion Matrix Pengujian Kedua Tahap Pertama	36
Gambar 4.15 Confusion Matrix Pengujian Ketiga Tahap Pertama	37
Gambar 4.16 Hasil Confusion Matrix Pengujian Ketiga Tahap Pertama	38
Gambar 4.17 Confusion Matrix Pengujian Keempat Tahap Pertama	39
Gambar 4.18 Hasil Confusion Matrix Pengujian Keempat Tahap Pertama	40
Gambar 4.19 Confusion Matrix Pengujian Kelima Tahap Pertama	41
Gambar 4.20 Hasil Confusion Matrix Pengujian Kelima Tahap Pertama	42
Gambar 4.21 Confusion Matrix Pengujian Kesatu Tahap Kedua	43
Gambar 4.22 Hasil Confusion Matrix Pengujian Kesatu Tahap Kedua	44
Gambar 4.23 Confusion Matrix Pengujian Kedua Tahap Kedua	45

Gambar 4.24 Hasil Confusion Matrix Pengujian Kedua Tahap Kedua	46
Gambar 4.25 Confusion Matrix Pengujian Ketiga Tahap Kedua	47
Gambar 4.26 Hasil Confusion Matrix Pengujian Ketiga Tahap Kedua	48
Gambar 4.27 Confusion Matrix Pengujian Keempat Tahap Kedua	49
Gambar 4.28 Hasil Confusion Matrix Pengujian Keempat Tahap Kedua	50
Gambar 4.29 Confusion Matrix Pengujian Kelima Tahap Kedua	51
Gambar 4.30 Hasil Confusion Matrix Pengujian Kelima Tahap Kedua	52
Gambar 4.31 Confusion Matrix Pengujian Kesatu Tahap Ketiga	53
Gambar 4.32 Hasil Confusion Matrix Pengujian Kesatu Tahap Ketiga	54
Gambar 4.33 Confusion Matrix Pengujian Kedua Tahap Ketiga	55
Gambar 4.34 Hasil Confusion Matrix Pengujian Kedua Tahap Ketiga	56
Gambar 4.35 Confusion Matrix Pengujian Ketiga Tahap Ketiga	57
Gambar 4.36 Hasil Confusion Matrix Pengujian Ketiga Tahap Ketiga	58
Gambar 4.37 Confusion Matrix Pengujian Keempat Tahap Ketiga	59
Gambar 4.38 Hasil Confusion Matrix Pengujian Keempat Tahap Ketiga	60
Gambar 4.39 Confusion Matrix Pengujian Kelima Tahap Ketiga	61
Gambar 4.40 Hasil Confusion Matrix Pengujian Kelima Tahap Ketiga	62
Gambar 4.41 Hasil Akurasi Naive Bayes Classifier	64

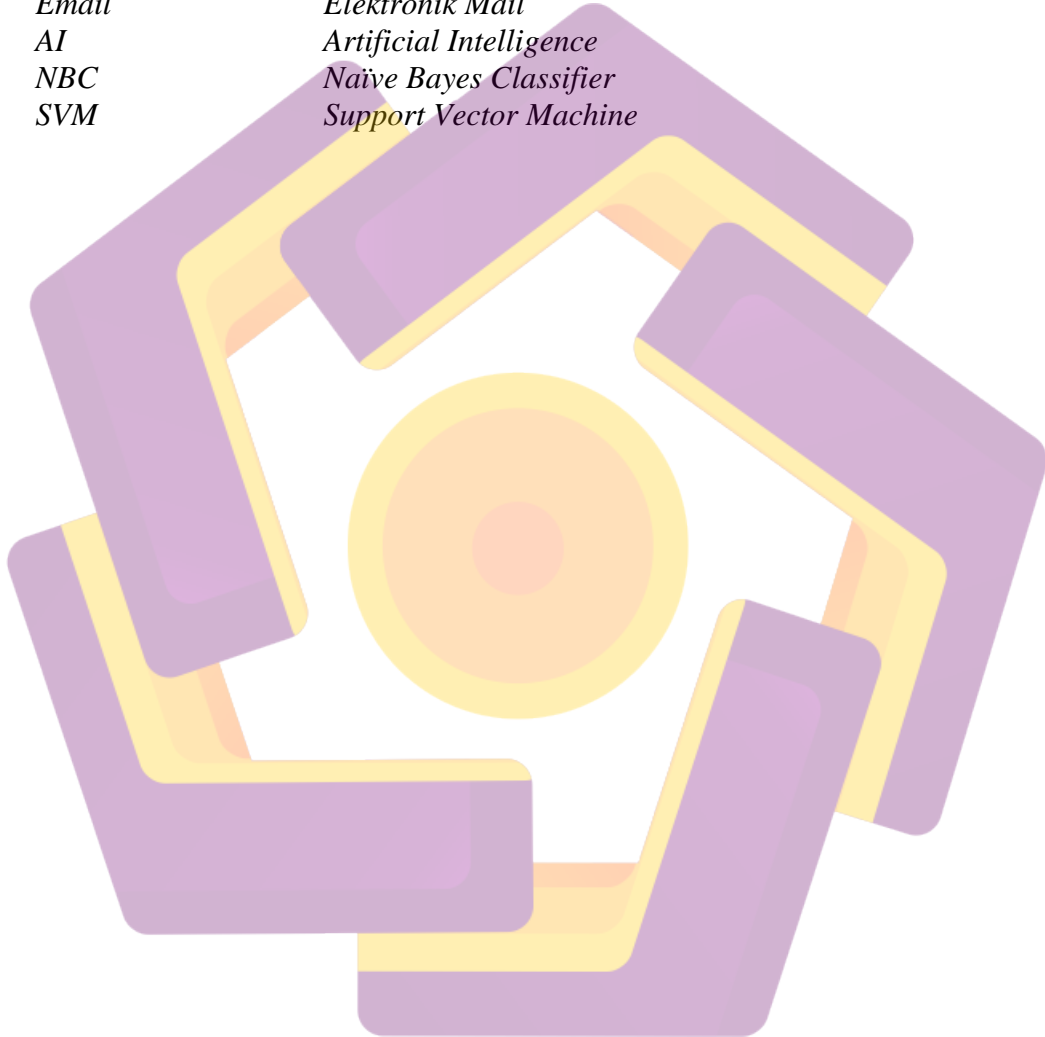
## DAFTAR LAMPIRAN

Lampiran 1 Source Code Labelling menggunakan TextBlob	1
Lampiran 2 Source Code Count Vectorize	2
Lampiran 3 Source Code Splitting Dataset	2
Lampiran 4 Source Code Multinomial Naive Bayes	2
Lampiran 5 Source Code Confusion Matrix	3



## DAFTAR LAMBANG DAN SINGKATAN

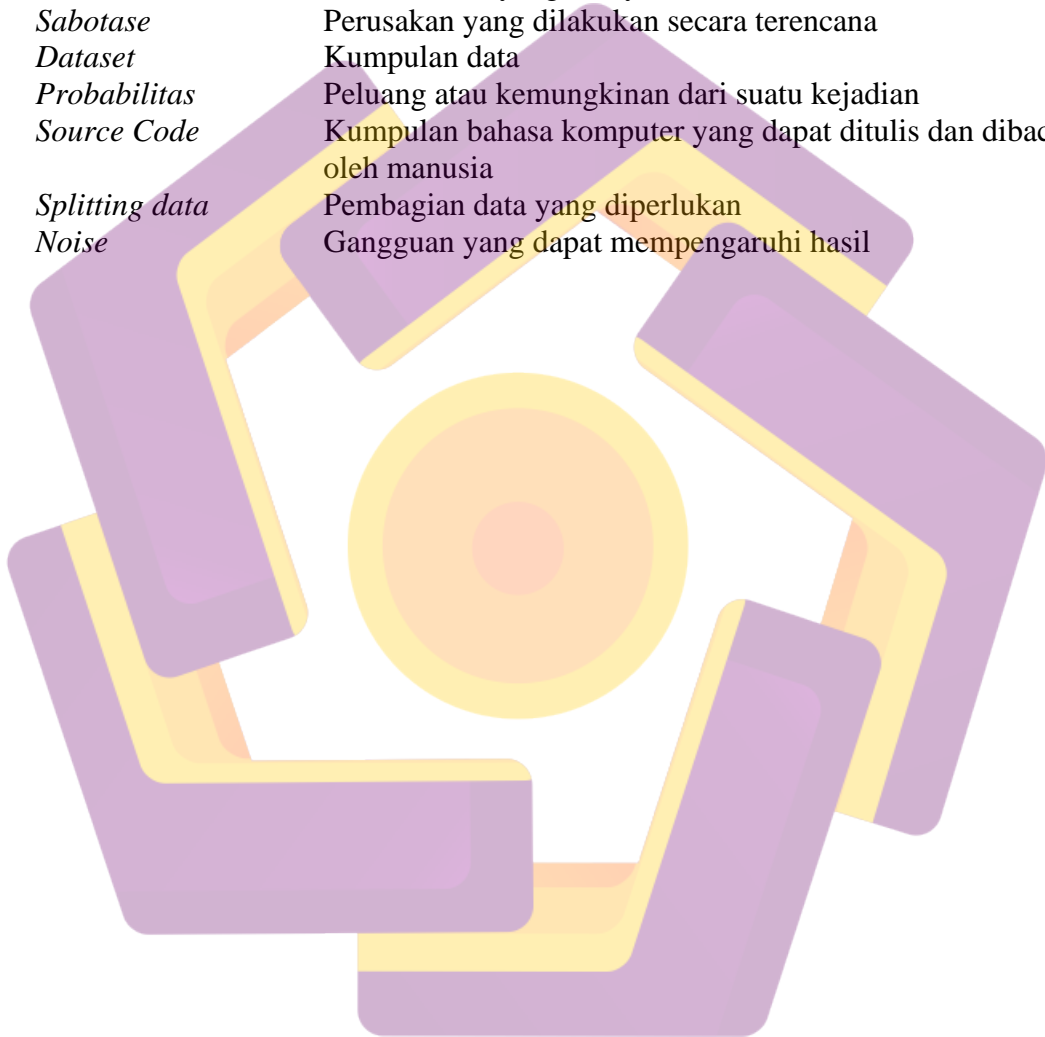
BPJS	Badan Penyelenggara Jaminan Sosial
COVID-19	<i>Corona Virus Disease-2019</i>
NIK	Nomor Induk Kependudukan
<i>Twitter API</i>	<i>Twitter Application Programming Interface</i>
SMS	<i>Short Message Service</i>
<i>Email</i>	<i>Elektronik Mail</i>
AI	<i>Artificial Intelligence</i>
NBC	<i>Naïve Bayes Classifier</i>
SVM	<i>Support Vector Machine</i>





## DAFTAR ISTILAH

<i>CyberCrime</i>	Tindak kejahatan yang dilakukan melalui jaringan internet
<i>Bug</i>	kerusakan atau error secara teknis pada suatu aplikasi
<i>Hacker</i>	Pelaku tindak kejahatan jaringan internet
<i>Deep Web</i>	Situs yang tidak bisa diakses dengan jaringan internet biasa
<i>Microblogging</i>	Penggabungan antara post singkat dan pesan instan
<i>Website</i>	Jalur internet yang menyediakan halaman informasi
<i>Sabotase</i>	Perusakan yang dilakukan secara terencana
<i>Dataset</i>	Kumpulan data
<i>Probabilitas</i>	Peluang atau kemungkinan dari suatu kejadian
<i>Source Code</i>	Kumpulan bahasa komputer yang dapat ditulis dan dibaca oleh manusia
<i>Splitting data</i>	Pembagian data yang diperlukan
<i>Noise</i>	Gangguan yang dapat mempengaruhi hasil



## INTISARI

Kemajuan teknologi akan terus berkembang bersamaan dengan kemajuan zaman. Begitu pula dengan data yang tersimpan di internet akan semakin banyak. Data-data tersebut akan dilindungi oleh pihak-pihak terkait agar tidak diambil dan disalahgunakan oleh pihak yang tidak bertanggung jawab. Kasus kebocoran data sudah beberapa kali terjadi di Indonesia, contohnya pada aplikasi PeduliLindungi. PeduliLindungi merupakan sebuah aplikasi yang dibuat pemerintah dengan tujuan untuk menyimpan data masyarakat yang sudah melakukan vaksinasi *COVID-19*. Dengan beredarnya kabar kasus kebocoran data pada aplikasi PeduliLindungi, membuat masyarakat menjadi resah terhadap sistem keamanan yang dibuat oleh pihak pengembang aplikasi. Banyak masyarakat yang mengutarakan tanggapan mereka di berbagai media sosial, contohnya di twitter.

Tanggapan-tanggapan tersebut dapat diolah menjadi sebuah informasi yang diperlukan. Namun sebelum menjadi sebuah informasi, harus dilakukannya tahap *preprocessing* dan klasifikasi menggunakan Algoritma *Naive Bayes Classifier*. Penelitian ini menggunakan *dataset* sebanyak 1501 *tweet* yang diuji melalui tiga tahapan dengan masing-masing tahap pengujiannya sebanyak lima kali, dengan perbandingan *splitting dataset* yang berbeda.

Hasil analisis dari lima kali pengujian disetiap tahap pengujian, terdapat satu pengujian yang menghasilkan skor akurasi yang tertinggi yakni 88% dengan perbandingan *splitting dataset* sebesar 80:20 atau 1200 data *training* dan 301 data *testing* pada pengujian kesatu tahap pertama.

**Kata kunci:** Kebocoran Data, PeduliLindungi, Naïve Bayes Classifier, Analisis Sentimen, Twitter

## **ABSTRACT**

*Technological progress will continue to develop along with the progress of the times. Likewise with the data stored on the internet will be more and more. These data will be protected by related parties so that they are not taken and misused by irresponsible parties. Cases of data leakage have occurred several times in Indonesia, for example in the PeduliLindung application. PeduliLindung is an application created by the government with the aim of storing data on people who have been vaccinated against COVID-19. With the news of a data leak in the PeduliLindung application circulating, the public becomes restless about the security system created by the application developer. Many people express their responses on various social media, for example on Twitter.*

*These responses can be processed into the necessary information. However, before it becomes information, it is necessary to carry out the preprocessing and classification stages using the Naive Bayes Classifier Algorithm. This study uses a dataset of 1501 tweets which were tested through three stages with each stage of the test five times, with different dataset splitting comparisons.*

*The results of the analysis of five tests in each testing stage, there is one test that produces the highest accuracy score of 88% with a dataset splitting ratio of 80:20 or 1200 training data and 301 testing data in the first stage of the first test.*

**Keyword:** *Data Leakage, Care to Protect, Naive Bayes Classifier, Sentiment Analysis, Twitter*