



LAPORAN SKRIPSI

**ANALISIS SENTIMEN TERHADAP APLIKASI PEDULI
LINDUNGI DARI DATA *TWITTER* MENGGUNAKAN
METODE *NAIVE BAYES CLASSIFIER***

**AHMAD NURUL MA'ARIF
NIM. 201751188**

**DOSEN PEMBIMBING
Anastasya Latubessy, S.Kom, M.Cs
Tutik Khotimah, S.Kom, M.Kom**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MURIA KUDUS
FEBRUARI 2023**

HALAMAN PERSETUJUAN

ANALISIS SENTIMEN TERHADAP APLIKASI PEDULI LINDUNGI DARI DATA *TWITTER* MENGGUNAKAN METODE *NAIVE BAYES CLASSIFER*

AHMAD NURUL MA'ARIF
NIM. 201751188

Kudus, 21 Januari 2023

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,



Anastasya Latubessy, S.Kom, M.Cs
NIDN. 0604048702



Tutik Khotimah, S.Kom, M.Kom
NIDN. 0608068502

HALAMAN PENGESAHAN

ANALISIS SENTIMEN TERHADAP APLIKASI PEDULI LINDUNGI DARI DATA *TWITTER* MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER*

AHMAD NURUL MA'ARIF
NIM. 201751188

Kudus, 14 Februari 2023

Menyetujui,

Ketua Penguji,

Ratih Nindyasari S.Kom, M.kom.
NIDN. 0625028501

Anggota Penguji I,

Ahmad Jazuli S.Kom, M.Kom.
NIDN. 0406107004

Anggota Penguji II,

Tutik Khotimah S.Kom, M.Kom.
NIDN. 0608068502

Mengetahui

Dekan Fakultas Teknik

Mohammad Daulan, S.T, M.T.
NIS: 061070100001141

Ketua Program Studi
Teknik Informatika

Mukhamad Nurkamid, S.Kom, M.Cs.
NIS. 0610701000001212

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini :

Nama : Ahmad Nurul Ma'arif
NIM : 201751188
Tempat & Tanggal Lahir : Jepara, 11 Juni 1999
Judul Skripsi : Analisis Sentimen Terhadap Aplikasi Peduli
Lindungi dari Data *Twitter* Menggunakan Metode
Naive bayes Classifier

Menyatakan dengan sebenarnya bahwa penulisan skripsi ini berdasarkan hasil penelitian, pemikiran dan pemaparan asli dari saya sendiri, baik untuk naskah laporan maupun kegiatan lain yang tercantum sebagai bagian dari Skripsi ini. Seluruh ide, pendapat, atau materi dari sumber lain telah dikutip dalam Skripsi dengan cara penulisan referensi yang sesuai.

Demikian pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar dan sanksi lain sesuai dengan peraturan yang berlaku di Universitas Muria Kudus.

Demikian pernyataan ini saya buat dalam keadaan sadar tanpa paksaan dari pihak manapun.

Kudus, 21 Januari 2023

Yang memberi pernyataan,



Ahmad Nurul Ma'arif
NIM. 201751188

KATA PENGANTAR

Segala puji syukur bagi Allah SWT, alhamdulillah penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya. Akhirnya penulis berhasil menyelesaikan skripsi yang berjudul "Analisis Sentimen Terhadap Aplikasi Peduli Lindungi Dari Data *Twitter* Menggunakan Metode *Naive Bayes Classifier*".

Penyusunan Skripsi/Tugas Akhir ini ditujukan untuk memenuhi salah satu syarat memperoleh gelar sarjana sastra satu (S1) dalam Fakultas Teknik Program studi Teknik Informatika pada Universitas Muria Kudus. Dalam proses penyelesaian Penulis menyadari bahwa keberhasilan penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, dengan segala kerendahan hati penulis menyampaikan terima kasih:

1. Prof. Dr.Ir. Darsono, M.Si. Selaku Rektor Universitas Muria Kudus
2. Mohammad Dahlan, ST., MT, selaku Dekan Fakultas Teknik Universitas Muria Kudus
3. Muhammad Malik Hakim, ST., MTI, selaku ketua Program studi Teknik Informatika Fakultas Teknik Universitas Muria Kudus
4. Rati Nindyasari, S.Kom., M.Kom., selaku Koordinator Skripsi
5. Anastasya Latubessy, S.Kom, M.Cs., selaku pembimbing utama dan Tutik Khotimah, S.Kom, M.Kom. selaku pembimbing pendamping yang telah memberikan bimbingannya, ilmu, saran, motivasi, nasihat, dan pengarahan selama penyusunan skripsi ini.
6. Kedua orang tua saya, Bapak dan Ibu saya tercinta, yang telah mengasuh dan membimbing serta memberikan arahan dan dorongan kepada penulis, baik moral maupun spiritual dalam menyelesaikan skripsi ini.

Penulis menyadari dengan segala keterbatasan dan kemampuan yang dimiliki dalam penyusunan skripsi ini yang masih jauh dari kesempurnaan, untuk itu penulis membuka diri dan menerima kritik, serta saran yang konstruktif dari pembaca agar skripsi ini dapat menjadi lebih baik lagi.

Akhirnya penulis hanya dapat membalas dengan do'a, semoga Allah SWT yang akan memberikan balasan atas kebaikan budi mereka. Akhirukalam, semoga karya skripsi ini dapat diambil manfaatnya bagi para pembaca, Aamiin.

Kudus, Januari 2021

Penulis



ANALISIS SENTIMEN TERHADAP APLIKASI PEDULI LINDUNGI DARI DATA *TWITTER* MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER*

Nama mahasiswa : Ahmad Nurul Ma'arif

NIM : 201751188

Pembimbing :

1. Anastasya Latubessy, S.Kom, M.Cs
2. Tutik Khotimah, S.Kom, M.Kom

RINGKASAN

Tahun 2020 merupakan tahun mulainya pandemi virus *covid-19* yang muncul pertama kali di Wuhan Cina, dan menyebabkan kepanikan luar biasa kepada semua penduduk dunia. Salah satu terobosan yang dilakukan oleh pemerintah Indonesia adalah dengan meluncurkan aplikasi PeduliLindungi, yang diharapkan mampu memberikan peringatan kepada masyarakat jika kalau memasuki daerah yang terdampak virus *covid-19*. *Twitter* merupakan salah satu situs *microblogging* dengan pengguna lebih dari 29,5 juta pengguna *twitter* di Indonesia. *Tweet* pada *twitter* bisa jadi sumber data untuk analisis sentimen. Tujuan dari penelitian ini adalah dapat melakukan proses pengklasifikasian *tweet* dari pengguna *twitter* terhadap aplikasi Peduli Lindungi. Proses dimulai dari *scraping*, *Preprosesing*, dan pelabelan. Data hasil *preprosesing* dan pelabelan dilakukan klasifikasi menggunakan *metode naive bayes clasifer*. Berdasarkan hasil klasifikasi dari 2500 data dengan nilai polaritas positif 1114, negatif 290, dan netral 1096 didapat akurasi 77,8.

Kata kunci : Aplikasi PeduliLindungi, twitter, analisis sentimen, naive bayes classifier

**SENTIMEN ANALYSIS OF THE APPLICATION PEDULILINDUNGI
FROM TWITTER DATA USING THE NAIVE BAYES CLASSIFIER
METHOD**

Student Name : Ahmad Nurul Ma'arif

Student Identity Number : 201751188

Supervisor :

1. Anastasya Latubessy, S.Kom, M.Cs

2. Tutik Khotimah, S.Kom, M.Kom

ABSTRACT

The year 2020 was the year the Covid-19 virus pandemic began, which first appeared in Wuhan, China, and caused tremendous panic to all the world's population. One of the breakthroughs made by the Indonesian government is by launching the Pedulilindungi application, which is expected to be able to provide a warning to the public if they enter an area affected by the COVID-19 virus. Twitter is a microblogging site with more than 29.5 million twitter users in Indonesia. Tweets on Twitter can be a source of data for sentiment analysis. The purpose of this research is to be able to carry out the process of classifying tweets from twitter users to the Peduli Protect application. The process starts from scraping, preprocessing, and labeling. Data from preprocessing and labeling were classified using the Naive Bayes Classifier method. Based on the classification results of 2500 data with a positive polarity value of 1114, negative 290, and neutral 1096, an accuracy of 77.8 was obtained.

Keywords :pedulilindungi application, twitter, sentiment analysis, naive bayes classifier

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
PERNYATAAN KEASLIAN	iv
KATA PENGANTAR	v
RINGKASAN	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR RUMUS	xiii
BAB I PENDAHULUAN	
1.1 Latar belakang	1
1.2 Perumusan Masalah	2
1.3 Batasan Masalah	2
1.3 Tujuan	2
1.5 Manfaat	3
BAB II TINJAUAN PUSTAKA	
2.1 Penelitian Terkait	4
2.2 Analisis Sentimen	5
2.3 Aplikasi Pedulilindungi	6
2.4 Twitter	6
2.5 Metode Naive Bayes Classifier	7
2.6 Preprosesing	7
2.7 Pembobotan Kata	8
2.8 Evaluasi Performasi	8
2.8.1 Accuracy	9
2.8.2 Precision	9
2.8.3 Recall	10
2.8.4 F1-Score	10
BAB III METODOLOGI PENELITIAN	
3.1 Obyek Penelitian	11
3.2 Metode Pengumpulan Data	11
3.2.1 Pengumpulan Data	11
3.2.2 Studi Literatur	11
3.3 Implementasi	11
3.3.1 Gambaran Umum	11
3.3.2 Scraping	12
3.3.3 Preprosesing	13
3.3.4 Labeling	22
3.3.5 Vektorisasi TF-IDF	23

3.3.6	Klasifikasi Naive Bayes	24
3.3.7	Evaluasi	26
3.3.8	Akurasi	26

BAB IV HASIL DAN PEMBAHASAN

4.1	Pengumpulan Dataset	27
4.2	Tahap Preprosesing	32
4.3	Tahap Labeling.....	38
4.4	Vektorisasi.....	39
4.5	Klasifikasi Naive Bayes Classifier.....	40
4.6	Tahap Evaluasi	42
4.7	Visualisasi	45
4.8	Prediksi Model	47

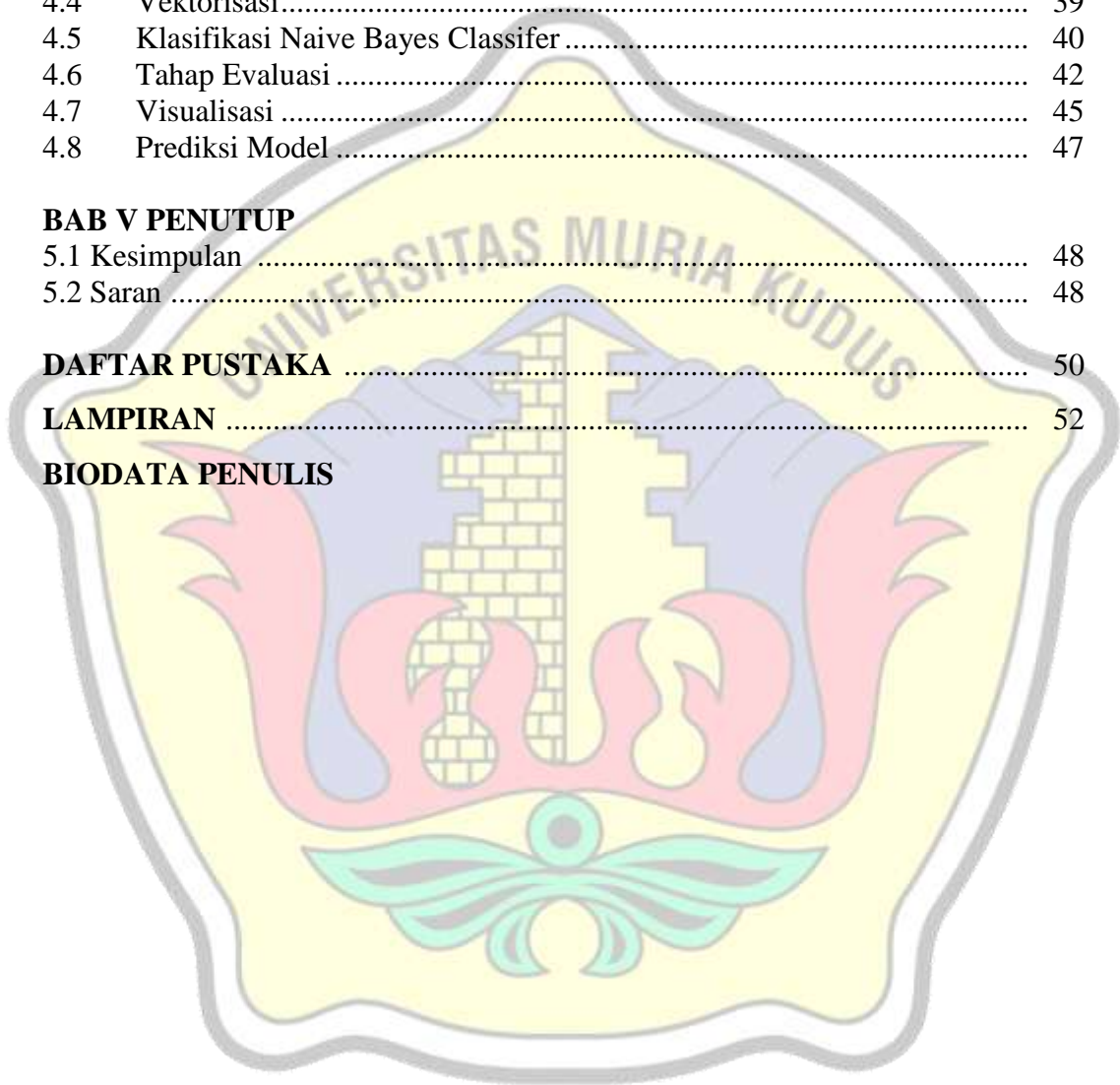
BAB V PENUTUP

5.1	Kesimpulan	48
5.2	Saran	48

DAFTAR PUSTAKA	50
-----------------------------	----

LAMPIRAN	52
-----------------------	----

BIODATA PENULIS

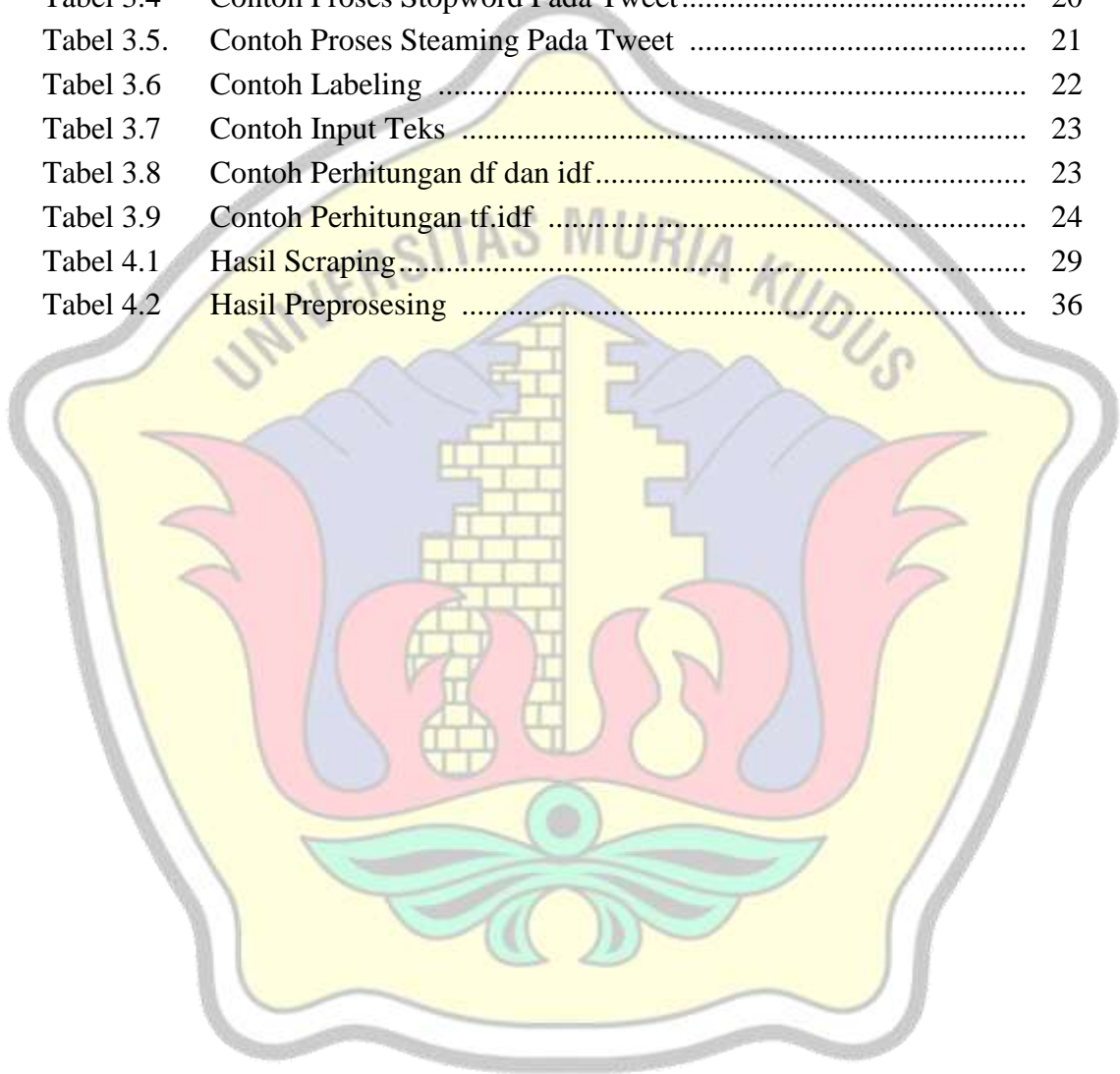


DAFTAR GAMBAR

Gambar 3.1	Diagram Alir Sentimen Analisis Secara Umum.....	12
Gambar 3.2	Diagram Alur Proses Scraping	13
Gambar 3.3	Diagram Alir Proses Preprosesing	14
Gambar 3.4	Diagram Alur Klasifikasi Naive Bayes	25
Gambar 4.1	Source Code Scraping	28
Gambar 4.2.	Halaman Scraping	29
Gambar 4.3	Hasil Scraping	29
Gambar 4.4	Source Code Preprosesing	33
Gambar 4.5	Source Code Cleansing.....	34
Gambar 4.6	Source Code Fungsi Case Folding	34
Gambar 4.7	Sourcode Tokenzing	34
Gambar 4.8	Sourcode Stop Removal	35
Gambar 4.9	Sourcode Steaming	35
Gambar 4.10	Halaman Preprosesing	36
Gambar 4.11	Hasil Preprosesing	36
Gambar 4.12	Proses Labeling Menggunakan Microsoft Exel	39
Gambar 4.13	Hasil Labeling Menggunakan Microsoft Exel	39
Gambar 4.14	Source Code Vectorize	40
Gambar 4.15	Source Code Klasifikasi Menggunakan Naive Bayes.....	41
Gambar 4.16	Halaman Klasifikasi	41
Gambar 4.17	Hasil Klasifikasi Naive Bayes	42
Gambar 4.18	Sourcode Confusion Matrix	44
Gambar 4.19	Hasil Confusion Matrix	45
Gambar 4.20	Grafik Tweet Nilai Positif	45
Gambar 4.21	Grafik Tweet Nilai Negatif	46
Gambar 4.22	Grafik Tweet Nilai Netral	46
Gambar 4.23	Word Cloud	47
Gambar 4.24	Prediksi Model	47

DAFTAR TABEL

Tabel 2.1	Struktur Confusion Matrix	9
Tabel 3.1	Contoh Proses Cleansing Pada Tweet	15
Tabel 3.2	Contoh Proses Case Folding Pada Tweet	16
Tabel 3.3	Contoh Proses Tokenzing Pada Tweet	18
Tabel 3.4	Contoh Proses Stopword Pada Tweet	20
Tabel 3.5.	Contoh Proses Steaming Pada Tweet	21
Tabel 3.6	Contoh Labeling	22
Tabel 3.7	Contoh Input Teks	23
Tabel 3.8	Contoh Perhitungan df dan idf	23
Tabel 3.9	Contoh Perhitungan tf.idf	24
Tabel 4.1	Hasil Scraping	29
Tabel 4.2	Hasil Preprosesing	36



DAFTAR RUMUS

Rumus 1	Rumus Metode Naive Bayes Classifier	7
Rumus 2	Rumus TF (Term Frequency)	8
Rumus 3	Rumus IDF (Inverse Document Frequency).....	8
Rumus 4	Rumus TFIDF (Term Frequency-Inverse Document Frequency).	8
Rumus 5	Rumus Accuracy	9
Rumus 6	Rumus Precision	10
Rumus 7	Rumus Recall	10
Rumus 8	Rumus F1-Score	10



DAFTAR LAMPIRAN

Lampiran 1	Lembar revisi sidang skripsi ketua penguji	52
Lampiran 2	Lembar revisi sidang skripsi anggota penguji 1	53
Lampiran 3	Lembar revisi sidang skripsi anggota penguji 2.....	54



DAFTAR ISTILAH DAN SINGKATAN

TF	: Term Frequency
IDF	: Inverse Documen Frequency
TP	: True Positive
FP	: False Positive
TN	: True Negative
FN	: False Negative

