

**1ANALISIS SENTIMEN TERHADAP APLIKASI PEDULILINDUNGI PADA
MEDIA SOSIAL TWITTER MENGGUNAKAN
ALGORITMA NAÏVE BAYES**

SKRIPSI



disusun oleh

Choirul Mahendrata Harpi

18.11.1869

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**ANALISIS SENTIMEN TERHADAP APLIKASI PEDULILINDUNGI
PADA MEDIA SOSIAL TWITTER MENGGUNAKAN
ALGORITMA NAÏVE BAYES**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Choirul Mahendrata Harpi

18.11.1869

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP APLIKASI PEDULILINDUNGI
PADA MEDIA SOSIAL TWITTER MENGGUNAKAN
ALGORITMA NAÏVE BAYES**

yang dipersiapkan dan disusun oleh

Choirul Mahendrata Harpi

18.11.1869

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 8 Juni 2022

Dosen Pembimbing,

M. Rudyanto Arief, M.T.

NIK. 190302098

PENGESAHAN

SKRIPSI

ANALISIS SENTIMEN TERDAHAP APLIKASI PEDULILINDUNGI PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES

yang dipersiapkan dan disusun oleh

Choirul Mahendrata Harpi
18.11.1869

telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Juni 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anggit Dwi Hartanto, M.Kom.
NIK. 190302163

Wiwi Widayani, M.Kom.
NIK. 190302272

M. Rudyanto Arief, M.T.
NIK. 190302098

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Juni 2022

DEKAN FAKULTAS ILMU KOMPUTER

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

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 3 Juni 2022



Choirul Mahendrata Harpi

NIM. 18.11.1869

PERSEMBAHAN

Peneliti mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung maupun tidak langsung dalam proses pembuatan skripsi.

1. Kedua orang tua dan keluarga, yang selalu mendoakan, selalu menyemangati dan selalu memberikan dukungan.
2. Bapak M. Rudyanto Arief, M.T. selaku dosen pembimbing yang telah membimbing dan mengarahkan dari awal sampai akhir pembuatan skripsi.
3. Bapak Anggit Dwi Hartanto, M.Kom. selaku dosen wali yang telah memberikan arahan selama kuliah.
4. Dosen – dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama kuliah.
5. Teman seperjuangan Fahmi Irsyad, Arya Putra, Aditya Bintang, Dwiki Rivan, Lukman Haryanto yang selalu memberikan solusi, tempat bertanya ketika dimasa sulit dan tempat curhat keluh kesah selama dalam pembuatan skripsi.
6. Teman – teman kelas IF 01 2018 yang selalu menemani perkuliahan, mendukung dan memberikan semangat sampai saat ini. Semoga kita bisa meraih apa yang kita angan-angan dan cita-cita.
7. Teman – teman UKM FOSSIL Universitas Amikom Yogyakarta yang selalu mendukung dan memberikan semangat sampai saat ini. Semoga kita selalu bahagia dan menjadi pribadi yang lebih baik lagi.
8. Orang – orang baik yang senantiasa mendoakan, memberi semangat dan bantuan.

KATA PENGANTAR

Puji dan syukur panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat, hidayah dan kekuatan sehingga peneliti dapat menyelesaikan skripsi yang berjudul Analisis Sentimen Terhadap Aplikasi PeduliLindungi Pada Media Sosial Twitter Menggunakan Algoritma *Naïve Bayes*.

Skripsi ini dibuat guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program strata satu dan untuk memperoleh gelar Sarjana Komputer. Dengan selesainya skripsi ini, maka pada kesempatan ini peneliti mengucapkan terima kasih kepada:

1. Tuhan Yang Maha Esa yang memberikan segala nikmat dan kasih sayangnya sampai sejauh ini.
2. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
3. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
4. Bapak M. Rudyanto Arief, M.T. selaku dosen pembimbing yang selalu bijaksana dalam memberikan bimbingan, nasehat serta waktunya selama pembuatan skripsi ini.
5. Dosen penguji dan segenap Dosen serta Karyawan Universitas AMIKOM Yogyakarta yang telah berbagi ilmu dan pengalamannya.
6. Kedua orang tua yang telah mendoakan, mendukung dan memberikan semangat kepada.
7. Orang – orang baik yang senantiasa mendoakan, memberi semangat dan bantuan kepada.
8. Teman – teman UKM FOSSIL Universitas Amikom Yogyakarta yang selalu memberi semangat dan bantuan.
9. Teman – teman IF 01 2018 yang telah menemani selama proses perkuliahan.

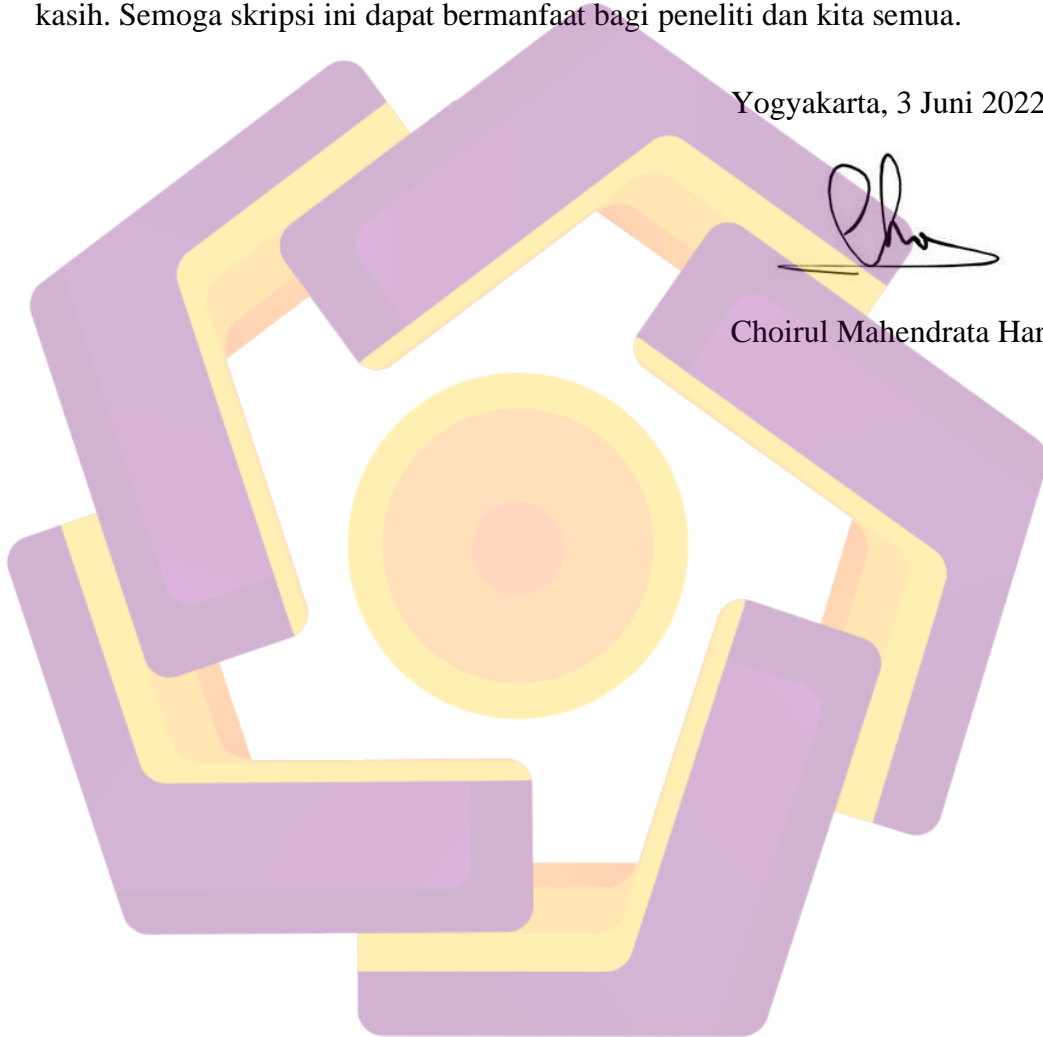
10. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

Semoga Allah SWT memberikan balasan yang lebih kepada semua yang telah ikut membantu peneliti hingga menyelesaikan skripsi ini. Demi perbaikan selanjutnya, saran dan kritik membangun diterima dengan senang hati dan terima kasih. Semoga skripsi ini dapat bermanfaat bagi peneliti dan kita semua.

Yogyakarta, 3 Juni 2022



Choirul Mahendrata Harpi



DAFTAR ISI

ANALISIS SENTIMEN TERHADAP APLIKASI PEDULILINDUNGI PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES	I
ANALISIS SENTIMEN TERHADAP APLIKASI PEDULILINDUNGI PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES	II
PERSETUJUAN	III
PENGESAHAN	IV
PERNYATAAN	V
PERSEMBAHAN	VI
KATA PENGANTAR	VII
DAFTAR ISI	IX
DAFTAR TABEL	XII
DAFTAR GAMBAR	XIII
DAFTAR LAMPIRAN	XIV
DAFTAR ISTILAH	XV
INTISARI	XVI
<i>ABSTRACT</i>	XVII
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH	3
1.3 BATASAN MASALAH	4
1.4 TUJUAN PENELITIAN	4
1.5 MANFAAT PENELITIAN	5
BAB II TINJAUAN PUSTAKA	6
2.1 STUDI LITERATUR	6
2.2 DASAR TEORI	8
2.2.1 PeduliLindungi	8
2.2.2 Media sosial	8

2.2.3	Analisis sentimen	8
2.2.4	Metode klasifikasi	8
2.2.5	<i>Data mining</i>	9
2.2.6	<i>Text mining</i>	10
2.2.7	<i>Preprocessing</i>	11
2.2.8	<i>Lexicon based features</i>	12
2.2.9	<i>Term frequency inverse document frequency (TF IDF)</i>	12
2.2.10	<i>Wordcloud</i>	14
2.2.11	<i>K-fold cross validation</i>	14
2.2.12	<i>Hyperparameter</i>	15
2.2.13	<i>Grid search</i>	15
2.2.14	<i>Naïve bayes classifier</i>	15
2.3	METODE PENGUJIAN	18
2.3.1	<i>Confussion matrix</i>	18
BAB III METODOLOGI PENELITIAN.....		20
3.1	PROSEDUR PENELITIAN	20
3.2	DATA PENELITIAN.....	21
3.2.1	Data primer.....	21
3.2.2	Data sekunder.....	21
3.3	INSTRUMEN PENGAMBILAN DATA.....	22
3.4	PARAMETER PENELITIAN.....	22
3.5	METODE ANALISIS	23
3.6	IMPLEMENTASI MULTINOMIAL NAÏVE BAYES	25
3.7	ANALISIS HASIL	27
BAB IV HASIL DAN PEMBAHASAN		28
4.1	PENGUMPULAN DATA TWITTER.....	28
4.2	<i>DATA PREPROCESSING</i>	29
4.2.1	<i>Cleaning data</i>	30
4.2.2	<i>Case folding</i>	31
4.2.3	<i>Tokenizing</i>	31
4.2.4	Normalisasi	32

4.2.5	<i>Filtering (stopwords removal)</i>	33
4.2.6	<i>Stemming</i>	34
4.3	<i>DATA LABELING</i> LEMKON INDONESIA	34
4.3.1	<i>Wordcloud</i>	35
4.4	<i>TERM FREQUENCY-INVERS DOCUMENT FREQUENCY (TF-IDF)</i>	36
4.4.1	<i>Statistical term frequency</i>	38
4.5	<i>KLASIFIKASI NAÏVE BAYES</i>	39
4.5.1	<i>Data splitting</i>	39
4.5.2	<i>Hyperparameter tuning</i>	39
4.5.3	<i>Hasil grid search CV</i>	40
4.6	<i>CONFUSION MATRIX</i>	40
4.7	<i>HASIL AKHIR</i>	43
BAB V PENUTUP		45
5.1	KESIMPULAN	45
5.2	SARAN	46
DAFTAR PUSTAKA		47
LAMPIRAN		50

DAFTAR TABEL

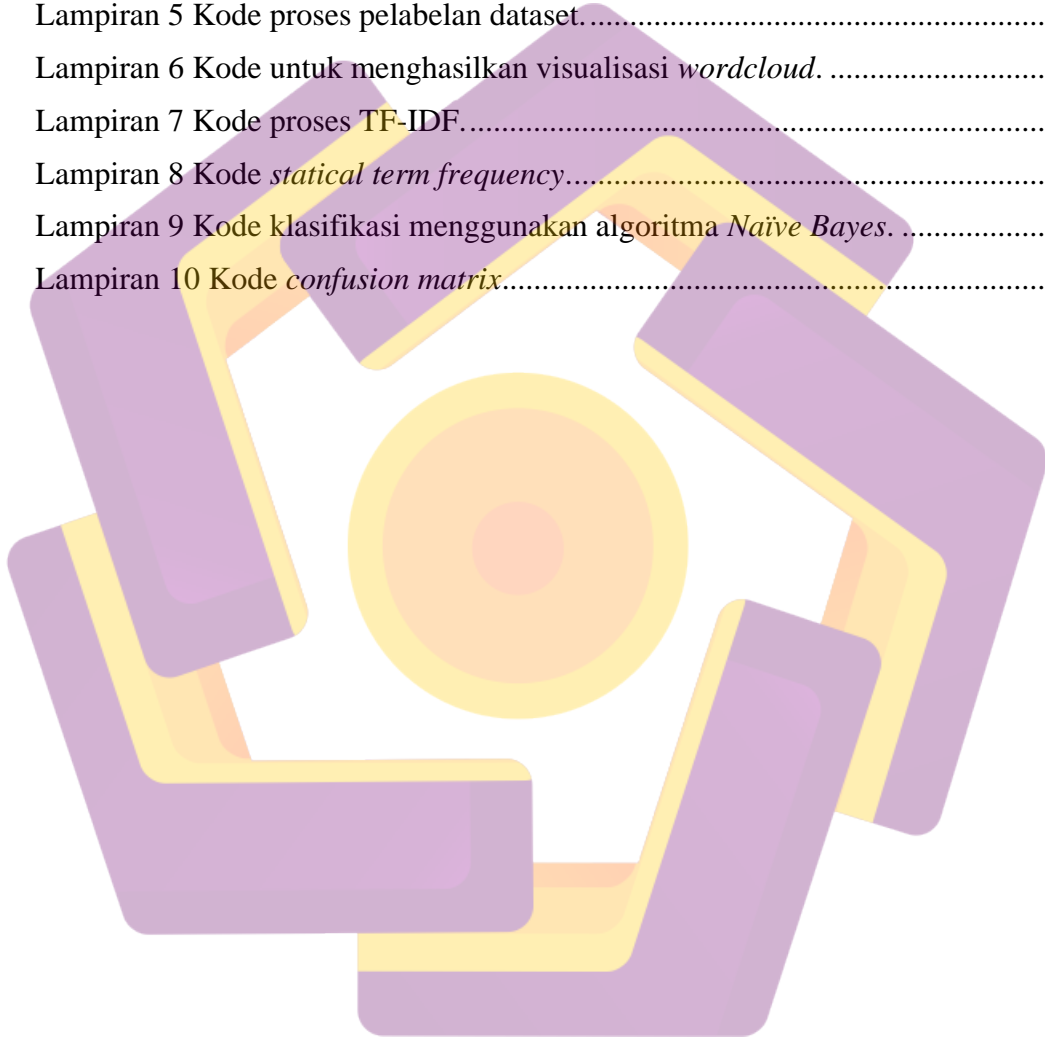
Tabel 2. 1 <i>Confussion Matrix</i>	18
Tabel 3. 1 Variabel & Definisi Operasional Variabel.....	22
Tabel 3. 2 Contoh hasil pembobotan TF pada data latih.....	25
Tabel 3. 3 Nilai Likelihood dari setiap <i>terms</i>	26
Tabel 3. 4 <i>Confusion Matrix</i>	27
Tabel 4. 1 Data <i>Tweet</i>	28
Tabel 4. 2 Data Awal Penelitian.....	29
Tabel 4. 3 Proses <i>Cleaning</i> Data.....	30
Tabel 4. 4 Proses <i>Case Folding</i>	31
Tabel 4. 5 Proses <i>Tokenizing</i>	31
Tabel 4. 6 Proses Normalisasi.....	32
Tabel 4. 7 Proses <i>Filtering</i>	33
Tabel 4. 8 Proses <i>Stemming</i>	34
Tabel 4. 9 Pelabelan data menggunakan Leksikon Indonesia.....	35
Tabel 4. 10 Contoh Dokumen.....	36
Tabel 4. 11 Perhitungan Nilai TF.....	36
Tabel 4. 12 Perhitungan Nilai IDF.....	37
Tabel 4. 13 Perhitungan Nilai TF-IDF.....	37
Tabel 4. 14 Pembagian data <i>training</i> dan data <i>testing</i>	39
Tabel 4. 15 <i>Parameter grid</i>	39
Tabel 4. 16 Hasil <i>Grid Search CV</i>	40
Tabel 4. 17 <i>Confusion Matrix</i> analisis sentimen tidak menggunakan <i>hyperparameter tuning</i>	41
Tabel 4. 18 <i>Confusion Matrix</i> analisis sentimen menggunakan <i>hyperparameter tuning</i>	42
Tabel 4. 19 Sampel hasil klasifikasi model algoritma naïve bayes.....	43
Tabel 4. 20 Kesalahan prediksi pada model.....	43

DAFTAR GAMBAR

Gambar 1. 1 Grafik respon masyarakat terhadap Aplikasi PeduliLindungi.	2
Gambar 2. 1 Diagram Proses Data Mining. (Fayyad, 1996).....	9
Gambar 2. 2 Visualiasi data menggunakan Wordcloud.....	14
Gambar 2. 3 <i>K-Fold Cross Validation</i> (Rahmadya Trias Handayanto, 2020).....	14
Gambar 3. 1 Tahapan Penelitian.	20
Gambar 3. 2 Bagan alur implementasi algoritma multinomial naïve bayes.	25
Gambar 4. 1 Perintah <i>Twint</i> untuk <i>scraping</i> pada media sosial Twitter.....	28
Gambar 4. 2 Visual wordcloud kelas positif dan negatif.....	35
Gambar 4. 3 Frekuensi kata yang muncul terhadap aplikasi PeduliLindungi.....	38
Gambar 4. 4 <i>Clasification report</i> tidak menggunakan <i>hyperparameter tuning</i>	41
Gambar 4. 5 Hasil <i>akurasi, presisi, recall dan f1-score</i> menggunakan <i>hyperparameter tuning</i>	42
Gambar 4. 6 Persentase hasil klasifikasi algoritma naïve bayes.....	44

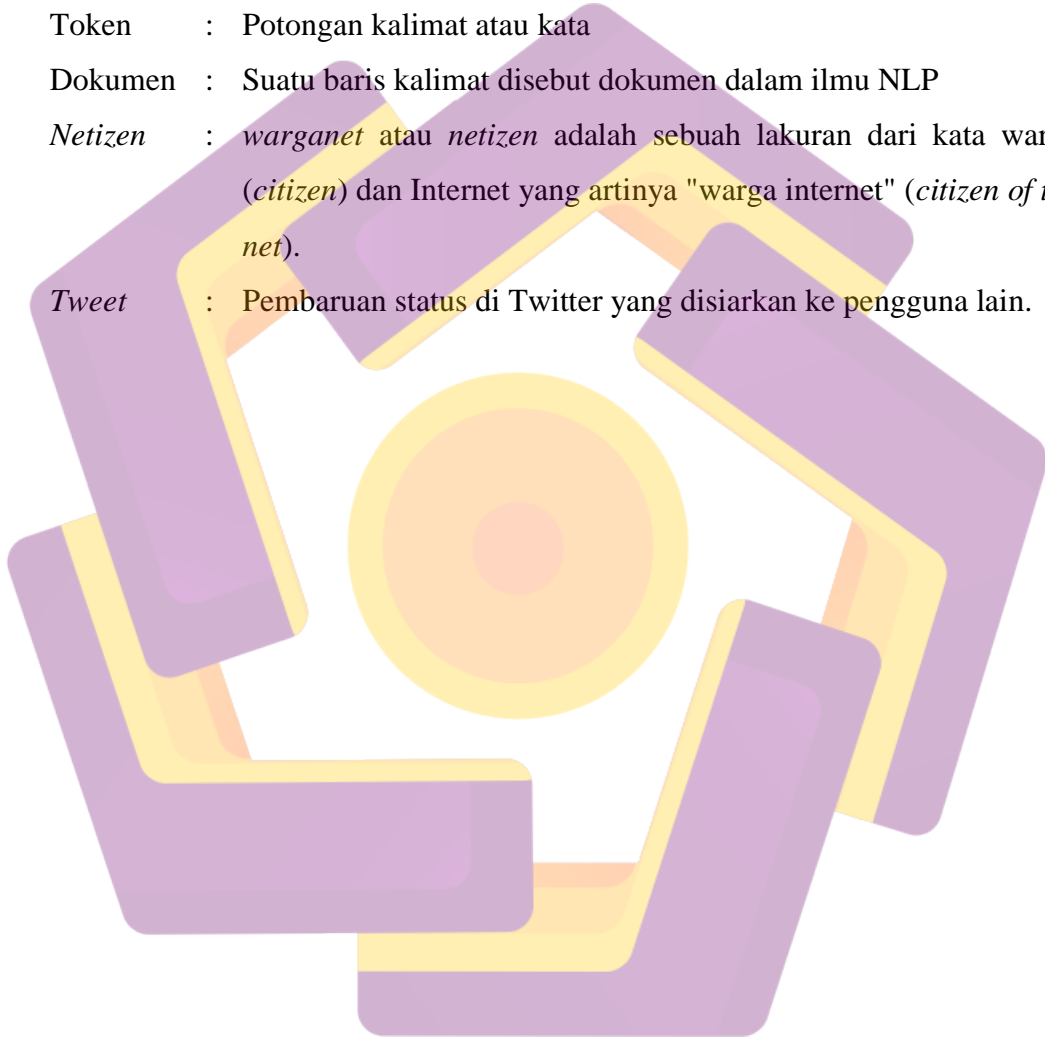
DAFTAR LAMPIRAN

Lampiran 1 Kode <i>load</i> data.....	50
Lampiran 2 Kode persiapan tahap <i>preprocessing</i>	50
Lampiran 3 Kode membuat fungsi <i>preprocessing</i>	51
Lampiran 4 Kode penerapan fungsi <i>preprocessing</i>	52
Lampiran 5 Kode proses pelabelan dataset.....	52
Lampiran 6 Kode untuk menghasilkan visualisasi <i>wordcloud</i>	53
Lampiran 7 Kode proses TF-IDF.....	53
Lampiran 8 Kode <i>statical term frequency</i>	54
Lampiran 9 Kode klasifikasi menggunakan algoritma <i>Naïve Bayes</i>	55
Lampiran 10 Kode <i>confusion matrix</i>	56



DAFTAR ISTILAH

- Dataset* : Set data (data set/himpunan data) merupakan kumpulan objek dan atributnya.
- Scraping* : Proses untuk mengekstraksi informasi dan data di dalam website secara otomatis.
- Token* : Potongan kalimat atau kata
- Dokumen* : Suatu baris kalimat disebut dokumen dalam ilmu NLP
- Netizen* : *warganet* atau *netizen* adalah sebuah lakuran dari kata warga (*citizen*) dan Internet yang artinya "warga internet" (*citizen of the net*).
- Tweet* : Pembaruan status di Twitter yang disiarkan ke pengguna lain.



INTISARI

PeduliLindungi adalah aplikasi yang dikembangkan untuk membantu instansi pemerintah terkait dalam melakukan pelacakan untuk menghentikan penyebaran Coronavirus Disease (COVID-19). Aplikasi ini mengandalkan partisipasi masyarakat untuk saling membagikan data lokasinya saat bepergian agar penelusuran riwayat kontak dengan penderita COVID-19 dapat dilakukan.

Aplikasi ini sudah digunakan masyarakat luas guna untuk mendapatkan sertifikat vaksin dan menjadi persyaratan verifikasi melakukan bepergian menggunakan transportasi publik, dan aktivitas di tempat publik, harapannya dengan adanya aplikasi PeduliLindungi dapat mempermudah administrasi juga mendapatkan informasi mengenai COVID-19 disekitarnya, namun tak jarang masyarakat mengalami kendala pada saat menggunakan aplikasi ini.

Banyaknya opini dan kritikan masyarakat di sosial media Twitter tentang aplikasi PeduliLindungi, banyaknya kritikan dan opini publik yang tersebar pada media sosial Twitter, data *tweets* tersebut nantinya akan diproses melalui *scraping*, *text processing* setelahnya dilakukan analisis sentimen menggunakan algoritma *Naïve Bayes* dan diklasifikasikan menjadi dua label positif dan negatif. Berdasarkan hasil klasifikasi untuk tanggapan tentang aplikasi PeduliLindungi menggunakan *Naïve Bayes Classifier* diperoleh hasil nilai *accuracy* 85.40%, nilai *recall* 89%, nilai *precision* 83% pada kelas negatif dan nilai *recall* 81%, nilai *precision* 88% pada kelas positive.

Kata Kunci: Analisis Sentimen, *Naïve Bayes*, *Data Mining*, COVID-19, Twitter, pedulilindungi, *e-hac*

ABSTRACT

PeduliLindungi is an application developed to assist relevant government agencies in tracking to stop the spread of Coronavirus Disease (COVID-19). This application relies on community participation to share location data with each other while traveling so that contact history tracing with COVID-19 sufferers can be carried out. Users of this application will also get a notification if they are in a crowd or are in a red zone, namely an area or sub-district where it has been recorded that there are people who are infected with COVID-19 positive or there are patients under surveillance.

The PeduliLindungi application has been used by the wider community, in addition to obtaining a vaccine certificate, this application is a verification requirement when traveling using air or land transportation and activities in public places, the public hopes that the PeduliLindungi application can facilitate administration as well as obtain information about COVID-19 around it, but not infrequently people experience problems when using this application.

There are many public opinions and criticisms on Twitter social media about the PeduliLindungi application, the criticisms and public opinions are spread. The tweets data will go through a scraping process, text processing after which sentiment analysis is carried out using the Naïve Bayes algorithm and classified into two parts, labelled positive and negative. Based on the classification results for responses about Pedulilindungi application using the Naïve Bayes Classifier, the results obtained are 85.40% accuracy, 89% recall value, 83% precision value in negative class and 81% recall value, 88% precision value in positive class.

Keywords: *Sentiment Analysis, Naïve Bayes, Data Mining, COVID-19, Twitter, pedulilindungi, e-hac*