

**ANALISIS SENTIMEN TERHADAP ULASAN PENGGUNA
APLIKASI PEDULILINDUNGI MENGGUNAKAN ALGORITMA
SUPPORT VECTOR MACHINE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



diajukan oleh

DIMAS PRAMANA INDRATAMA

18.11.2407

Kepada

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2022

**ANALISIS SENTIMEN TERHADAP ULASAN PENGGUNA
APLIKASI PEDULILINDUNGI MENGGUNAKAN ALGORITMA
SUPPORT VECTOR MACHINE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



diajukan oleh

DIMAS PRAMANA INDRATAMA

18.11.2407

Kepada

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2022

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP ULASAN PENGGUNA APLIKASI
PEDULILINDUNGI MENGGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE**

yang disusun dan diajukan oleh

Dimas Pramana Indratama

18.11.2407

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 Agustus 2022

Dosen Pembimbing,

Theopilus Bayu Sasongko, S.Kom., M.Eng.

NIK. 190302375

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP ULASAN PENGGUNA APLIKASI
PEDULILINDUNGI MENGGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE**

yang disusun dan diajukan oleh

Dimas Pramana Indratama

18.11.2407

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Agustus 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anggit Dwi Hartanto, M.Kom
NIK. 190302163

Erni Seniwati, S.Kom, M.Cs
NIK. 190302231

Theopilus Bayu Sasongko, S.Kom, M.Eng
NIK. 190302375

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

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Dimas Pramana Indratama**
NIM : **18.11.2407**

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi PeduliLindungi Menggunakan Algoritma Support Vector Machine

Dosen Pembimbing : **Theopilus Bayu Sasongko, 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, 24 Agustus 2022

Yang Menyatakan,



Dimas Pramana Indratama

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam atas barokah, rahmat dan karunia Allah SWT yang telah memberi kemudahan, kesehatan dan kelancaran kepada penulis dalam menyelesaikan skripsi ini dengan sebaik-baiknya. Karya ini menjadi persembahan bagi orang-orang yang penulis kasihi dan sayangi.

1. Kepada Papah dan Mamah, yang selalu memberi motivasi, dukungan dan tentunya do'a yang tidak pernah putus. Terimakasih Pah, terimakasih Mah atas semua yang telah kalian berikan. Semoga Allah SWT selalu memberi kalian kesehatan dan umur yang panjang untuk menemaniku hingga menuju kesuksesan.
2. Kepada seluruh Keluarga yang tidak dapat saya sebutkan satu – persatu, terimakasih saya ucapkan atas kebaikan dan kasih sayang yang telah diberikan kepada saya.
3. Kepada Bapak Theopilus Bayu Sasongko, S.Kom., M.Eng. atas bimbingannya selama penelitian ini berlangsung.
4. Kepada mas Satrio Yudho Pangestu, S.Kom. yang telah bersedia membagikan ilmu yang bermanfaat bagi peneliti. Terimakasih untuk ilmunya, semoga keakraban kita selalu terjaga.
5. Kepada Bapak dan Ibu kost, terimakasih telah menjaga penulis selama tinggal di Yogyakarta.
6. Kepada para penghuni Kost Bloops, terimakasih atas keramahan, canda tawa, dan bantuan yang telah diberikan selama ini. Semoga kita dapat meraih kesuksesan bersama.
7. Kepada seluruh teman 18 Informatika 09, yang sudah menjadi teman yang baik untuk saya.
8. Serta semua pihak yang tidak dapat disebutkan satu per-satu.

KATA PENGANTAR

Alhamdulillahirobbil'alamin, Puji dan syukur bagi Allah SWT atas berkah, rahmat, dan nikmat yang diberikan kepada hamba-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “*Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi PeduliLindungi Menggunakan Algoritma Support Vector Machine*”. Skripsi ini disusun sebagai syarat mendapatkan gelar Sarjana Komputer dari Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Tidak dapat disangkal bahwa butuh usaha yang keras dalam penyelesaian penyusunan skripsi ini. Namun, karya ini tidak akan selesai tanpa orang-orang tercinta di sekeliling penulis yang mendukung dan membantu. Terima kasih penulis sampaikan kepada:

1. Allah SWT atas rahmat, hidayah, serta karunia-Nya yang telah diberikan kepada penulis sehingga tugas akhir ini dapat terselesaikan.
2. Bapak Prof. Dr. M. Suyanto, MM. Selaku Rektor Universitas Amikom Yogyakarta.
3. Bapak Theopilus Bayu Sasongko, S.Kom., M.Eng. selaku dosen pembimbing yang telah memberikan banyak arahan bagi penulis dalam penyusunan skripsi ini.
4. Kedua orang tua penulis dan keluarga yang telah mendukung dan selalu mendo'akan penulis selama ini.
5. Para Dosen dan Staff Universitas Amikom Yogyakarta yang telah berperan penting dalam memberikan ilmu pengetahuan dan pengalaman selama masa perkuliahan.
6. Serta seluruh pihak yang telah membantu penulis, mohon maaf tidak dapat penulis sebutkan satu per-satu.

Yogyakarta, 24 Agustus 2022

Penulis

DAFTAR ISI

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 LAMPIRAN.....	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1. Literature Review.....	5
2.2. Landasan Teori.....	8
2.2.1. Data Mining.....	8
2.2.2. Analisis Sentimen.....	9
2.2.3. Text Mining.....	9
2.2.4. Preprocessing.....	10
2.2.5. Term Frequency-Inverse Document Frequency (TF-IDF).....	10
2.2.6. Support Vector Machine	11
2.2.7. K-Fold Cross Validation	14
2.2.8. Confusion Matrix	14
2.2.9. Flowchart.....	15
2.2.10. Python.....	17
2.2.11. Twitter API.....	17

BAB III METODOLOGI PENELITIAN.....	19
3.1. Gambaran Umum Objek Penelitian	19
3.2. Instrument Penelitian	19
3.2.1. Bahan Penelitian.....	19
3.2.2. Alat Penelitian	19
3.3. Alur Penelitian	20
3.4. Langkah-Langkah Penelitian	20
3.4.1. Pengumpulan Data	20
3.4.2. Pelabelan Data.....	21
3.4.3. Preprocessing.....	21
3.4.4. Pembagian Data.....	22
3.4.5. Pembobotan TF-IDF.....	22
3.4.6. Model Klasifikasi SVM.....	22
3.5. Analisis Algoritma	24
3.5.1. Pengumpulan Data	24
3.5.2. Pelabelan Data.....	24
3.5.3. Case Folding.....	25
3.5.4. Cleaning Data	26
3.5.5. Tokenizing.....	26
3.5.6. Normalisasi Kata	27
3.5.7. Stopword Removal	28
3.5.8. Stemming.....	29
3.5.9. Pembobotan Term Frequency-Inverse Document Frequency	30
3.5.10. Klasifikasi Support Vector Machine	36
3.6. Rancangan Antarmuka	39
3.6.1. Halaman Upload File.....	40
3.6.2. Halaman Klasifikasi	40
BAB IV HASIL DAN PEMBAHASAN	42
4.1 Pengumpulan Data Tweet	42
4.1.1. API Twitter.....	42
4.1.2. Pengunduhan Data Tweet.....	43
4.2 Pelabelan Data.....	44

4.3	Preprocessing	45
4.3.1.	Case Folding	45
4.3.2.	Cleaning.....	46
4.3.3.	Tokenize	46
4.3.4.	Normalisasi.....	46
4.3.5.	Stopword Removal	47
4.3.6.	Stemming.....	48
4.4	Pembagian Data	48
4.5	Pelatihan Model Klasifikasi Support Vector Machine.....	49
4.5.1.	Pembobotan Term Frequency-Inverse Document Frequency	49
4.5.2.	Klasifikasi Support Vector Machine	49
4.6	Klasifikasi dengan Model Support Vector Machine.....	50
4.6.1.	Perancangan Antarmuka.....	50
4.6.2.	Klasifikasi Data	52
4.7	Evaluasi Model Klasifikasi Support Vector Machine.....	53
4.7.1.	Confusion Matrix	53
4.7.2.	K-Fold Cross Validation	55
4.8	Visualisasi Data Hasil Klasifikasi.....	57
4.8.1.	Wordcloud	57
4.8.1.	Pie Chart	58
BAB V KESIMPULAN DAN SARAN.....		59
5.1.	Kesimpulan	59
5.2.	Saran.....	59
DAFTAR PUSTAKA		60
LAMPIRAN.....		62

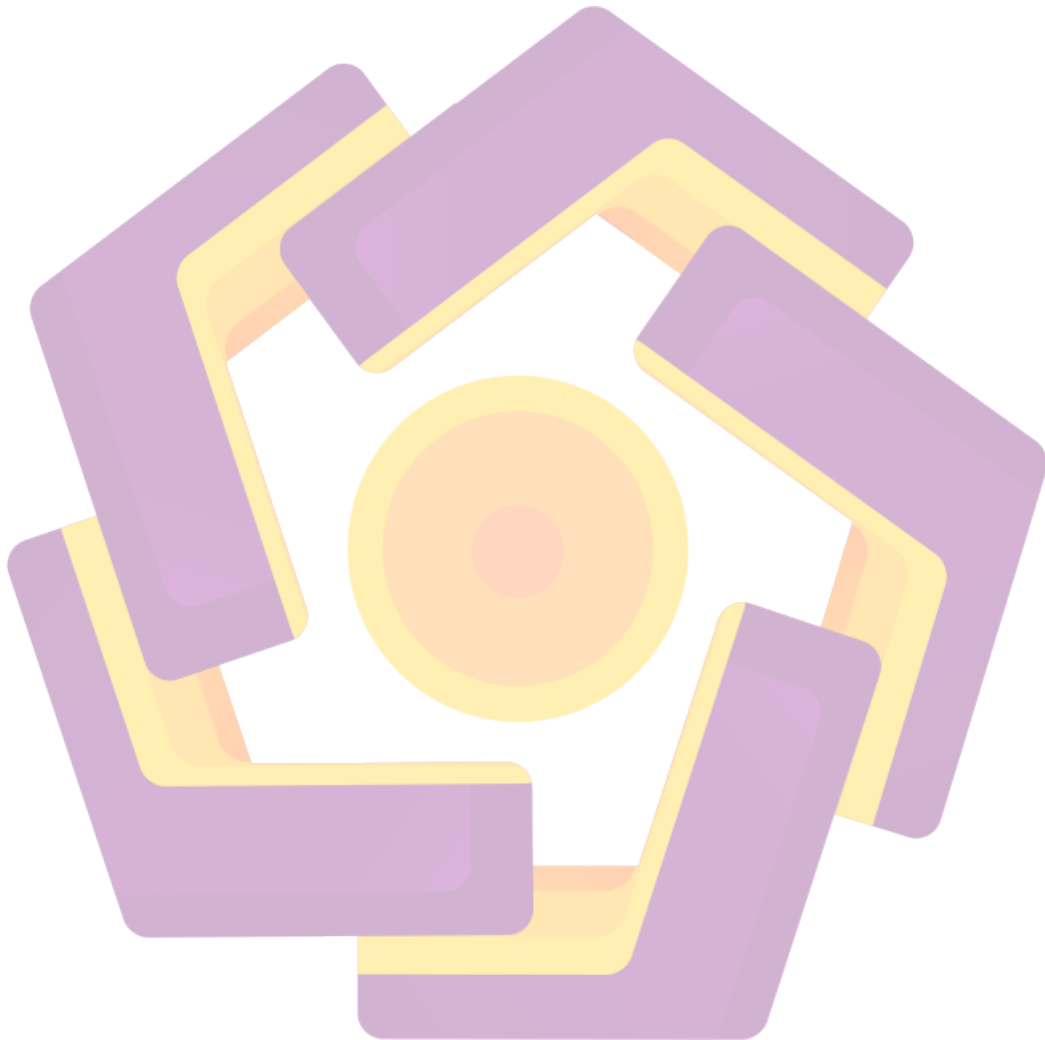
DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian	6
Tabel 2.2 Confusion Matrix	14
Tabel 2.3 Daftar Simbol Dokumen Flowchart.....	16
Tabel 3.1 Data tweet yang telah diperoleh.....	24
Tabel 3.2 Data tweet yang telah diberikan label.....	24
Tabel 3.3 Case Folding	25
Tabel 3.4 Cleaning Data	26
Tabel 3.5 Tokenizing	26
Tabel 3.6 Normalisasi Kata.....	27
Tabel 3.7 Stopword Removal.....	28
Tabel 3.8 Stemming	29
Tabel 3.9 Term Hasil Preprocessing	30
Tabel 3.10 Term Frequency Kelas Positif	30
Tabel 3.11 Term Frequency Kelas Negatif.....	31
Tabel 3.12 Document Frequency	32
Tabel 3.13 Inverse Document Frequency	33
Tabel 3.14 Term Frequency-Inverse Document Frequency Positif.....	34
Tabel 3.15 Term Frequency-Inverse Document Frequency Negatif	35
Tabel 3.16 Hasil Akhir TF-IDF	36
Tabel 3.17 Persamaan w_1 , w_2 dan b	38
Tabel 3.18 Klasifikasi SVM	38
Tabel 3.19 Confusion Matrix	39
Tabel 3.20 Accuracy, Precision, Recall Dan F-Score.....	39
Tabel 4.1 Confusion Matrix	54
Tabel 4.2 Nilai Accuracy, Precision, Recall dan F-Score.....	54
Tabel 4.3 Hasil Evaluasi K-Fold Cross Validation.....	56

DAFTAR GAMBAR

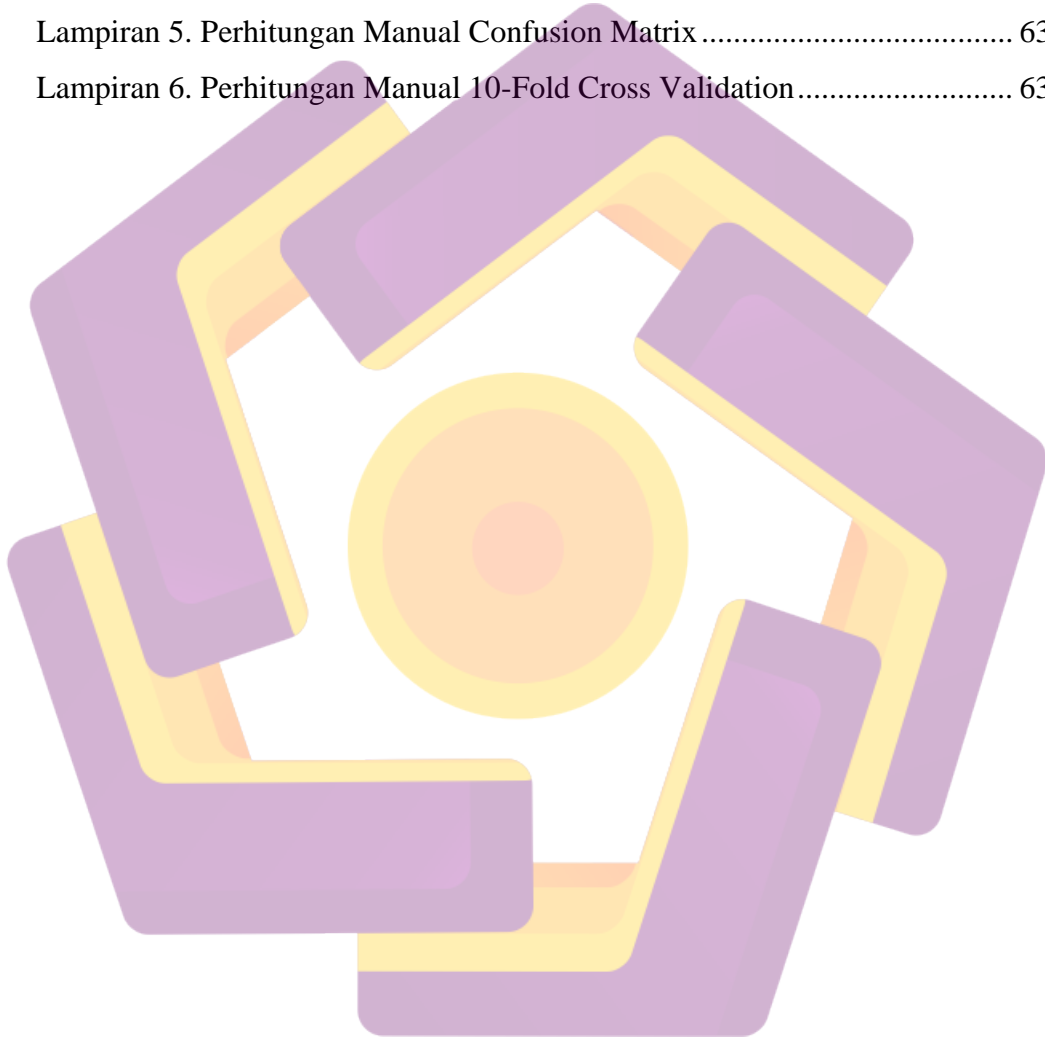
Gambar 2.1 Contoh Pencarian Hyperplane Dua Dimensi	12
Gambar 2.2 Contoh Hyperplane Terbaik.....	12
Gambar 2.3 Hyperplane Dengan Margin Maksimum.....	13
Gambar 3.1 Flowchart Alur Penelitian	20
Gambar 3.2 Diagram Alir Pengumpulan Data.....	21
Gambar 3.3 Diagram Alir Preprocessing.....	21
Gambar 3.4 Diagram Alir Preprocessing.....	23
Gambar 3.5 Halaman Upload File	40
Gambar 3.6 Halaman Klasifikasi	41
Gambar 4.1 Tampilan API Twitter	42
Gambar 4.2 Script Pengunduhan Data Twitter	43
Gambar 4.3 Hasil Pengunduhan Data Tweet	44
Gambar 4.4 Hasil Pelabelan Data	45
Gambar 4.5 Script Case Folding	45
Gambar 4.6 Script Cleaning.....	46
Gambar 4.7 Script Tokenize	46
Gambar 4.8 Contoh Kamus Normalisasi	47
Gambar 4.9 Script Normalisasi.....	47
Gambar 4.10 Script Stopword Removal	47
Gambar 4.11 Script Stemming.....	48
Gambar 4.12 Script Pembagian Data	48
Gambar 4.13 Script Term Frequency-Inverse Document Frequency	49
Gambar 4.14 Script Pembuatan Model Support Vector Machine	50
Gambar 4.15 Script Save Model	50
Gambar 4.16 Halaman Upload File	51
Gambar 4.17 Halaman Klasifikasi	51
Gambar 4.18 Script Import Model Klasifikasi.....	52
Gambar 4.19 Script Prediksi	52
Gambar 4.20 Script Confusion Matrix.....	53

Gambar 4.21 Script K-Fold Cross Validation..... 55
Gambar 4.22 Grafik Hasil Evaluasi K-Fold Cross Validation..... 56
Gambar 4.23 Wordcloud Sentimen Positif 57
Gambar 4.24 Wordcloud Sentimen Negatif..... 58
Gambar 4.25 Pie Chart..... 58



DAFTAR LAMPIRAN

Lampiran 1. Perhitungan Manual TF-IDF Term Positif	62
Lampiran 2. Perhitungan Manual TF-IDF Term Negatif	62
Lampiran 3. Hasil TF-IDF	62
Lampiran 4. Perhitungan dan Prediksi SVM	63
Lampiran 5. Perhitungan Manual Confusion Matrix	63
Lampiran 6. Perhitungan Manual 10-Fold Cross Validation	63



INTISARI

Coronavirus Disease 2019 adalah wabah virus yang menyebar ke seluruh dunia, termasuk Indonesia. Pemerintah Indonesia sendiri telah melakukan berbagai macam upaya dalam menghentikan penyebaran virus ini. Salah satunya adalah meluncurkan aplikasi yang bernama PeduliLindungi melalui Kementerian Komunikasi dan Informatika (KOMINFO). Aplikasi ini bertujuan untuk melakukan penelusuran, pelacakan dan pemberian peringatan dalam menghentikan penyebaran virus Covid-19 di Indonesia.

Dalam penerapannya, aplikasi PeduliLindungi tentunya mengundang berbagai macam opini dari masyarakat Indonesia. Twitter adalah salah satu media sosial populer yang biasanya digunakan untuk membagikan informasi seperti opini, kritik, dan saran. Dari ketersediaan informasi tersebut, dapat dilakukan sebuah penelitian yang bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi PeduliLindungi pada Twitter. Sentimen yang terkandung akan diklasifikasikan kedalam 2 kelas, yaitu positif dan negatif. Metode yang akan digunakan meliputi *Preprocessing*, pembobotan kata TF-IDF, klasifikasi dengan *Support Vector Machine* dan evaluasi dengan *K-Fold Cross Validation*.

Hasil dari penelitian ini menunjukkan metode *Support Vector Machine* menggunakan *10-Fold Cross Validation* menghasilkan tingkat rata-rata *accuracy* sebesar 85%. Adapun *accuracy* tertinggi berada pada *fold* ke-3 dengan nilai *accuracy* sebesar 91%.

Kata Kunci: *Coronavirus Disease 2019*, PeduliLindungi, Analisis Sentimen, *Support Vector Machine*

ABSTRACT

Coronavirus Disease 2019 is a virus outbreak that has spread throughout the world, including Indonesia. The Indonesian government itself has made various efforts to stop the spread of this virus. One of them launched an application called PeduliLindungi by the Ministry of Communication and Information (KOMINFO). This application aims to tracing, tracking, warning and fencing in stopping the spread of the Covid-19 virus in Indonesia.

In its application, the PeduliLindungi application certainly invites various kinds of opinions from the Indonesian people. Twitter is one of the popular social media that is usually used to share information such as opinions, criticisms, and suggestions. From the availability of this information, study can be carried out which aims to conduct sentiment analysis on user reviews of the PeduliLindungi application on Twitter. The sentiments contained will be classified into 2 classes, namely positive and negative. The methods to be used include Preprocessing, TF-IDF word weighting, classification with Support Vector Machine and evaluation with K-Fold Cross Validation.

The results of this study indicate the Support Vector Machine algorithm using 10-Fold Cross Validation produces an average accuracy rate of 85%. The highest accuracy is in the 3rd fold with an accuracy value of 91%.

Keywords: *Coronavirus Disease 2019, PeduliLindungi, Sentiment Analysis, Support Vector Machine*