

**KLASIFIKASI SENTIMEN PUBLIK TERHADAP PROSES
VAKSINASI COVID-19 DI INDONESIA BERDASARKAN
METODE *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE***

SKRIPSI



RITA SUSANTI

**PROGRAM STUDI S-1 STATISTIKA
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA**

2022

**KLASIFIKASI SENTIMEN PUBLIK TERHADAP PROSES
VAKSINASI COVID-19 DI INDONESIA BERDASARKAN
METODE *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE***

SKRIPSI



RITA SUSANTI

**PROGRAM STUDI S-1 STATISTIKA
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA**

2022

**KLASIFIKASI SENTIMEN PUBLIK TERHADAP PROSES
VAKSINASI COVID-19 DI INDONESIA BERDASARKAN
METODE *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE***

SKRIPSI

Sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Statistika
pada Fakultas Sains dan Teknologi
Universitas Airlangga

Rita Susanti
NIM 081811833040

Disetujui oleh:

Pembimbing I,



Dr. Nur Chamidah, S.Si., M.Si.

NIP. 197206021998022001

Pembimbing II,



Dr. Toha Saifudin, S.Si., M.Si.

NIP. 197501061999031002

LEMBAR PENGESAHAN NASKAH SKRIPSI

Judul : Klasifikasi Sentimen Publik Terhadap Proses Vaksinasi
Covid-19 di Indonesia Berdasarkan Metode *Naïve Bayes*
dan *Support Vector Machine*

Penyusun : Rita Susanti

NIM : 081811833040

Pembimbing I : Dr. Nur Chamidah, S.Si., M.Si.

Pembimbing II : Dr. Toha Saifudin, S.Si., M.Si.

Tanggal Seminar : 13 Juli 2022

Disetujui oleh:

Pembimbing I,



Dr. Nur Chamidah, S.Si., M.Si.

NIP. 197206021998022001

Pembimbing II,



Dr. Toha Saifudin, S.Si., M.Si.

NIP. 197501061999031002

Mengetahui,

Ketua Departemen Matematika
Fakultas Sains dan Teknologi
Universitas Airlangga



Dr. Herry Suprajitno, S.Si., M.Si.

NIP. 196804041994031020

Koordinator Prodi S-1 Statistika
Fakultas Sains dan Teknologi
Universitas Airlangga



Drs. H. Sediono, M.Si.

NIP. 196107121987011001

PEDOMAN PENGGUNAAN SKRIPSI

Skripsi ini tidak dipublikasikan, namun tersedia di perpustakaan dalam lingkungan Universitas Airlangga, diperkenankan untuk dipakai sebagai referensi kepustakaan, tetapi pengutipan harus seizin penyusun dan harus menyebutkan sumbernya sesuai kebiasaan ilmiah. Dokumen skripsi ini merupakan hak milik Universitas Airlangga.

SURAT PERNYATAAN TENTANG ORISINALITAS

Yang bertanda tangan di bawah ini, saya:

Nama : Rita Susanti

NIM : 081811833040

Program Studi : Statistika

Fakultas : Sains dan Teknologi Universitas Airlangga

Jenjang : Sarjana (S1)

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul:

Klasifikasi Sentimen Publik Terhadap Proses Vaksinasi Covid-19 di Indonesia Berdasarkan Metode *Naïve Bayes* dan *Support Vector Machine*.

Apabila suatu saat nanti melakukan tindakan plagiat, maka saya akan menerima sanksi yang telah ditetapkan. Demikian surat pernyataan ini saya buat dengan sebenar-benarnya.

Surabaya, 19 Juli 2022



Rita Susanti

NIM. 081811833040

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah Swt. karena rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Sentimen Publik Terhadap Proses Vaksinasi Covid-19 di Indonesia Berdasarkan Metode *Naïve Bayes* dan *Support Vector Machine*”. Pada kesempatan ini penulis juga ingin mengucapkan terima kasih kepada:

1. Prof. Dr. Mohammad Nasih, S.E., M.T., Ak., CMA. selaku Rektor Universitas Airlangga, Prof. Dr. Moh Yasin, M. Si. selaku Dekan Fakultas Sains dan Teknologi Universitas Airlangga, dan Drs. H. Sediono, M.Si. selaku Ketua Program Studi S-1 Statistika Universitas Airlangga yang telah memberikan kesempatan pada penulis untuk menimba ilmu di Program Studi Statistika Universitas Airlangga.
2. Orangtua tercinta yang selalu memberikan dukungan dan doa untuk penulis
3. Ibu Dr. Nur Chamidah, S.Si., M.Si. selaku dosen pembimbing I dan Bapak Dr. Toha Saifudin, S.Si., M.Si. selaku dosen pembimbing II yang senantiasa memberikan nasehat, bimbingan serta dukungan, serta Bapak Suliyanto, Drs., M.Si. selaku dosen wali penulis. Seluruh dosen Statistika Universitas Airlangga yang telah memberikan ilmu dan pengalaman yang bermanfaat selama perkuliahan.
4. Teman-teman penulis, khususnya Meika dan Najma dan semua pihak yang selalu memberikan dukungan dalam penyusunan skripsi ini.
5. Karakter Na Hee Do dalam drama *Twenty-Five Twenty-One* yang menginspirasi penulis untuk terus berjuang menyelesaikan naskah skripsi ini.

Penulis menyadari bahwa dalam penyusunan skripsi ini terdapat banyak kekurangan sehingga penulis mengharapkan kritik dan saran yang membangun dari semua pihak guna menyempurnakan naskah skripsi ini.

Surabaya, 1 Juli 2022

Penulis,

Rita Susanti

Rita Susanti, 2022. **Klasifikasi Sentimen Publik Terhadap Proses Vaksinasi Covid-19 di Indonesia Berdasarkan Metode *Naïve Bayes* dan *Support Vector Machine***. Skripsi di bawah Dr. Nur Chamidah, S.Si., M.Si. dan Dr. Toha Saifudin, S.Si., M.Si., Program Studi S-1 Statistika, Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya.

ABSTRAK

Pandemi Covid-19 telah menyebabkan krisis global yang berdampak pada seluruh aspek kehidupan individu. Tidak ada pengobatan yang efektif kecuali vaksinasi skala besar. Namun, proses vaksinasi di Indonesia terkesan lambat dibandingkan negara-negara maju yang telah mencapai *herd immunity*. Hal ini dikarenakan banyak permasalahan yang terjadi di Indonesia seperti kelangkaan pasokan vaksin, pendistribusian vaksin yang lama, tidak adanya laporan data vaksinasi secara *real-time* serta adanya keraguan di masyarakat untuk melakukan vaksinasi. Hal ini membuktikan perlunya evaluasi proses vaksinasi guna menganalisis aspek-aspek yang perlu diperbaiki. Dalam mencapai tujuan tersebut, klasifikasi sentimen dilakukan menggunakan *text mining* berdasarkan metode *Naïve Bayes* dan *Support Vector Machine*. Penelitian ini menggunakan 1000 *tweets* periode 18 Agustus sampai 16 September 2021 yang diperoleh dari media sosial *Twitter* dengan teknik *crawling* menggunakan API. Data tersebut dibagi menjadi 90% data *training* dan 10% data *testing* dengan *10-fold cross validation*. Hasil analisis deskriptif menunjukkan bahwa kebijakan vaksinasi Covid-19 dan keterbatasan informasi yang benar merupakan aspek permasalahan yang dikritik oleh publik. Sementara hasil klasifikasi berdasarkan *Naïve Bayes* menunjukkan akurasi sebesar 95,43% sedangkan SVM menunjukkan akurasi sebesar 76,30% yang menunjukkan *Naïve Bayes* lebih baik dalam mengklasifikasikan sentimen positif dan negatif. Selain itu, *Naïve Bayes* juga lebih baik dalam mengklasifikasikan sentimen positif dengan nilai *sensitivity* sebesar 95,72% namun SVM lebih baik dalam mengklasifikasikan sentimen negatif dengan nilai *specificity* sebesar 100%. Dengan demikian, dapat disimpulkan bahwa *Naïve Bayes* secara umum menghasilkan performansi yang lebih baik dan sesuai dengan studi pada kasus penelitian ini dibandingkan SVM.

Kata Kunci: *Proses Vaksinasi Covid-19, Analisis Sentimen, Text Mining, Naïve Bayes, Support Vector Machine*

Rita Susanti, 2022. *Classification of Public Sentiment on Covid-19 Vaccination Process in Indonesia Based on Naïve Bayes and Support Vector Machine Method*. The undergraduate thesis is guided by Dr. Nur Chamidah, S.Si., M.Si., and Dr. Toha Saifudin, S.Si., M.Si., Statistics Study Program, Mathematics Department, Faculty of Science and Technology, Airlangga University, Surabaya.

ABSTRACT

The Covid-19 pandemic has caused a global crisis that has affected all aspects of individual life. Vaccinations are the only effective treatment. However, the vaccination process in Indonesia appears to be slow in comparison to developed countries that have achieved herd immunity. This is due to a number of issues in Indonesia, such as vaccine supply scarcity, long distribution times, a lack of real-time data reports, and public skepticism about vaccination. This shows the importance of evaluating the vaccination process to identify areas for improvement. Sentiment classification based on Naïve Bayes and Support Vector Machine is used to achieve this goal. This study uses 1000 tweets obtained from Twitter using a crawling technique with the API from August 18 to September 16, 2021. The data is divided into 90% training data and 10% testing data, with 10-fold cross validation. The descriptive analysis result shows that public criticizes the Covid-19 vaccination policy and the lack of accurate information. Meanwhile, the accuracy of the classification based on Naïve Bayes is 95.43%, while SVM is 76.30%, indicating that Naïve Bayes is better at classifying positive and negative sentiments. Furthermore, while Naïve Bayes is better at classifying positive sentiments with a sensitivity value of 95.72%, SVM is better at classifying negative sentiments with a specificity value of 100%. As a result, it is possible to conclude that Naïve Bayes produces better overall performance and is more consistent with the study case in this research than SVM.

Keywords: *Covid-19 Vaccination Process, Sentiment Classification, Text Mining, Naïve Bayes, Support Vector Machine*

DAFTAR ISI

LEMBAR JUDUL	i
HALAMAN PERNYATAAN	ii
LEMBAR PENGESAHAN NASKAH SKRIPSI.....	iii
PEDOMAN PENGGUNAAN SKRIPSI	iv
SURAT PERNYATAAN TENTANG ORISINALITAS	v
KATA PENGANTAR	vi
ABSTRAK.....	vii
DAFTAR ISI.....	viix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian.....	7
1.4 Manfaat Penelitian.....	7
BAB II TINJAUAN PUSTAKA.....	8
2.1 Analisis Sentimen.....	8
2.2 <i>Text Mining</i>	9
2.3 <i>Naïve Bayes</i>	13
2.3.1 Simulasi <i>Naïve Bayes</i>	17
2.4 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	20
2.4.1 Simulasi TF-IDF	21
2.5 <i>Support Vector Machine</i>	23
2.5.1 Simulasi <i>Support Vector Machine</i>	29
2.6 Evaluasi Kerja	31

2.6.1	<i>K-fold Cross Validation</i>	31
2.6.2	Kurva <i>Receiver Operating Characteristic</i> (ROC).....	32
2.7	<i>Word Cloud</i>	34
2.8	Proses Vaksinasi Covid-19 di Indonesia	35
2.9	<i>Twitter</i>	36
2.10	<i>Software R</i>	37
BAB III METODE PENELITIAN.....		41
3.1	Data dan Sumber Data.....	41
3.2	Variabel Penelitian	41
3.3	Langkah Analisis Data	41
BAB IV HASIL DAN PEMBAHASAN		41
4.1	Karakteristik Sentimen Publik Terhadap Proses Vaksinasi Covid-19 di Indonesia Menggunakan Statistika Deskriptif dan <i>Word Cloud</i>	41
4.1.1	Evaluasi Proses Vaksinasi Covid-19 di Indonesia	50
4.2	Klasifikasi Sentimen Publik Terhadap Proses Vaksinasi Covid-19 di Indonesia Berdasarkan <i>Naïve Bayes</i>	52
4.2.1	Klasifikasi Sentimen Publik pada Data <i>Training</i>	53
4.2.2	Klasifikasi Sentimen Publik pada Data <i>Testing</i>	55
4.3	Klasifikasi Sentimen Publik Terhadap Proses Vaksinasi Covid-19 di Indonesia Berdasarkan <i>Support Vector Machine</i>	57
4.3.1	Klasifikasi Sentimen Publik pada Data <i>Training</i>	57
4.3.2	Klasifikasi Sentimen Publik pada Data <i>Testing</i>	59
4.4	Komparasi Performansi <i>Naïve Bayes</i> dan SVM dalam Mengklasifikasikan Sentimen Publik Terhadap Proses Vaksinasi Covid-19 di Indonesia	60
BAB V PENUTUP.....		62
5.1	Kesimpulan.....	62
5.2	Saran.....	63
DAFTAR PUSTAKA		64
LAMPIRAN		73

DAFTAR TABEL

Nomor	Judul Tabel	Halaman
2.1	Contoh Proses <i>Cleaning</i>	10
2.2	Contoh Proses <i>Case Folding</i>	10
2.3	Contoh Proses <i>Tokenization</i>	11
2.4	Contoh Proses <i>Stemming</i>	12
2.5	Contoh Proses <i>Stopword Removal</i>	13
2.6	Contoh Data <i>Training</i> dan Data <i>Testing</i>	17
2.7	Frekuensi Kemunculan Kata pada Setiap Dokumen	17
2.8	Nilai Probabilitas Kata pada Setiap Kategori	18
2.9	Contoh Dokumen Setelah Tahap <i>Preprocessing</i>	21
2.10	Hasil Kata Penting pada Data Simulasi	22
2.11	Hasil Simulasi Perhitungan Pembobotan Kata TF-IDF	23
2.12	Data Inputan Klasifikasi SVM	29
2.13	Nilai Fungsi Keputusan Simulasi Perhitungan Model Klasifikasi SVM	31
2.14	Hasil Klasifikasi Sentimen pada Data Simulasi Menggunakan SVM	31
2.15	<i>Confusion Matrix</i>	32
3.1	Contoh Struktur Data Penelitian	41
4.1	Evaluasi Proses Vaksinasi Covid-19 di Indonesia	50
4.2	Pembagian Data untuk <i>Training</i> dan <i>Testing</i> pada <i>Naïve Bayes</i>	53
4.3	Performansi <i>Naïve Bayes</i> pada Data <i>Training</i>	54
4.4	Performansi <i>Naïve Bayes</i> pada Data <i>Testing</i>	55
4.5	Pembagian Data untuk <i>Training</i> dan <i>Testing</i> pada SVM	57
4.6	Performansi SVM pada Data <i>Training</i>	58
4.7	Performansi SVM pada Data <i>Testing</i>	59
4.8	Komparasi Performansi <i>Naïve Bayes</i> dan SVM	61

DAFTAR GAMBAR

Nomor	Judul Gambar	Halaman
2.1	<i>Hyperplane</i> Pemisah Dua Kelas Data	24
2.2	<i>Hyperplane</i> dan <i>Margin</i>	26
2.3	Kurva ROC	34
2.4	Contoh <i>Word Cloud</i>	35
3.1	Diagram Alir Langkah Analisis Data	45
4.1	Persentase Jenis Sentimen Publik terhadap Proses Vaksinasi Covid-19 di Indonesia	46
4.2	<i>Word Cloud</i> pada Sentimen Positif	48
4.3	Frekuensi Kemunculan Kata Khusus pada Sentimen Positif	48
4.4	<i>Word Cloud</i> pada Sentimen Negatif	49
4.5	Frekuensi Kemunculan Kata Khusus pada Sentimen Negatif	49
4.6	Kurva ROC pada Data <i>Training</i> Berdasarkan <i>Naïve Bayes</i>	54
4.7	Kurva ROC pada Data <i>Testing</i> Berdasarkan <i>Naïve Bayes</i>	56
4.8	Kurva ROC pada Data <i>Training</i> Berdasarkan SVM	58
4.9	Kurva ROC pada Data <i>Testing</i> Berdasarkan SVM	60

DAFTAR LAMPIRAN

Nomor	Judul
1	<i>Script Program Crawling Data Tweet</i> tentang Proses Vaksinasi Covid-19 di Indonesia
2	<i>Script Program Preprocessing Data Tweet</i> tentang Proses Vaksinasi Covid-19 di Indonesia
3	Data Penelitian Komentar Publik tentang Proses Vaksinasi Covid-19 di Indonesia
4	Hasil <i>Preprocessing</i> Dokumen Komentar Publik tentang Proses Vaksinasi Covid-19 di Indonesia
5	Frekuensi Kata Khusus pada Sentimen Positif Publik terhadap Proses Vaksinasi Covid-19 di Indonesia
6	Frekuensi Kata Khusus pada Sentimen Negatif Publik terhadap Proses Vaksinasi Covid-19 di Indonesia
7	<i>Script Program Partisi Data</i>
8	Probabilitas Kemunculan Masing-Masing Kata
9	<i>Script Program Klasifikasi Naïve Bayes</i> pada Data <i>Tweet</i> tentang Proses Vaksinasi Covid-19 di Indonesia
10	Hasil Klasifikasi Sentimen Publik terhadap Proses Vaksinasi Covid-19 di Indonesia pada Data <i>Training</i> Berdasarkan <i>Naïve Bayes</i>
11	<i>Confusion Matrix</i> pada Data <i>Training</i> untuk Semua <i>Fold</i> Berdasarkan <i>Naïve Bayes</i>
12	Hasil Klasifikasi Sentimen Publik terhadap Proses Vaksinasi Covid-19 di Indonesia pada Data <i>Testing</i> Berdasarkan <i>Naïve Bayes</i>
13	<i>Confusion Matrix</i> pada Data <i>Testing</i> untuk Semua <i>Fold</i> Berdasarkan <i>Naïve Bayes</i>
14	Hasil Pembobotan Kata dengan TF-IDF pada Dokumen Komentar Publik tentang Proses Vaksinasi Covid-19 di Indonesia
15	<i>Script Program Klasifikasi SVM</i> pada Data <i>Tweet</i> tentang Proses Vaksinasi Covid-19 di Indonesia

- 16 Nilai Koefisien Lagrange pada *Support Vector* untuk Sentimen Publik terhadap Proses Vaksinasi Covid-19 di Indonesia
- 17 Nilai Fungsi Keputusan Model Klasifikasi SVM
- 18 Hasil Klasifikasi Sentimen Publik terhadap Proses Vaksinasi Covid-19 di Indonesia pada Data *Training* Berdasarkan SVM
- 19 *Confusion Matrix* pada Data *Training* untuk Semua *Fold* Berdasarkan SVM
- 20 Hasil Klasifikasi Sentimen Publik terhadap Proses Vaksinasi Covid-19 di Indonesia pada Data *Testing* Berdasarkan SVM
- 21 *Confusion Matrix* pada Data *Testing* untuk Semua *Fold* Berdasarkan SVM