



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

**PERBANDINGAN ALGORITMA SVM DAN DECISION TREE
DALAM KLASIFIKASI SENTIMEN JENIS VAKSIN COVID-19
DI INDONESIA PADA SOSIAL MEDIA TWITTER**



SKRIPSI

Kevin

00000020527

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2022**

**PERBANDINGAN ALGORITMA SVM DAN DECISION TREE
DALAM KLASIFIKASI SENTIMEN JENIS VAKSIN COVID-19
DI INDONESIA PADA SOSIAL MEDIA TWITTER**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Sistem Infomasi

Kevin

(00000020527)

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2022

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Kevin
Nomor Induk Mahasiswa : **00000020527**
Program studi : Sistem Informasi

Skripsi dengan judul:

Perbandingan Algoritma SVM dan Decision Tree dalam Klasifikasi Sentimen Jenis Vaksin COVID-19 di Indonesia pada Media Sosial Twitter, merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 28 Maret 2022



(Kevin)

UMIN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

“Perbandingan Algoritma SVM dan Decision Tree dalam Klasifikasi Sentimen
Jenis Vaksin COVID-19 di Indonesia pada Media Sosial Twitter”

Oleh

Nama : Kevin
NIM : 00000020527
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Kamis, 14 April 2022
Pukul 13.00 s/d 15.00 dan dinyatakan
LULUS

Dengan susunan pengujian sebagai berikut.

Ketua Sidang



Johan Setiawan, S.Kom., MM.
327106402

Penguji



Samuel Ady Sanjaya, S.T., M.T.
0305049402/075049

Pembimbing



Wella S.Kom, M.MSI
305119101/033470

Ketua Program Studi Sistem Informasi



Ririn Ikana Desanti, S. Kom., M.Kom
313058001/051314

iii

Perbandingan Algoritma SVM dan Decision Tree dalam Klasifikasi Sentimen Jenis Vaksin
COVID-19 di Indonesia pada Media Sosial Twitter. Kevin, Universitas Multimedia Nusantara

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas academica Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Kevin
NIM : 00000020527
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Jenis Karya : *Tesis/~~Skripsi~~/Tugas Akhir (*coret salah satu)

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul.

“Perbandingan Algoritma SVM dan Decision Tree dalam Klasifikasi Sentimen Jenis Vaksin COVID-19 di Indonesia pada Media Sosial Twitter” Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 28 Maret 2022

Yang menyatakan,



UNIVERSITAS
MULTIMEDIA
NUSANTARA (Kevin)

KATA PENGANTAR

Puji Syukur atas selesainya penulisan Skripsi ini dengan judul: “Perbandingan Algoritma SVM dan Decision Tree dalam Klasifikasi Sentimen Jenis Vaksin COVID-19 di Indonesia pada Media Sosial Twitter” dilakukan untuk memenuhi salah satu syarat untuk lulus pada matakuliah Metodologi Riset Sistem Informasi jurusan Sistem Informasi Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono. selaku Rektor Universitas Multimedia Nusantara.
2. Ibu Dr. Friska Natalia, S.Kom., M.T., selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S. Kom., M.Kom. selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ibu Wella, S.Kom., M.MSI., COBIT5 sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Keluarga, pasangan, sahabat dan juga teman-teman saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.

Semoga karya ilmiah ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 28 Maret 2022



Kevin

PERBANDINGAN ALGORITMA SVM DAN DECISION TREE DALAM KLASIFIKASI SENTIMEN JENIS VAKSIN COVID-19 DI INDONESIA PADA SOSIAL MEDIA TWITTER

(Kevin)

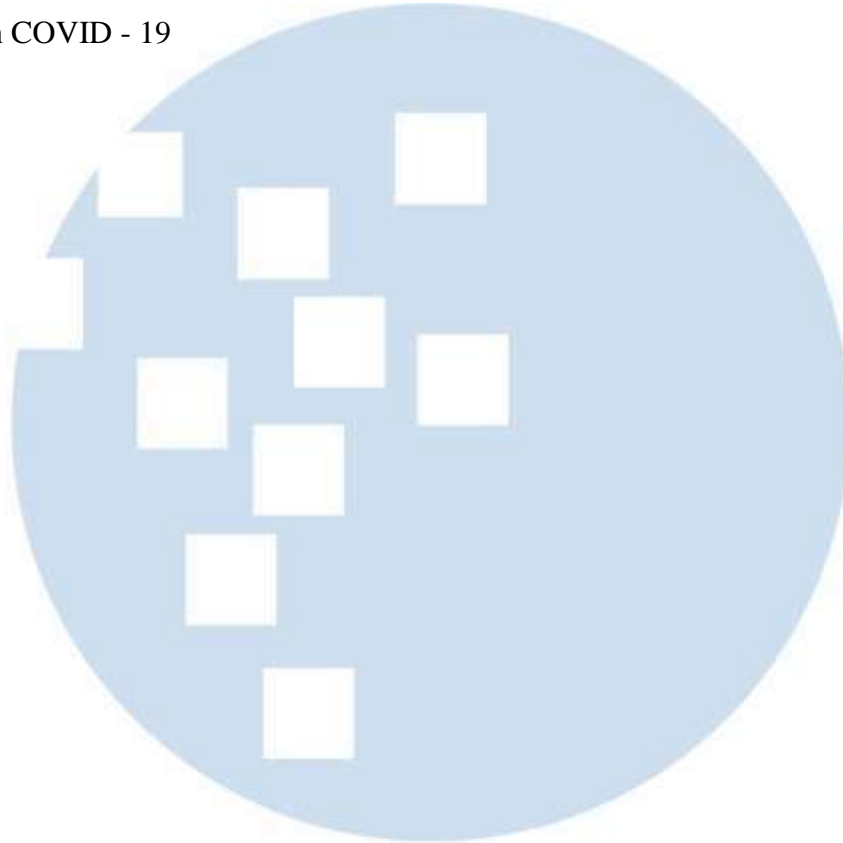
ABSTRAK

COVID-19 adalah virus yang hingga saat ini masih melanda seluruh dunia, pertama kali ditemukan pada tahun 2019 tepatnya di kota Wuhan, China. Cepatnya persebaran membuat semua orang dengan mudah terjangkit, efek yang ditimbulkan yang paling parah adalah kematian. Vaksin adalah salah satu cara yang bisa digunakan untuk menghalau pesebaran virus COVID-19. Namun, kehadiran sesuatu biasanya akan memicu opini public baik pro maupun kontra, termasuk juga vaksin COVID-19. Opini tentunya bisa disampaikan dimanapun salahsatunya adalah media sosial *Twitter*.

Pada penelitian ini akan dilakukan analisa sentimen masyarakat terkait jenis vaksin COVID-19 dengan menggunakan algoritma *Decision Tree* dan SVM. Data diambil pada bulan Februari dan April 2022 dengan total data sebanyak 7.316. selanjutnya data tersebut akan melewati tahap *preprocessing* dan klasifikasi dengan pembagian *data training* dan *testing* 80:20 dengan bantuan *K-fold Cross Validation*. Hasil akan ditampilkan dengan metode *confusion matrix*.

Hasil penelitian menunjukkan algoritma SVM memiliki akurasi yang lebih tinggi dibandingkan dengan *Decision Tree*, dimana dari kelima data vaksin, SVM mendapatkan akurasi yang lebih tinggi untuk kelima jenis vaksin seperti AstraZeneca, Moderna, Pfizer, Sinopharm, dan Sinovac dengan akurasi 81,75%, 73,22%, 75,48%, 76,97%, 71,14%. Sedangkan untuk hasil akurasi *Decsion Tree* lebih rendah pada kelima vaksin dengan angka akurasi 72,45% untuk AstraZeneca, 61,23% untuk Moderna, 61,49% untuk Pfizer, 73,35% untuk Sinopharm, 68,38% untuk Sinovac. Sehingga dapat disimpulkan SVM lebih baik dibandingkan dengan *Decision Tree*. Untuk hasil sentiment menunjukkan bahwa masyarakat saat ini sudah terbuka dengan jenis-jenis vaksin yang beredar dimana setiap vaksin yang diteliti mendapatkan sentiment positif lebih banyak dibandingkan negatif.

Kata kunci: Analisis Sentimen, *Confusion Matrix*, *Decision Tree*, SVM, Twitter, Vaksin COVID - 19



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

**COMPARISON BETWEEN SVM AND DECISION TREE
ALGORITHM IN CLASSIFICATION OF SENTIMENT TYPES
OF COVID-19 VACCINE IN INDONESIA ON TWITTER**

(Kevin)

ABSTRACT (English)

COVID-19 is a virus that has hit the whole world until now, first discovered in 2019 in the city of Wuhan, China. The fast spread makes everyone easily infected, the most severe effect is death. Vaccines are one way that can be used to prevent the spread of the COVID-19 virus. However, the presence of something will usually trigger public opinion both for and against, including the COVID-19 vaccine. Opinion can certainly convey one of which is social media Twitter.

In this study, an analysis of public sentiment related to the type of COVID-19 vaccine will be carried out using the Decision Tree and SVM algorithms. Data was taken in February and April 2022 with a total of 7,316 data. Furthermore, the data will go through the preprocessing and classification stages by distributing 80:20 training and testing data with the help of K-fold Cross Validation. The results will be displayed using the confusion matrix method.

The results showed the SVM algorithm had higher accuracy than Decision Tree, where from the five vaccine's data, SVM got higher accuracy for the five types of vaccines such as AstraZeneca, Moderna, Pfizer, Sinopharm, and Sinovac with an accuracy of 81.75%, 73.22%, 75.48%, 76.97%, 71.14%. Meanwhile, the accuracy of the Decision Tree was lower for the five vaccines with an accuracy rate of 72.45% for AstraZeneca, 61.23% for Moderna, 61.49% for Pfizer, 73.35% for Sinopharm, 68.38% for Sinovac. So it can be concluded that SVM is better than Decision Tree. Sentiment results show that the public is now open to the types of vaccines in circulation, where each vaccine studied has more positive sentiments than negative ones.

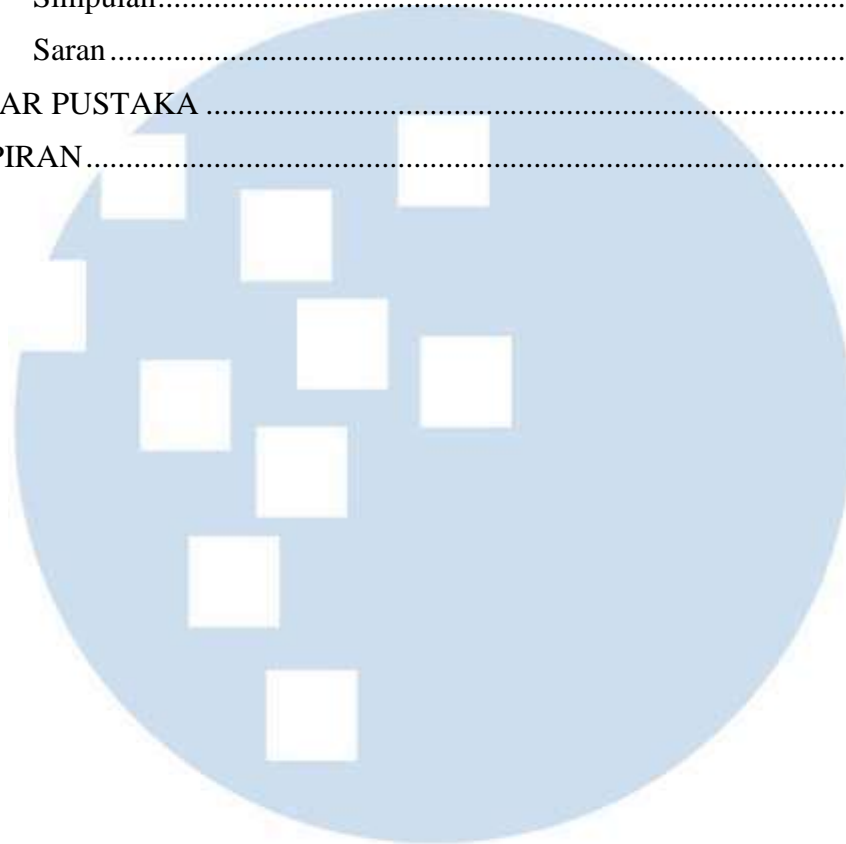
Keywords: *Confusion Matrix, COVID-19 Vaccine, Decision Tree, Sentiment Analysis, SVM, Twitter*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	iv
KATA PENGANTAR	v
ABSTRAK	vi
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR RUMUS.....	1
BAB I PENDAHULUAN	2
1.1. Latar Belakang.....	2
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan dan Manfaat Penelitian.....	6
1.4.1. Tujuan Penelitian	6
1.4.2. Manfaat Penelitian	6
BAB II LANDASAN TEORI.....	8
2.1. <i>Sentiment Analysis</i>	8
2.2. <i>Twitter</i>	8
2.3. <i>Text Mining</i>	9
2.4. <i>Text Preprocessing</i>	9
2.5. <i>RapidMiner</i>	10
2.6. <i>Decision Tree</i>	10
2.7. Support Vector Machine.....	11
2.8. K-Fold Cross Validation.....	11
2.9. <i>Confusion Matrix</i>	12
2.10. Penelitian Terdahulu	14

BAB III METODOLOGI PENELITIAN.....	20
3.1. Objek Penelitian	20
3.2. Variabel Penelitian	20
3.3. Teknik Pengumpulan Data	20
3.4. Metode Penelitian.....	21
3.4.1. Kerangka Pikir	22
3.4.2. <i>Text Preprocessing</i>	23
3.4.3. <i>Modelling</i>	24
3.4.4. Evaluasi	24
BAB IV	25
ANALISIS DAN HASIL PENELITIAN.....	25
4.1. <i>Text Scrapping</i>	25
4.2. <i>Data Cleansing</i>	32
4.2.1. <i>Remove URL</i>	32
4.2.2. <i>Remove Non-alphabet</i>	33
4.2.3. <i>Remove RT</i>	34
4.2.4. <i>Remove Duplicates dan Write Excel</i>	34
4.3. <i>Labelling</i>	36
4.4. <i>Text Preprocessing</i>	39
4.4.1. <i>Case Folding</i>	40
4.4.2. <i>Tokenize</i>	40
4.4.3. <i>Filtering</i>	41
4.4.4. <i>Stemming</i>	42
4.4.5. TD-IDF.....	43
4.5. <i>Modelling</i>	44
4.6. Hasil Pengujian dan Evaluasi	48
4.6.1. AstraZeneca.....	48
4.6.2. Moderna	50
4.6.3. Pfizer	53
4.6.4. Sinopharm	56
4.6.5. Sinovac.....	59
4.6.6. Evaluasi	62

BAB V.....	65
SIMPULAN DAN SARAN	65
5.1. Simpulan.....	65
5.2. Saran.....	65
DAFTAR PUSTAKA	66
LAMPIRAN.....	73



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2. 1 Contoh Visualisasi Confusion Matrix	12
Tabel 2. 2 Penelitian Terdahulu	14
Tabel 3. 1 Tabel Perbandingan Decision Tree dan SVM.....	21
Tabel 4. 1 Contoh Pemberian Label Pada Data Tweets	39
Tabel 4. 2 Contoh Pembobotan Labelling.....	39
Tabel 4. 3 Tabel Case Folding	40
Tabel 4. 4 Tabel Tokenize.....	41
Tabel 4. 5 Tabel Filtering.....	42
Tabel 4. 6 Tabel Stemming	42
Tabel 4. 7 Tabel Perbandingan Sentimen Vaksin COVID-19 Decision Tree.....	62
Tabel 4. 8 Tabel Perbandingan Sentimen Vaksin COVID-19 SVM.....	62
Tabel 4. 9 Perbandingan Akurasi Tiap Jenis Vaksin	63



DAFTAR GAMBAR

Gambar 1. 1 Platform Sosial Media Yang Paling Sering Digunakan di Indonesia	3
Gambar 3. 1 Kerangka Pikir Penelitian Modifikasi	22
Gambar 4. 1 Create Connection	25
Gambar 4. 2 Memilih Jenis Koneksi, Letak, dan Nama	26
Gambar 4. 3 Meminta Access Token	26
Gambar 4. 4 Koneksi Berhasil di Buat	27
Gambar 4. 5 Flow Text Scrapping	28
Gambar 4. 6 Parameters Search Twitter	29
Gambar 4. 7 Parameters Select Attributes	30
Gambar 4. 8 Contoh Hasil Text Scrapping AstraZeneca	31
Gambar 4. 9 Flow Data Cleansing	32
Gambar 4. 10 Parameter Remove URL	33
Gambar 4. 11 Parameter Non-alphabet	33
Gambar 4. 12 Parameter Remove RT	34
Gambar 4. 13 Contoh Hasil Data Cleansing	35
Gambar 4. 14 Flow Automation Labelling	36
Gambar 4. 15 Parameters Operator Extract Sentiment	37
Gambar 4. 16 Contoh Hasil Pelabelan Otomatis	38
Gambar 4. 17 Parameters Operator Transform Cases	40
Gambar 4. 18 Operator Tokenize	41
Gambar 4. 19 Operator Filter Stopwords	41
Gambar 4. 20 Operator Stem	42

Gambar 4. 21 Parameter TD-IDF.....	43
Gambar 4. 22 Pembagian Data atau Data Split.....	44
Gambar 4. 23 Flow Modelling.....	45
Gambar 4. 24 Parameters Cross Validation.....	45
Gambar 4. 25 Cross Validation Decision Tree.....	46
Gambar 4. 26 Cross Validation SVM.....	46
Gambar 4. 27 Parameters Decision Tree.....	47
Gambar 4. 28 Confusion Matrix AstraZeneca (Decision Tree).....	48
Gambar 4. 29 Confusion Matris AstraZeneca (SVM).....	48
Gambar 4. 30 Diagram Donut AZ Decision Tree.....	49
Gambar 4. 31 Diagram Donut AZ SVM.....	50
Gambar 4. 32 Confusion Matrix Moderna (Decision Tree).....	50
Gambar 4. 33 Confusion Matrix Moderna (SVM).....	51
Gambar 4. 34 Diagram Donut Moderna Decision Tree.....	52
Gambar 4. 35 Diagram Donut Moderna SVM.....	53
<i>Gambar 4. 36 Confusion Matrix Pfizer (Decision Tree).....</i>	<i>53</i>
Gambar 4. 37 Confusion Matrix Pfizer (SVM).....	54
Gambar 4. 38 Diagram Donut Pfizer Decision Tree.....	55
<i>Gambar 4. 39 Diagram Donut Pfizer SVM.....</i>	<i>56</i>
Gambar 4. 40 Confusion Matrix Sinopharm (Decision Tree).....	56
Gambar 4. 41 Confusion Matrix Sinopharm (SVM).....	57
Gambar 4. 42 Diagram Donut Sinopharm Decision Tree.....	58
Gambar 4. 43 Diagram Donut Sinopharm SVM.....	59

<i>Gambar 4. 44 Confusion Matrix Sinovac (Decision Tree)</i>	59
Gambar 4. 45 Confusion Matrix Sinovac (SVM)	60
Gambar 4. 46 Diagram Donut Sinovac Decision Tree	61
Gambar 4. 47 Diagram Donut Sinovac SVM	62
Gambar 4. 48 Perbandingan Algoritma Decision Tree dan SVM	64



DAFTAR RUMUS

Rumus 2. 1 Rumus Perhitungan Akurasi	13
Rumus 2. 2 Rumus Perhitungan Precision.....	13
Rumus 2. 3 Rumus Perhitungan Recall.....	13



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA