

**ANALISIS SENTIMEN PADA DOKUMEN *TWITTER*  
MENGENAI VAKSIN COVID-19  
DENGAN *DAMERAU LEVENSHTAIN DISTANCE*  
DAN *SUPPORT VECTOR MACHINE***

**TESIS**



oleh  
**I MADE DWI RAKA MUDIARTA**  
**NIM 1829101006**

**UNIVERSITAS PENDIDIKAN GANESHA  
PASCASARJANA  
PROGRAM STUDI ILMU KOMPUTER  
JANUARI 2022**



**ANALISIS SENTIMEN PADA DOKUMEN *TWITTER*  
MENGENAI VAKSIN COVID-19  
DENGAN *DAMERAU LEVENSHTEIN DISTANCE*  
DAN *SUPPORT VECTOR MACHINE***

**TESIS**

**Diajukan kepada**

**Universitas Pendidikan Ganesha**

**Untuk Memenuhi Sebagian Persyaratan**

**Memperoleh Gelar Magister Komputer**

**Program Studi Ilmu Komputer**

oleh

**I MADE DWI RAKA MUDIARTA**

**NIM 1829101006**



**UNIVERSITAS PENDIDIKAN GANESHA  
PASCASARJANA  
PROGRAM STUDI ILMU KOMPUTER  
JANUARI 2022**

Tesis oleh I Made Dwi Raka Mudiarta ini telah diperiksa dan disetujui untuk mengikuti ujian Tesis

Singaraja, Januari 2022

Pembimbing I



Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom.

NIP. 198407242015041002

Pembimbing II



Dr. Komang Setemen, S.Si., M.T.

NIP. 197603152001121002



Tesis oleh I Made Dwi Raka Mudiarta ini telah berhasil dipertahankan di depan tim penguji dan dinyatakan diterima sebagai salah satu persyaratan untuk memperoleh gelar Magister Ilmu Komputer di Program Studi Ilmu Komputer, Pascasarjana, Universitas Pendidikan Ganesha.

Disetujui pada tanggal:        Februari 2022

Oleh

.....  
.....  
.....  
.....  
.....  
.....

Tim Penguji  
Dr. Gede Indrawan, S.T., M.T.  
NIP.197601022003121001  
Dr. I Nyoman Sukajaya, M.T.  
NIP.196711151993031001  
Dr. I Gede Aris Gunadi, S.Si. M.Kom.  
NIP.197703182008121004  
Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom.  
NIP.198407242015041002  
Dr. Komang Setemen, S.Si., M.T.  
NIP.197603152001121002

Mengetahui Direktur

Pascasarjana Universitas Pendidikan Ganesha,



Prof.Dr. I Gusti Putu Suharta, M.Si.

NIP.196212151988031022

## LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Komputer dari Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila di kemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, 17 Maret 2022

Yang memberi pernyataan,



I Made Dwi Raka Mudiarta



## PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas anugrah-Nya, sehingga tesis yang berjudul: “Analisis Sentimen pada Dokumen Twitter Mengenai Vaksin COVID-19 dengan Damerau Levenshtein Distance dan Support Vector Machine”, dapat diselesaikan sesuai dengan yang direncanakan.

Tesis ini ditulis untuk memenuhi salah satu persyaratan untuk memperoleh gelar Megister Komputer Pascasarjana Universitas Pendidikan Ganesha pada Program Studi Ilmu Komputer. terselesaikannya tesis ini telah banyak memperoleh uluran tangan dari berbagai pihak. Untuk itu, ijinakan penulis menyampaikan terima kasih dan penghargaan kepada pihak-pihak berikut.

1. Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom., sebagai pembimbing I yang telah dengan sabar membimbing, mengarahkan, dan memberikan motivasi yang demikian bermakna, sehingga penulis mampu melewati berbagai hambatan dalam menjalankan studi dan penyelesaian tesis ini;
2. Dr. Komang Setemen, S.Si., M.T., sebagai pembimbing II yang memberikan motivasi, semangat, harapan kepada penulis selama penelitian dan penulisan naskah laporan tesis ini, sehingga tesis dapat terwujud dengan baik dan sesuai harapan;
3. Ketua Program Studi Ilmu Komputer dan staf dosen pengajar yang telah banyak membantu dan memotivasi penulis selama penyusunan tesis ini;
4. Direktur Pascasarjana Undiksha dan staf, yang telah banyak membantu selama penulis menyelesaikan tesis ini;

5. Rektor Universitas Pendidikan Ganesha, yang telah memberikan bantuan secara moral dan memfasilitasi berbagai kepentingan penulis dalam menyelesaikan tesis ini;
6. Rekan-rekan seangkatan di Program Studi Ilmu Komputer yang dengan karakternya masing-masing telah banyak berkontribusi membentuk kemandirian penulis selama menjalani studi dan penyelesaian tesis ini;
7. Bapak, Ibu, Istri dan anak yang telah banyak membantu secara material dan moral selama penyelesaian tesis ini.

Semoga semua bantuan yang telah mereka berikan dalam menyelesaikan studi ini, mereka diberkati imbalan yang sepadan oleh Tuhan Yang Maha Esa, kesehatan, dan keharmonisan dalam menjalani kehidupan.

Penulis menyadari bahwa Tesis ini belum sempurna. Namun, kehadirannya dalam konstelasi masyarakat akademis akan menambah perbendaharaan ilmu dalam perkembangan ilmu pengetahuan. Semoga tesis ini bermanfaat bagi masyarakat akademis, terutama mereka yang menyatakan diri bernaung di bawah kebesaran panji-panji pendidikan.

Singaraja, Maret 2022

Penulis



## DAFTAR ISI

	Halaman
ABSTRAK.....	iv
ABSTRACT.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR.....	viii
DAFTAR TABEL.....	ix
DAFTAR LAMPIRAN.....	xi
BAB I PENDAHULUAN.....	1
1. 1. Latar Belakang.....	1
1. 2. Identifikasi Masalah.....	6
1. 3. Pembatasan Masalah.....	6
1. 4. Rumusan Masalah.....	6
1. 5. Tujuan Penelitian.....	7
1. 6. Manfaat Penelitian.....	7
BAB II KAJIAN PUSTAKA.....	9
2. 1. Media Sosial Twitter.....	9
2. 2. Teks Preprocessing.....	10
2. 3. Algoritma Damerau Levenshtein Distance.....	12
2. 4. Feature Extraction.....	14
2. 5. Support Vector Machine.....	16
2. 6. Kajian Hasil Penelitian yang Relevan.....	17
BAB III METODE PENELITIAN.....	22
3. 1. Rancangan Penelitian.....	22
3. 2. Pengumpulan Data.....	25
3. 3. Preprocessing Data.....	26
3. 4. Feature Extraction.....	31
3. 5. SVM Training.....	33
3. 6. SVM Testing.....	33
3. 7. Hasil Klasifikasi SVM.....	33
3. 8. Pengujian.....	36
BAB IV HASIL DAN PEMBAHASAN.....	38
4. 1. Pengumpulan Data Twitter.....	38
4. 2. <i>Text Preprocessing</i> Data Komentar Twitter.....	45

4. 3. 1. Convert Emoticon .....	45
4. 3. 2. Cleansing .....	46
4. 3. 3. Case Folding .....	48
4. 3. 4. Terjemahan Kata Slangword .....	49
4. 3. 5. Stemming .....	49
4. 3. 6. Stopword Removal .....	50
4. 3. 7. Penerapan Algoritma Damerau Levenshtein Distance .....	52
4. 3. 8. Convert Negation .....	54
4. 3. 9. Tokenizing .....	55
4. 3. Feature Extraction .....	56
4. 3. 1. Term Frequency (TF) .....	56
4. 3. 2. Term Frequency – Invers Document Frequency (TF-IDF) .....	58
4. 4. SVM .....	59
4. 4. 1. Pengujian SVM .....	63
4. 4. 1. 1. Pengujian SVM Menggunakan Algoritma Damerau Levenshtein Distance pada Text Preprocessing .....	64
4. 4. 1. 2. Pengujian SVM Tanpa Menggunakan Algoritma Damerau Levenshtein Distance pada Text Preprocessing .....	65
4. 4. 2. Analisis Hasil .....	66
<b>BAB V PENUTUP .....</b>	<b>72</b>
5. 1. Rangkuman .....	72
5. 2. Simpulan .....	73
5. 3. Saran .....	74
<b>DAFTAR PUSTAKA .....</b>	<b>76</b>
<b>LAMPIRAN .....</b>	<b>81</b>

## DAFTAR GAMBAR

	Halaman
Gambar 3.1 Rancangan Penelitian Analisis Sentimen Twitter .....	22
Gambar 3.2 Proses <i>Preprocessing</i> teks data <i>twitter</i> .....	26
Gambar 3. 3 Visualisasi data SVM.....	35
Gambar 4. 1 Data Twitter .....	44
Gambar 4. 2 Grafik hasil pengujian SVM .....	66
Gambar 4. 3 Grafik Perbandingan Waktu Teks <i>Preprocessing</i> .....	70



## DAFTAR TABEL

Tabel 2.2 Perhitungan Damerau Levenshtein Distance .....	13
Tabel 2.3 Hasil perbandingan SVM dan <i>Naïve Bayes</i> (Rahat dkk., 2019) .....	19
Tabel 3.2 Pembersihan data <i>twitter</i> .....	27
Tabel 3.3 Contoh proses <i>case folding</i> data <i>twitter</i> .....	28
Tabel 3.3 Terjemahan kata-kata gaul .....	29
Tabel 3.4 Contoh proses <i>stop word removal</i> data <i>twitter</i> .....	29
Tabel 3.4 Contoh proses <i>stemming</i> data <i>twitter</i> .....	30
Tabel 3.5 Contoh proses algoritma Damerau Levenshtein Distance .....	30
Tabel 3.6 Contoh proses <i>tokenizing</i> data <i>twitter</i> .....	31
Tabel 3. 7 Contoh Data Perhitungan SVM .....	34
Tabel 3.8 Tabel pengujian dengan <i>confusion matrix</i> .....	37
Tabel 4. 1 Key API Twitter .....	38
Tabel 4. 2 Hasil Pengumpulan Data .....	39
Tabel 4. 3 Contoh Data yang Dihapus .....	40
Tabel 4. 4 Hasil Pemberian Label oleh Judges .....	41
Tabel 4. 5 Jumlah hasil akhir pemberian label .....	43
Tabel 4. 6 Jumlah Data Trainning dan Data Testing .....	44
Tabel 4. 7 Contoh Komentar Pada Twitter .....	45
Tabel 4. 8 Daftar Emoticon .....	46
Tabel 4. 9 Hasil Proses Cleansing .....	47
Tabel 4. 10 Hasil Tahapan Case Folding .....	48
Tabel 4. 11 Daftar Kata Slangword .....	49
Tabel 4. 12 Hasil dari Proses <i>Stemming</i> .....	50
Tabel 4. 13 Kamus <i>Stopword</i> .....	51
Tabel 4. 14 Hasil Tahapan <i>Stopword Removal</i> .....	51
Tabel 4. 15 Hasil Nilai Jarak Algoritma <i>Damerau Levensthein Distance</i> .....	52
Tabel 4. 16 Kata yang Sering Muncul pada Dokumen .....	53
Tabel 4. 17 Kata yang gagal diperbaiki .....	54
Tabel 4. 18 Daftar Kata Negasi .....	55
Tabel 4. 19 Hasil Proses Tokenizing .....	56

Tabel 4. 20 Data Komentar .....	57
Tabel 4. 21 Hasil <i>Term Frequency</i> .....	57
Tabel 4. 22 Tabel Hasil TF-IDF.....	59
Tabel 4. 23 Data <i>term</i> dan nilai <i>TF-IDF</i> .....	60
Tabel 4. 24 <i>Confusion Matrix</i> Pengujian SVM Menggunakan Algoritma <i>Damerau Levensthein Distance</i> pada <i>Text Preprocessing</i> .....	64
Tabel 4. 25 <i>Confusion Matrix</i> Pengujian SVM Tanpa Menggunakan Algoritma <i>Damerau Levensthein Distance</i> pada <i>Text Preprocessing</i> .....	65



## DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Biodata Responden.....	81
Lampiran 2. Source Code pengambilan Data Twitter.....	84
Lampiran 3. Source Code Pemrosesan Data.....	86
Lampiran 4. Data Tweet Hasil Perbaikan .....	89

