

**PERBANDINGAN METODE NAÏVE BAYES CLASSIFIER DAN
SUPPORT VECTOR MACHINE PADA KASUS ANALISIS SENTIMEN
TERHADAP DATA VAKSIN COVID-19 DI TWITTER**

TESIS

Oleh :

RIZKI ANOM RAHARJO

NIM 1929101020



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

**PERBANDINGAN METODE NAÏVE BAYES CLASSIFIER DAN
SUPPORT VECTOR MACHINE PADA KASUS ANALISIS SENTIMEN
TERHADAP DATA VAKSIN COVID-19 DI TWITTER**

TESIS

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Magister Komputer
Pada Program Studi Ilmu Komputer



RIZKI ANOM RAHARJO

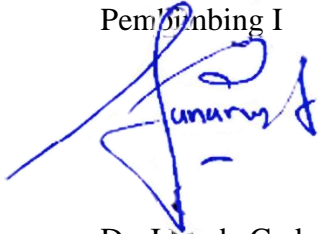
NIM 1929101020

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

Tesis oleh Rizki Anom Raharjo ini telah diperiksa dan disetujui untuk mengikuti Ujian Tesis.

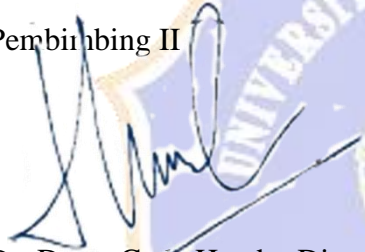
Singaraja, 2 Agustus 2022

Pembimbing I



Dr. I Made Gede Sunarya, S.Kom., M.Cs.
NIP. 198307252008011008

Pembimbing II



Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom.
NIP. 198407242015041002



LEMBAR PERSETUJUAN TIM PENGUJI

Tesis oleh Rizki Anom Raharjo ini telah dipertahankan didepan 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 tanggal : 8 Agustus 2022

Oleh
Tim Penguji

Ketua

Dr. Gede Indrawan, S.T., M.T.
NIP. 197601022003121001

Anggota

Dr. I Made Gede Sunarya, S.Kom., M.Cs.
NIP. 198307252008011008

Anggota

Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom.
NIP. 198407242015041002

Anggota

Dr. Luh Joni Erawati Dewi, S.T., M.Pd.
NIP. 197606252001122001

Mengetahui Direktur
Pascasarjana Undiksha,



Prof. Dr. I Gusti Putu Suharta, M.Si
NIP. 19621215 198803 1 002

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis saya 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, 9 Agustus 2022

Yang memberi pernyataan,



5A545AJX017204510
(Rizki Anom Raharjo)

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa atas segala rahmat dan hidayah-Nya sehingga Tesis dengan judul “Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Kasus Analisis Sentimen Terhadap Data Vaksin Covid-19 Di Twitter” dapat terselesaikan dengan baik. Tesis ini disusun untuk memenuhi sebagian persyaratan dalam menyelesaikan studi pada Program Pascasarjana Universitas Pendidikan Ganesha, Program Studi Ilmu Komputer. Tak lupa penulis menyampaikan terimakasih dan penghargaan yang setinggi-tingginya kepada pihak-pihak terkait yang telah berkontribusi dalam penyusunan tesis ini.

1. Bapak Dr. I Made Gede Sunarya, S.Kom., M.Cs. sebagai pembimbing I yang telah membimbing, memberikan ilmu pengetahuan, mengarahkan dan memotivasi, sehingga penulis mampu menyelesaikan tesis ini.
2. Bapak Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom. sebagai pembimbing II yang telah membimbing, memberikan ilmu pengetahuan, mengarahkan dan memotivasi, sehingga penulis mampu menyelesaikan tesis ini.
3. Rektor Universitas Pendidikan Ganesha, Direktur Program Pascasarjana, Kepala Program Studi Ilmu Komputer, seluruh dosen pengajar dan staf Pascasarjana Universitas pendidikan Ganesha yang telah berjasa atas bimbingan dan arahan selama penulis menjalani studi dan menyelesaikan tesis ini.
4. Kedua orang tua bapak dan ibu tercinta, mertua, istri, anak serta seluruh keluarga yang selalu memberi dukungan dan do'a yang tiada henti untuk kesuksesan dan kelancaran tesis ini.
5. Ibu Dr. Novena Ade Fredyarini Soedjiwo, S.S., M.Hum. sebagai Anotator/Pakar yang telah membantu dalam memberikan pelabelan pada setiap data, sehingga penulis dapat dengan cepat melakukan proses analisa pada data tersebut.

6. Rekan-rekan angkatan 2019 Program Studi Ilmu Komputer yang telah banyak membantu selama menjalani studi dan selama penyelesaian tugas akhir ini.
7. Semua pihak yang tidak dapat disebutkan satu persatu, yang telah banyak memberikan dukungan dan bantuan dalam penyusunan tugas akhir ini.

Semoga semua bantuan dan dukungan yang diberikan mendapatkan balasan berupa keberkahan dan keafiatan oleh Tuhan Yang Maha Esa. Penulis sepenuhnya menyadari bahwa laporan tesis ini masih jauh dari kata sempurna. Oleh karena itu saran dan kritik yang membangun sangat penulis harapkan.

Akhir kata, semoga tesis ini dapat memberikan manfaat dan memberikan wawasan bagi penulis sendiri dan para pembaca serta bermanfaat bagi masyarakat akademis.



Denpasar, 17 September 2022
Penulis

DAFTAR ISI

COVER	i
HALAMAN JUDUL	ii
LEMBAR PERSETUJUAN	iii
LEMBAR PERSETUJUAN TIM PENGUJI	iv
LEMBAR PERNYATAAN	v
KATA PENGANTAR	vi
ABSTRAK	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	3
1.3 Batasan Masalah	3
1.4 Rumusan Masalah.....	3
1.5 Tujuan Penelitian	4
1.6 Manfaat Penelitian	4
BAB II KAJIAN PUSTAKA	5
2.1 Kajian Teori	5
2.2 Analisis Sentimen	5
2.3 <i>Text Preprocessing</i>	5
2.3.1 <i>Case Folding</i>	6
2.3.2 <i>Tokenizing</i>	6
2.3.3 <i>Filtering (Stopword Removal)</i>	6
2.3.4 <i>Stemming</i>	7
2.4 Klasifikasi	7

2.5	Metode <i>Naïve Bayes Classifier</i> (NBC)	7
2.6	<i>Term Frequency - Inverse Document Frequency</i> (TF-IDF)	12
2.7	Metode <i>Support Vector Machine</i> (SVM)	12
2.8	<i>Twitter Crawling</i>	14
2.9	Validasi Model	14
2.9.1	<i>Hold-Out Validation</i>	14
2.9.2	<i>K-Fold Cross Validation</i>	15
2.10	<i>Confusion Matrix</i>	15
2.10.1	<i>Accuracy</i>	16
2.10.2	<i>Precision</i>	16
2.10.3	<i>Recall</i>	17
2.10.4	<i>F1-Score</i>	17
2.11	Kajian Hasil Penelitian yang Relevan	17
2.12	Kerangka Berpikir	20
BAB III METODE PENELITIAN		22
3.1	Metode Pengumpulan Data	22
3.2	Data Penelitian	22
3.3	Tahapan Penelitian	23
3.4	Kendala Penelitian	25
3.4.1	Kendala Mendapatkan Data	25
3.4.2	Kendala Menganalisis Data	25
BAB IV HASIL PENELITIAN DAN PEMBAHASAN		27
4.1	Populasi dan Sampel Penelitian	27
4.2	Pengambilan Data	27
4.3	<i>Text Preprocessing</i>	28
4.3.1	<i>Case Folding</i>	28
4.3.2	<i>Tokenizing</i>	30
4.3.3	<i>Filtering (Stopword Removal)</i>	31
4.3.4	<i>Stemming</i>	32
4.4	Ekstraksi Fitur	34
4.5	Pelatihan dan Pengujian Data	38

4.6	Klasifikasi	39
4.6.1	Klasifikasi <i>Naïve Bayes Classifier</i> (NBC)	39
4.6.2	Klasifikasi <i>Support Vector Machine</i> (SVM).....	41
4.7	Evaluasi Model	43
4.8	Perbandingan Hasil Metode NBC dan SVM	46
BAB V PENUTUP		48
5.1	Kesimpulan	48
5.2	Saran	48
DAFTAR PUSTAKA		49
LAMPIRAN.....		53



DAFTAR TABEL

Tabel 2. 1 Tabel <i>Confusion Matrix</i>	16
Tabel 2. 2 Tabel Penelitian Terkait	18
Tabel 3. 1 Contoh Dataset	21
Tabel 4. 1 Proses <i>Case Folding</i>	29
Tabel 4. 2 Proses <i>Tokenizing</i>	30
Tabel 4. 3 Proses <i>Filtering (Stopword Removal)</i>	31
Tabel 4. 4 Proses <i>Stemming</i>	33
Tabel 4. 5 <i>Term Frequency (TF)</i>	35
Tabel 4. 6 <i>Document Frequency (DF)</i>	36
Tabel 4. 7 <i>Inverse Document Frequency (IDF)</i>	37
Tabel 4. 8 Nilai TF-IDF	37
Tabel 4. 9 Data <i>Training</i> dan Data <i>Testing</i>	39
Tabel 4. 10 Probabilitas Prior	40
Tabel 4. 11 Hasil <i>Confusion Matrix SVM</i>	44
Tabel 4. 12 Nilai <i>Accuracy, Precision, Recall, dan f1-Score</i> Metode SVM	43
Tabel 4. 13 Hasil <i>Confusion Matrix NBC</i>	45
Tabel 4. 14 Nilai <i>Accuracy, Precision, Recall, dan f1-Score</i> Metode NBC	44
Tabel 4. 15 Hasil Perbandingan Performansi NBC dan SVM	46

DAFTAR GAMBAR

Gambar 2. 1 <i>Hyperplane</i> yang memisahkan dua kelas positif dan negatif.....	13
Gambar 2. 2 Kerangka Berpikir	21
Gambar 3. 1 Tahapan Penelitian	23
Gambar 4. 1 Kode Program <i>Crawling</i> data dengan <i>Library Snsrape</i>	27
Gambar 4. 2 Kode Program Proses <i>Case Folding</i>	29
Gambar 4. 3 Kode Program Proses <i>Tokenizing</i>	31
Gambar 4. 4 Kode Program Proses <i>Filtering</i>	32
Gambar 4. 5 Kode Program Proses <i>Stemming</i>	34
Gambar 4. 6 Kode Program Proses TF-IDF.....	38
Gambar 4. 7 Kode Program Klasifikasi <i>Naïve Bayes Classifier</i>	41
Gambar 4. 8 <i>Hyperplane</i> pada SVM.....	42
Gambar 4. 9 Kode Program Klasifikasi <i>Support Vector Machine</i>	43



DAFTAR LAMPIRAN

Lampiran 1. Proses Mempersiapkan Dataset	52
Lampiran 2. Proses <i>Text Preprocessing</i>	53
Lampiran 3. Proses Pembobotan Kata	55
Lampiran 4. Proses Klasifikasi	55

