

**PERBANDINGAN AKURASI METODE *NAIVE BAYES CLASSIFIER*
DAN *SUPPORT VECTOR MACHINE (SVM)* PADA ANALISIS
SENTIMEN ULASAN APLIKASI MYTELKOMSEL**

SKRIPSI



Disusun Oleh :

VERONIKA PASKALIA SINAMBELA
NPM.19081010083

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR**

2024

**PERBANDINGAN AKURASI METODE *NAIVE BAYES CLASSIFIER*
DAN *SUPPORT VECTOR MACHINE* (SVM) PADA ANALISIS
SENTIMEN ULASAN APLIKASI MYTELKOMSEL**

SKRIPSI

**Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar
Sarjana Komputer Program Studi Informatika**



Oleh:

**VERONIKA PASKALIA SINAMBELA
NPM.19081010083**

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2024

LEMBAR PENGESAHAN SKRIPSI

Judul : PERBANDINGAN AKURASI METODE NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE (SVM) PADA ANALISIS SENTIMEN ULASAN APLIKASI MYTELKOMSEL

Oleh : VERONIKA PASKALIA SINAMBELA

NPM : 19081010083

Telah Diseminarkan Dalam Ujian Skripsi Pada:

Hari Jumat, Tanggal 05 Januari 2024

Mengetahui

Dosen Pembimbing

Dosen Penguji

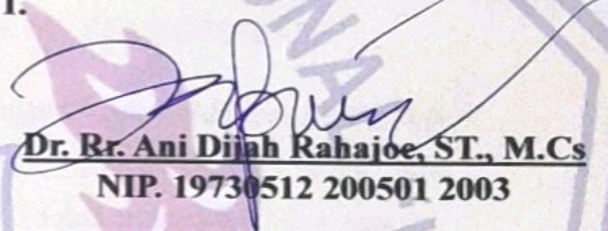
1.



Eva Yulia Puspaningrum, S.Kom,
M.Kom

NIP. 19890705 2021212 002

1.



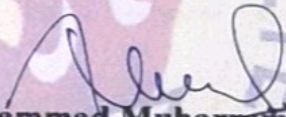
Dr. Rr. Ani Djjah Rahajoe, ST., M.Cs
NIP. 19730512 200501 2003

2.



Chrystia Aji Putra, S.Kom, M.T
NIP. 19861008 2021211 001

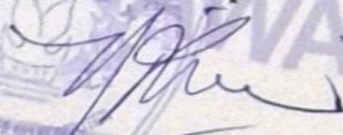
2.



Muhammad Muharrom Al Haromainy,
S.Kom, M.Kom
NIP. 19950601 202203 1006

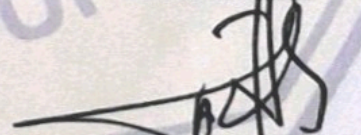
Menyetujui

Dekan
Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, M.T.
NIP. 19681126 199403 2 001

Koordinator Program Studi
Informatika



Fetty Tri Anggraeni, S.Kom, M.Kom
NIP. 19820211 2021212 005

SURAT PERNYATAAN ANTI PLAGIAT

Saya mahasiswa program studi Informatika UPN “Veteran” Jawa Timur, yang bertanda tangan dibawah ini:

Nama : VERONIKA PASKALIA SINAMBELA

NPM : 19081010083

Menyatakan bahwa Judul Skripsi/Tugas Akhir yang saya ajukan dan akan dikerjakan, yang berjudul:

“PERBANDINGAN AKURASI METODE NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE (SVM) PADA ANALISIS SENTIMEN ULASAN APLIKASI MYTELKOMSEL”

Bukan merupakan plagiat dari Skripsi/Tugas Akhir/Penelitian orang lain juga bukan merupakan produk dan atau software yang dibeli dari pihak lain. Saya juga menyatakan Skripsi/Tugas Akhir ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN “Veteran” Jawa Timur maupun institusi Pendidikan lain.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka saya siap menerima segala konsekuensinya.

Surabaya, 22 Januari 2024

Hormat Saya,

A red digital stamp featuring a signature in cursive script. Below the signature, the text "VERONIKA PASKALIA SINAMBELA" and "NPM. 19081010083" is printed. To the left of the signature, there is a vertical stamp with the text "KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN" and "RIPRAK" (RIPRAK is partially visible). At the bottom of the stamp, the alphanumeric code "C8DAJX559329805" is visible.

VERONIKA PASKALIA SINAMBELA
NPM. 19081010083

PERBANDINGAN AKURASI METODE *NAIVE BAYES CLASSIFIER* DAN *SUPPORT VECTOR MACHINE (SVM)* PADA ANALISIS SENTIMEN ULASAN APLIKASI MYTELKOMSEL

Nama Mahasiswa : Veronika Paskalia Sinambela
NPM : 19081010083
Program Studi : Informatika
Dosen Pembimbing : Eva Yulia Puspaningrum, S.Kom, M.Kom.
Chrystia Aji Putra, S.Kom, M.T.

Abstrak

MyTelkomsel merupakan suatu aplikasi yang dirancang untuk membantu pengguna *provider* Telkomsel dalam melakukan aktivitas *provider* seperti, mengecek pulsa, mengecek kuota, membeli kuota, memeriksa tanggal aktif kuota dan masa aktif kartu SIM, dan beberapa fitur layanan Telkomsel lainnya. Bagi calon pengguna aplikasi MyTelkomsel dapat melihat ulasan yang terdapat pada Google Playstore dan pengalaman pengguna lain selama menggunakan aplikasi tersebut. Selain itu, ulasan-ulasan tersebut juga dapat dijadikan sebagai evaluasi bagi developer Telkomsel untuk memperbaiki aspek-aspek yang masih kurang sesuai dengan ulasan. Algoritma Support Vector Machine (SVM) dan Naive Bayes Classifier merupakan dua algoritma yang dapat digunakan untuk mengklasifikasikan ulasan tersebut. Kedua algoritma ini adalah dua algoritma yang populer digunakan dalam mengklasifikasikan data dalam bentuk teks.

Penelitian ini melakukan perbandingan antara algoritma SVM dan Naive Bayes Classifier pada ulasan Aplikasi MyTelkomsel. Data yang diambil merupakan data ulasan selama satu tahun yakni dari 08 Juni 2022 hingga 07 Juni 2023 sebanyak 6019 ulasan berbahasa Indonesia. Pada analisis sentimen ini, sentimen diklasifikasikan menjadi tiga kelas, positif, negatif, dan netral. Pelabelan dilakukan dengan leksikon *Sentistrength_id*. Ekstraksi fitur dilakukan dengan word embedding TF-IDF. Metode yang digunakan adalah SVM dan *Naive Bayes*

Classifier, yang kemudian akan dibandingkan hasilnya. Dari 6019 data yang didapat data sentimen negatif ada sebanyak 2699 data, netral sebanyak 1870 data, dan positif sebanyak 1450 data. Setelah kedua algoritma diterapkan pada tiga skenario akurasi SVM lebih unggul dengan rata-rata 76% sementara Naive Bayes hanya mencapai 61%.

Kata kunci: Analisis sentimen, SVM, Naive Bayes, TF-IDF, Sentistrength, MyTelkomsel

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Kuasa, karena atas berkat dan kasih karunia-Nya, penulis dapat menyelesaikan penyusunan skripsi ini dengan judul "Perbandingan Akurasi Metode *Naive Bayes Classifier* Dan *Support Vector Machine* (SVM) Pada Analisis Sentimen Ulasan Aplikasi MyTelkomsel".

Penulisan skripsi ini sebagai bentuk dari salah satu syarat memperoleh gelar Sarjana Komputer di Program Studi Informatika, Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur. Penulis menyadari bahwa masih banyak kekurangan dalam penulisan skripsi ini. Oleh karena itu, penulis mengharapkan kritik dan saran dari pembaca yang bersifat membangun demi penyempurnaan laporan skripsi ini.

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan bantuan, dukungan, dan kesabaran selama proses penulisan skripsi ini. Penulis berharap semoga skripsi ini dapat bermanfaat baik bagi pembaca maupun penelitian selanjutnya.

Surabaya, 22 Januari 2024

Penulis

Veronika Paskalia Sinambela

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya bagi seluruh pihak yang telah berpartisipasi dan terlibat dalam membantu penulis sehingga penulis dapat menyelesaikan penelitian dan penyusunan tugas akhir/skripsi ini. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Tuhan Yesus Kristus, karena hanya oleh kasih karunia, penyertaan, dan izin-Nya penulis dapat menyelesaikan menyusun skripsi ini.
2. Kedua Orang Tua penulis, Bunda Theresia Simbolon dan Bapak Roy Tono Sinambela yang dengan penuh kasih sayang dan dukungan dalam penyusunan tugas akhir/skripsi. Terima kasih atas dedikasi tanpa pamrih, kesabaran, dan cinta yang luar biasa yang telah diberikan kepada penulis.
3. Prof. Dr. Ir. Akhmad Fauzi, MMT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Ibu Fetty Tri Anggraeny, S.Kom., M. Kom. selaku Koordinator Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
6. Ibu Eva Yulia Puspaningrum, S.Kom, M.Kom. selaku dosen pembimbing pertama dan Bapak Chrystia Aji Putra, S.Kom, M.T. selaku dosen pembimbing kedua yang telah meluangkan waktu dan tenaga dalam membimbing, mengarahkan dan memberi pencerahan kepada penulis selama penyusunan tugas akhir/skripsi.
7. Kepada kedua kakak kandung penulis yang selalu memberikan perhatian dan dukungan kepada penulis agar selalu semangat dan termotivasi dalam menyelesaikan perkuliahan.
8. Teman-teman penulis di prodi Informatika Angkatan 2019 yang telah membantu penulis serta memberikan semangat, wawasan, motivasi, perhatian dan pencerahan dapat menyelesaikan penyusunan tugas

akhir/skripsi ini juga berbagi cerita suka dan duka selama masa perkuliahan.

9. Sahabat-sahabat penulis yang telah menjadi tempat yang nyaman untuk pulang dari kehidupan perkuliahan, mencairkan stres dan memotivasi penulis untuk tetap semangat dan bertahan selama kuliah.
10. Dan kepada diri sendiri yang telah bertahan dan berjuang selama perkuliahan. Terima kasih karena tidak menyerah.

Penulis berharap semoga Tuhan Yang Maha Kuasa membalas seluruh kebaikan yang telah diberikan selama proses penyusunan tugas akhir/skripsi. Semoga tugas akhir/skripsi ini dapat memberikan manfaat bagi pembaca ataupun penulis sendiri.

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
SURAT PERNYATAAN ANTI PLAGIAT	ii
Abstrak.....	iii
KATA PENGANTAR	v
UCAPAN TERIMA KASIH	vi
DAFTAR ISI	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DAFTAR KODE PROGRAM	xiii
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
BAB II	5
TINJAUAN PUSTAKA	5
2.1 Penelitian Pendahulu	5
2.2 Aplikasi MyTelkomsel.....	7
2.3 Text Mining.....	7
2.4 Analisis Sentimen.....	8
2.5 CRISP-DM.....	9
2.6 Text Preprocessing	12
2.6.1 Cleaning	12
2.6.2 Case Folding.....	12
2.6.3 Tokenizing.....	12
2.6.4 Stopwords Removal	12

2.7	Pelabelan Sentristrength_id.....	13
2.8	TF-IDF	13
2.9	Support Vector Machine (SVM).....	14
2.10	Naive Bayes	16
2.11	Confusion Matrix	18
2.12	Visualisasi	19
BAB III.....		20
METODE PENELITIAN		20
3.1	Tahap Penelitian.....	20
3.2	Business Understanding	21
3.3	Data Understanding.....	21
3.4	Data Preparation.....	23
3.4.1.	Case Folding	24
3.4.2.	Cleaning	25
3.4.3.	Filter Bahasa	25
3.4.4.	Stopword Removal	26
3.4.5.	Tokenizing	27
3.4.6.	Pelabelan Data	27
3.4.7.	TF-IDF (<i>Term Frequency-Inverse Dense Frequency</i>)	28
3.4.8.	Split Data	31
3.5	Modeling	32
3.6	Evaluation	40
3.7	Deployment.....	41
BAB IV.....		43
HASIL DAN PEMBAHASAN		43
4.1	Pengambilan Data	43
4.2	Preprocessing dan Filtering Data	44
4.3	Pelabelan Data.....	48
4.4	Split Data.....	51
4.5	Word Embedding (TF-IDF)	51
4.6	Mengklasifikasi dengan Metode SVM.....	52

4.7	Mengklasifikasikan dengan Metode Naive Bayes	53
4.8	Hasil dan Pembahasan Analisis Perbandingan Uji Coba Menggunakan SVM dan Naive Bayes.....	54
4.8.1	Hasil Pengujian pada Skenario 1	55
4.8.2	Hasil Pengujian pada Skenario 2	57
4.8.3	Hasil Pengujian pada Skenario 3	58
4.8.4	Perbandingan Hasil Akurasi Algoritma SVM dan Naive Bayes..	60
4.8.5	Visualisasi Perbandingan Akurasi Algoritma SVM dan Naive Bayes	61
4.9	Uji Coba Analisis Sentimen Menggunakan Data Baru.....	62
4.10	Desain Antarmuka Pengembangan Aplikasi.....	63
BAB V		65
PENUTUP		65
5.1	Kesimpulan	65
5.2	Saran.....	65
DAFTAR PUSTAKA.....		66

DAFTAR GAMBAR

<i>Gambar 3. 1 Tahap penelitian</i>	20
<i>Gambar 3. 2 Data ulasan pada google playstore</i>	22
<i>Gambar 3. 3 Hasil scraping data</i>	22
<i>Gambar 3. 4 Preprocessing</i>	24
<i>Gambar 3. 5 Bagan pengklasifikasian</i>	32
<i>Gambar 3. 6 Desain antarmuka analisis sentimen (1)</i>	41
<i>Gambar 3. 7 Desain antarmuka analisis sentimen (2)</i>	42
<i>Gambar 3. 8 Desain antarmuka analisis sentimen (3)</i>	42
<i>Gambar 4. 1 Hasil case folding</i>	45
<i>Gambar 4. 2 Data kosong setelah proses cleaning</i>	47
<i>Gambar 4. 3 Beberapa leksikon Sentistrent_id dan nilainya</i>	49
<i>Gambar 4. 4 Contoh pelabelan satu kalimat sentimen dengan Sentistrength_id</i> 49	
<i>Gambar 4. 5 Data yang telah diberi label</i>	50
<i>Gambar 4. 6 Visualisasi jumlah data tiap label</i>	50
<i>Gambar 4. 7 Hasil ekstraksi fitur TF-IDF</i>	52
<i>Gambar 4. 8 Hasil pengklasifikasian SVM</i>	53
<i>Gambar 4. 9 Pemodelan Naive Bayes</i>	53
<i>Gambar 4. 10 Classification report SVM skenario 1</i>	56
<i>Gambar 4. 11 Classification report Naive Bayes skenario 1</i>	56
<i>Gambar 4. 12 Classification report SVM skenario 2</i>	57
<i>Gambar 4. 13 Classification report Naive Bayes skenario 2</i>	58
<i>Gambar 4. 14 Classification report SVM skenario 3</i>	59
<i>Gambar 4. 15 Classification report Naive Bayes skenario 3</i>	59
<i>Gambar 4. 16 Diagram perbandingan akurasi SVM dan Naive Bayes</i>	61
<i>Gambar 4. 17 Wordcloud sentimen aplikasi MyTelkomsel</i>	62
<i>Gambar 4. 18 Memasukkan input kalimat baru</i>	62
<i>Gambar 4. 19 Mengklasifikasikan data baru</i>	62
<i>Gambar 4. 20 Tampilan UI pengembangan aplikasi yang menampilkan diagram</i>	63
<i>Gambar 4. 21 Tampilan UI Aplikasi yan menampilkan tabel jumla data setiap label dan wordcloud</i>	64
<i>Gambar 4. 22 Tampilan UI aplikasi yang menampilkan hasil klasifikasi SVM dan Naive Bayes setiap data sentimen</i>	64

DAFTAR TABEL

<i>Tabel 2. 1 Matrix model Confusion Matrix</i>	18
<i>Tabel 2. 2 Contoh Hasil Scraping data Sentimen</i>	22
<i>Tabel 3. 1 Contoh Cleaning</i>	24
<i>Tabel 3. 2 Contoh Case Folding</i>	25
<i>Tabel 3. 3 Filter bahasa</i>	25
<i>Tabel 3. 4 Contoh Stopword removal</i>	26
<i>Tabel 3. 5 Contoh tokenizing</i>	27
<i>Tabel 3. 6 Contoh korpus</i>	28
<i>Tabel 3. 7 Perhitungan TF</i>	28
<i>Tabel 3. 8 Hasil Perhitungan DF</i>	29
<i>Tabel 3. 9 Tabel Perhitungan IDF</i>	30
<i>Tabel 3. 10 Tabel Perhitungan TF-IDF</i>	30
<i>Tabel 3. 11 tabel perhitungan TF-IDF Normalisasi</i>	31
<i>Tabel 3. 12 Contoh data latih dan data uji</i>	32
<i>Tabel 3. 13 Ekstraksi fitur TF-IDF terhadap data latih dan data uji</i>	33
<i>Tabel 3. 14 Menetapkan nilai y pada setiap model pelatihan</i>	33
<i>Tabel 3. 15 Tabel perhitungan matriks kernel</i>	35
<i>Tabel 3. 16 Tabel perhitungan matrix hessian</i>	35
<i>Tabel 3. 17 Hasil Perhitungan $E1$</i>	36
<i>Tabel 3. 18 Hasil perhitungan $\delta\alpha1$</i>	36
<i>Tabel 3. 19 Hasil perhitungan $\alpha1$</i>	36
<i>Tabel 3. 20 Hasil perhitungan Ei</i>	37
<i>Tabel 3. 21 Hasil perhitungan $\delta\alpha i$</i>	37
<i>Tabel 3. 22 Hasil Perhitungan αi</i>	37
<i>Tabel 3. 23 Hasil Perhitungan Kernel Data Latih terhadap Data Uji</i>	38
<i>Tabel 3. 24 Tabel hasil fungsi diskriminan</i>	38
<i>Tabel 3. 25 TF setiap Term pada data latih dan data uji</i>	38
<i>Tabel 3. 26 Peluang setiap term terhadap setiap kelas</i>	39
<i>Tabel 4. 1 Hasil scraping data</i>	44
<i>Tabel 4. 2 Hasil case folding</i>	45
<i>Tabel 4. 3 Hasil proses cleaning</i>	46
<i>Tabel 4. 4 Hasil deteksi bahasa</i>	47
<i>Tabel 4. 5 Data yang telah dilabeli</i>	50
<i>Tabel 4. 6 Pembagian data setiap skenario</i>	51
<i>Tabel 4. 7 Tabel confusion matrix skenario 1</i>	55
<i>Tabel 4. 8 Tabel confusion matrix skenario 2</i>	57
<i>Tabel 4. 9 Tabel confusion matrix skenario 3</i>	58

DAFTAR KODE PROGRAM

<i>Kode Program 4. 1 Scraping data</i>	43
<i>Kode Program 4. 2 Kode program case folding.....</i>	45
<i>Kode Program 4. 3 Kode program proses cleaning.....</i>	46
<i>Kode Program 4. 4 Kode program menghapus data kosong</i>	47
<i>Kode Program 4. 5 Kode program deteksi bahasa</i>	47
<i>Kode Program 4. 6 Mengambil data berbahasa Indonesia</i>	48
<i>Kode Program 4. 7 stopword removal</i>	48
<i>Kode Program 4. 8 Melabeli dataset sentimen</i>	49
<i>Kode Program 4. 9 Split data skenario 3</i>	51
<i>Kode Program 4. 10 Word embedding TF-IDF</i>	52
<i>Kode Program 4. 11 Pemodelan SVM.....</i>	52
<i>Kode Program 4. 12 Pengujian SVM terhadap data uji</i>	53
<i>Kode Program 4. 13 Pengujian Naive Bayes pada data uji.....</i>	54
<i>Kode Program 4. 14 Menampilkan tabel confusion matrix</i>	54
<i>Kode Program 4. 15 Menampilkan classification report</i>	55