

**PEMBANGUNAN *CLASSIFIER* UNTUK KLASIFIKASI
JUDUL BERITA BERDASARKAN SENTIMEN**



**OLEH
NI LUH PUTU RISMA DEWI
NIM 1915101008**

**PROGRAM STUDI ILMU KOMPUTER
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA**

2023



**PEMBANGUNAN *CLASSIFIER* UNTUK KLASIFIKASI
JUDUL BERITA BERDASARKAN SENTIMEN**

SKRIPSI

Diajukan Kepada

Universitas Pendidikan Ganesha

Untuk memenuhi salah satu persyaratan Dalam Menyelesaikan

Program Sarjana Ilmu Komputer



Oleh

NI LUH PUTU RISMA DEWI

NIM 1915101008

**PROGRAM SRUDI ILMU KOMPUTER
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA**

2023

SKRIPSI

**DIAJUKAN UNTUK MELENGKAPI TUGAS
DAN MEMENUHI SYARAT-SYARAT UNTUK
MENCAPAI GELAR SARJANA KOMPUTER**

Menyetujui

Pembimbing I,



I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs.
NIP. 198910262019031004

Pembimbing II,



I Ketut Purnamawan, S.Kom., M.Kom.
NIP.197905112006041004

Skripsi oleh Ni Luh Putu Risma Dewi ini
telah di pertahankan di depan dewan penguji
pada tanggal 22 Juni 2023

Dewan Penguji,



I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs.
NIP. 198910262019031004

(Ketua)



I Ketut Purnamawan, S.Kom., M.Kom.
NIP. 197905112006041004

(Anggota)



Ni Wayan Marti, S.Kom., M.Kom.
NIP. 197711282001122001

(Anggota)



Ni Putu Novita Puspa Dewi, S.Kom., M.Cs.
NIP. 199410032020122015

(Anggota)

Diterima oleh Panitia Ujian Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha
Guna memenuhi syarat-syarat untuk mencapai gelar sarjana

Pada:

Hari : Kamis.....

Tanggal : 22 Juni 2023

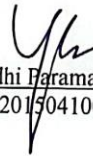
Mengetahui,

Ketua Ujian,



Prof. Dr. Ketut Agustini, S.Si., M.Si.
NIP. 197408012000032001

Sekretaris Ujian,



A. A. Gede Yudhi Paramartha, S.Kom., M.Kom.
NIP. 198806222015041003

Mengesahkan,

Dekan Fakultas Teknik dan Kejuruan



Prof. Dr. I Gede Sudirtha, S.Pd., M.Pd.
NIP. 197106161996021001

PERNYATAAN

Dengan ini saya menyatakan bahwa karya tulis yang berjudul "**Pembangunan Classifier Untuk Klasifikasi Judul Berita Berdasarkan Sentimen**" beserta keseluruhan isinya adalah benar-benar hasil karya sendiri dan tidak melakukan penjiplakan dan pengutipan dengan cara yang tidak sesuai dengan etika yang berlaku pada buku pendoman skripsi. Jika dikemudian hari ditemukan adanya pelanggaran saya siap menanggung resiko atau sanksi yang di jatuhkan kepada saya.

Singaraja, 22 Juni 2023

Yang membuat pernyataan,



Ni Luh Putu Risma Dewi

NIM. 1915101008

MOTTO

"If you want to change things differently, start with yourself."



PRAKATA

Puji syukur peneliti panjatkan ke hadapan Ida Sang Hyang Widhi Wasa/Tuhan Yang Maha Esa karena berkat rahmat-Nya lah sehingga memungkinkan peneliti menyelesaikan skripsi dengan judul "**Pembangunan Classifier Untuk Klasifikasi Judul Berita Berdasarkan Sentimen**" sebagai salah satu syarat untuk memperoleh gelar sarjana Ilmu Komputer di Universitas Pendidikan Ganesha. Dalam menyelesaikan skripsi ini, peneliti banyak mendapatkan motivasi dan bantuan baik berupa moral maupun material dari berbagai pihak. Untuk itu dalam kesempatan ini peneliti mengucapkan rasa terima kasih kepada :

1. Bapak Prof.Dr.I Gede Sudirtha, M.Pd., selaku Dekan Fakultas Teknik dan Kejuruan atas motivasi dan fasilitas yang diberikan sehingga peneliti dapat menyelesaikan studi sesuai dengan rencana.
2. Ibu Dr. Luh Joni Erawati Dewi, ST., M.Pd., selaku Ketua Jurusan Teknik Informatika atas motivasi dan fasilitas yang diberikan sehingga peneliti dapat menyelesaikan studi ini dengan tepat waktu.
3. Bapak A.A. Gede Yudhi Paramartha, S.Kom., M.Kom., selaku Koorprodi Ilmu Komputer atas motivasi dan fasilitas yang diberikan kepada peneliti dalam penyelesaian skripsi ini.
4. Bapak I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs., selaku pembimbing I yang telah memberikan banyak motivasi, bimbingan, arahan, petunjuk skripsi, dengan penuh kesabaran kepada peneliti dalam menyelesaikan skripsi ini sehingga peneliti dapat menyelesaikan skripsi dengan tepat waktu.
5. Bapak I Ketut Purnamawan, S.Kom., M.Kom., selaku pembimbing II yang telah banyak memberikan motivasi, semangat, bimbingan, arahan dengan penuh kesabaran kepada peneliti dalam menyelesaikan skripsi ini sehingga peneliti dapat menyelesaikan skripsi dengan tepat waktu.
6. Seluruh staff dosen di lingkungan Jurusan Teknik Informatika/ Program Studi Ilmu Komputer yang telah memberikan ilmu, pengalaman, motivasi, serta semangat selama menjalani studi di Universitas Pendidikan Ganesha

7. Bapak I Putu Edy Suardiyana Putra, PH.D., selaku CEO Bit House yang telah membantu memberikan ide kepada peneliti dalam menyelesaikan skripsi ini sehingga peneliti dapat menyelesaikan skripsi dengan tepat waktu.
8. Ungkapan terimakasih dan penghargaan yang sangat spesial peneliti haturkan kepada kedua orang tua tercinta peneliti (Bpk. I Made Sujana dan Ibu Rosalina Sainah) serta kedua saudara peneliti, yang selalu memberikan doa restu, nasihat, dukungan, semangat, bimbingan dan motivasi serta cinta kasih kepada peneliti, sehingga skripsi ini dapat selesai dengan tepat waktu.
9. Untuk teman seperjuangan peneliti Subiksa, Prima, Gita dan teman-teman lainnya yang tidak bisa peneliti sebutkan satu persatu, terima kasih banyak telah memberikan semangat, dukungan, bantuan sehingga peneliti sampai di titik ini dan dapat menyelesaikan skripsi dengan tepat waktu.
10. Untuk Ayu Suartini, Dewi Saras, Manika dan Ary Putra yang telah selalu memberikan semangat dan menghibur peneliti dalam penyusunan skripsi ini sehingga peneliti dapat menyelesaikan skripsi dengan tepat waktu.
11. Untuk seluruh teman – teman prodi Ilmu Komputer Angkatan 2019 yang telah memberikan berbagai pengalaman dan kebersamaan selama peneliti menempuh pendidikan di Universitas Pendidikan Ganesha.
12. Seluruh pihak yang tidak bisa di sebutkan satu persatu yang telah membantu peneliti dalam penyusunan skripsi sehingga dapat selesai tepat waktu.

Semoga Ida Sang Hyang Widhi Wasa/Tuhan Yang Maha Esa senantiasa memberikan karunianya dan membantu peneliti dalam menyelesaikan skripsi ini. Peneliti menyadari sepenuhnya bahwa apa yang disajikan dalam skripsi ini masih jauh dari kata sempurna mengingat keterbatasan peneliti. Peneliti mengharapkan segala masukan, kritik dan saran yang bersifat membangun dari berbagai pihak. Peneliti berharap apa yang dibuat oleh peneliti dapat bermanfaat dan berharga bagi kita semua, khususnya untuk mengembangkan dunia pendidikan

Singaraja, 22 Juni 2023
Peneliti

**PEMBANGUNAN *CLASSIFIER* UNTUK KLASIFIKASI
JUDUL BERITA BERDASARKAN SETIMEN**

Oleh

Ni Luh Putu Risma Dewi, NIM 1915101008

Jurusan Teknik Informatika

ABSTRAK

Perkembangan teknologi dan *platform* digital telah memberikan dampak yang signifikan pada industri pariwisata, termasuk dalam penyebaran berita pariwisata. Artikel-artikel berita yang banyak dipublikasikan oleh portal media online dapat memiliki dampak positif atau negatif yang besar bagi masyarakat pembaca. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sebuah *classifier* yang dapat melakukan klasifikasi terhadap judul berita pariwisata berdasarkan sentimen. Tujuan utamanya adalah menganalisis dan mengklasifikasikan sentimen atau sikap emosional yang terkandung dalam judul berita. Metode *machine learning* yang digunakan adalah *Support Vector Machine* (SVM). Data penelitian dikumpulkan dari portal berita detik.com dan dilabeli secara manual sesuai dengan sentimen yang terkandung di dalamnya. Proses *preprocessing* teks dilakukan untuk mempersiapkan data judul berita pariwisata Indonesia agar dapat diekstraksi fiturnya menggunakan pendekatan *Binary Term Presence* dan *Term Frequency*. Proporsi pembagian dataset dalam penelitian ini adalah 90% untuk pelatihan (*training process*) dan 10% untuk pengujian. *K-Fold Cross Validation* digunakan untuk membagi data dalam proses pelatihan. Hasil penelitian menunjukkan bahwa pendekatan *Binary Term Presence* mencapai akurasi 87,80% dan pendekatan *Term Frequency* mencapai akurasi 85,95% dengan menggunakan kernel RBF. Penelitian ini memberikan kontribusi dalam pemahaman informasi yang mendukung tentang respon publik terhadap topik atau isu pariwisata yang di publikasikan oleh media.

Kata Kunci : Pariwisata, Sentimen , Klasifikasi, SVM , Detik.com

**CLASSIFIER DEVELOPMENT FOR NEWS TITLE CLASSIFICATION
BASED ON SENTIMENT**

By

Ni Luh Putu Risma Dewi, NIM 1915101008

Computer Science Major

ABSTRACT

The development of technology and digital platforms has had a significant impact on the tourism industry, including the dissemination of tourism-related news. The numerous news articles published by online media portals can have a profound positive or negative impact on readers. Therefore, this research aims to develop a classifier capable of sentiment classification for tourism news headlines. The main objective is to analyze and classify the sentiment or emotional stance conveyed in the news titles. The machine learning method employed is Support Vector Machine (SVM). Research data is collected from the detik.com news portal and manually labeled according to the sentiments expressed. Text preprocessing is performed to prepare the Indonesian tourism news headline data for feature extraction, using the Binary Term Presence and Term Frequency approaches. The dataset is divided with a proportion of 90% for the training process and 10% for testing. K-Fold Cross Validation is employed for data division during the training process. The research results indicate that the Binary Term Presence approach achieves an accuracy of 87.80%, while the Term Frequency approach achieves an accuracy of 85.95% using the kernel RBF. This research contributes to understanding information supporting the public response to tourism topics or issues published by the media.

Keywords: *Tourism, Sentiment, Classification, SVM, Detik.com*

DAFTAR ISI

PRAKATA	i
ABSTRAK	iii
DAFTAR ISI	v
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Batasan Masalah	5
1.5 Manfaat Penelitian	6
BAB II LANDASAN TEORI	7
2.1 Penelitian Terkait	7
2.2 <i>Web Scraping</i>	11
2.3 <i>Text Mining</i>	12
2.4 Klasifikasi Teks	13
2.5 Analisis Sentimen	13
2.6 <i>Text Preprocessing</i>	14
2.7 <i>Binary Term Presence</i>	17
2.8 <i>Term Frequency</i>	17
2.9 <i>CountVectorizer</i>	18
2.10 N – Gram	18
2.11 <i>WordCloud</i>	19
2.12 <i>Training set, Validation set dan Testing set</i>	20
2.13 <i>Support Vector Machine</i>	21
2.14 <i>K-fold Cross Validation</i>	27
2.15 <i>Confusion Matrix</i>	28

BAB III METODOLOGI PENELITIAN	31
3.1 Kerangka Penelitian.....	31
3.2 Studi Literatur.....	32
3.3 Pengambilan Data.....	32
3.4 Pelabelan Data.....	32
3.5 Text <i>Preprocessing</i>	34
3.6 Ekstraksi Fitur	36
3.7 Pembagian Data.....	38
3.8 Klasifikasi <i>Support vector machine (SVM)</i>	39
3.9 Visualisasi Data.....	40
3.10 Pengujian.....	40
BAB IV HASIL DAN PEMBAHASAN	41
4.1 Pengambilan Data.....	41
4.2 Pelabelan dan Pembagian Data	42
4.3 Text <i>Preprocessing</i>	44
4.3.1 <i>Cleaning</i>	45
4.3.2 <i>Case Folding</i>	46
4.3.3 <i>Tokenizing</i>	47
4.3.4 <i>Stopword</i>	49
4.3.5 <i>Stemming</i>	50
4.4 Ekstraksi Fitur	51
4.4.1 <i>Binary Term Presence</i>	51
4.4.2 <i>Term Frequency</i>	52
4.5 Klasifikasi Sentimen.....	54
4.5.1 <i>K-Fold Cross Validation (training dan validation)</i>	54
4.5.2 <i>Support Vector Machine Model</i>	56
4.5.2.1 SVM dengan <i>Binary Term Presence</i>	58
4.5.2.2 SVM dengan <i>Term Frequency</i>	59
4.5.3 <i>Confusion Matrix</i> dari Model Terbaik	61
4.6 Visualisasi	66

4.6.1 Sentimen Positif.....	67
4.6.2 Sentimen Netral	69
4.6.3 Sentimen Negatif	72
4.7 Analisis Perbandingan Hasil Ekstraksi Fitur <i>Binary Term Presence</i> dan <i>Term Frequency</i>	74
BAB V PENUTUP	76
5.1 Kesimpulan.....	76
5.2 Saran	77
DAFTAR PUSTAKA.....	78
LAMPIRAN.....	83
RIWAYAT HIDUP.....	90



DAFTAR GAMBAR

Gambar 2.1 Alur <i>Web Scrapping</i>	11
Gambar 2.2 Alur <i>Preprocessing</i>	15
Gambar 2.3 Visualisasi Word Cloud	20
Gambar 2.4 <i>Hyperplane</i> SVM memisahkan class positif dan negatif.....	23
Gambar 2.5 Pemetaan Input Space Berdimensi Dua dengan Pemetaan ke Dimensi Tinggi.....	23
Gambar 2.6 Struktur One-against-all.....	26
Gambar 2.7 Struktur one-against-one	26
Gambar 2.8 Struktur Directed Acyclic Graph Support Vector Machine.....	27
Gambar 2.9 Contoh <i>K-Fold Cross Validation</i>	28
Gambar 3.1 Kerangka Penelitian.....	31
Gambar 3.2 Flowchart Proses <i>Preprocessing</i>	34
Gambar 3.3 Flowchart Klasifikasi SVM.....	39
Gambar 4.1 Halaman judul berita detik.com	41
Gambar 4.2 Hasil Presentase Distribusi Sentimen	44
Gambar 4.3 Upload Data excel.....	45
Gambar 4.4 Code menampilkan data.....	45
Gambar 4.5 Algoritma <i>Binary Term Presence</i>	52
Gambar 4.6 Hasil Vector <i>Binary Term Presence</i>	52
Gambar 4.7 Algoritma <i>Term Frequency</i>	53
Gambar 4.8 <i>Confusion Matrix</i> SVM dengan <i>Binary Term Presence</i>	62
Gambar 4.9 Hasil Barplot Sentimen Positif	68
Gambar 4.10 Hasil <i>WordCloud</i> Sentimen Positif.....	68
Gambar 4.11 Hasil Barplot Sentimen Netral.....	70
Gambar 4.12 Hasil <i>WordCloud</i> Sentimen Netral	70
Gambar 4.13 Hasil Barplot Sentimen Negatif.....	72
Gambar 4.14 Hasil <i>WordCloud</i> Sentimen Negatif	73

DAFTAR TABEL

Tabel 2.1 Jenis-jenis kernel SVM.....	25
Tabel 2.2 Confusion Matrix	29
Tabel 3.1 Kriteria Pelabelan Data.....	33
Tabel 3.2 Contoh Hasil Pelabelan Data	33
Tabel 3.3 Contoh kalimat dokumen CountVectorizer	36
Tabel 3.4 Daftar kata dengan nomer indeks.....	36
Tabel 3.5 Frekuensi Term Dokumen 1	37
Tabel 3.6 Frekuensi Term dokumen 2	37
Tabel 3.7 Hasil pembobotan CountVectorizer manual.....	38
Tabel 4. 1 Data Judul berita januari - desember 2022	42
Tabel 4. 2 Hasil Pelabelan Data.....	43
Tabel 4. 3 Perbandingan Jumlah Data Pada Kelas Sentimen.....	43
Tabel 4.4 Hasil Cleaning Data.....	46
Tabel 4.5 Hasil Casefolding Data.....	47
Tabel 4.6 Hasil Tokenizing Data.....	48
Tabel 4 7 Hasil Stopword Data	49
<i>Tabel 4.8 Hasil Stemming Data.....</i>	<i>50</i>
Tabel 4.9 Percobaan K-Fold Cross Validation.....	55
Tabel 4.10 Hasil Training dan Testing SVM with Binary Term Presence	58
Tabel 4.11 Hasil Training dan Testing SVM dengan Term Frequency.....	59
Tabel 4.12 Hasil Akurasi 2 Label	61
Tabel 4. 13 Perbandingan Hasil Model.....	61
Tabel 4. 14 Confusion Matrix dengan Binary Term Presence	62
Tabel 4. 15 Confusion Matrix Kelas Positif.....	63
Tabel 4.16 Confusion Matrix Kelas Netral	63
Tabel 4. 17 Confusion Matrix Kelas Negatif	63
Tabel 4.18 Misclassification Testing Data.....	64

DAFTAR LAMPIRAN

Lampiran 1. Lampiran 1 Data Mahasiswa Pelabel Data.....	84
Lampiran 2. <i>Source Code Program</i>	85

