

**PERBANDINGAN ANALISIS SENTIMEN JUDUL
BERITA PARIWISATA BALI MENGGUNAKAN SVM
(*SUPPORT VECTOR MACHINE*)**



OLEH

I PUTU SOMA DARMAYASA

NIM 1915101017

PROGRAM STUDI ILMU KOMPUTER

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK DAN KEJURUAN

UNIVERSITAS PENDIDIKAN GANESHA

SINGARAJA

2023

**PERBANDINGAN ANALISIS SENTIMEN JUDUL
BERITA PARIWISATA BALI MENGGUNAKAN SVM
(*SUPPORT VECTOR MACHINE*)**

SKRIPSI



PROGRAM STUDI ILMU KOMPUTER
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA
2023

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

Menyetujui,

Pembimbing I,



I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs.
NIP. 19891026 201903 1 004

Pembimbing II,



I Ketut Resika Arthana, S.T., M.Kom.
NIP. 19841201 201212 1 002


Skripsi oleh I Putu Soma Darmayasa ini
telah di pertahankan di depan dewan penguji
pada tanggal 18 September 2023

Dewan Penguji,



I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs.
NIP. 19891026 201903 1 004

(Ketua)



I Ketut Resika Arthana, S.T., M.Kom.
NIP. 19841201 201212 1 002

(Anggota)



Dr. Ni Ketut Kertiasih, S.Si., M.Pd.
NIP. 19701118 199703 2 001

(Anggota)



Agus Aan Jiwa Permana, S.Kom., M.Cs.
NIP. 19870804 201504 1 001

(Anggota)

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

Pada:

Hari : Senin

Tanggal : 18 September 2023

Mengetahui,

Ketua Ujian,



Made Windu Antara Kesiman, S.T., M.Sc., Ph.D.
NIP. 19821111 200812 1 001

Sekretaris Ujian,



I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs.
NIP. 19891026 201903 1 004

Mengesahkan,

Dekan Fakultas Teknik dan Kejuruan




Dr. Adek Rihendra Dantes, S.T., M.T.
NIP. 19791201 200604 1 001

PERNYATAAN

Dengan ini, saya menyatakan bahwa karya tulis yang berjudul "**Perbandingan Analisis Sentimen Judul Berita Pariwisata Bali Menggunakan SVM (*Support Vector Machine*)**" beserta seluruh isinya adalah hasil karya pribadi saya sendiri, dan tidak melibatkan penjiplakan atau pengutipan yang melanggar etika dan aturan yang berlaku dalam penulisan skripsi. Jika di masa depan terungkap adanya pelanggaran, saya bersedia menerima konsekuensi atau sanksi yang diberlakukan kepada saya..

Singaraja, 18 September 2023

Yang membuat pernyataan,



I Putu Soma Darmayasa

NIM. 1915101017

MOTTO

“Kerjakan Saja”



PRAKATA

Puji syukur peneliti panjatkan kehadapan Ida Sang Hyang Widhi Wasa/Tuhan Yang Maha Esa karena berkat rahmat-Nya lah sehingga memungkinkan peneliti menyelesaikan skripsi dengan judul "Perbandingan Analisis Sentimen Judul Berita Pariwisata Bali Menggunakan SVM (*Support Vector Machine*)" 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 Wayan Lasmawan, M.Pd., selaku Rektor Universitas Pendidikan Ganesha beserta stafnya yang telah memberikan motivasi dan fasilitas sehingga peneliti dapat menyelesaikan studi dan melakukan penelitian sesuai dengan rencana yang dicanangkan.
2. Bapak Dr. Kadek Rihendra Dantes, S.T., M.T., selaku Dekan Fakultas Teknik dan Kejuruan atas motivasi dan fasilitas yang diberikan sehingga peneliti dapat menyelesaikan studi sesuai dengan rencana.
3. Bapak Dr. Putu Hendra Suputra, S.Kom., M.Cs., selaku Ketua Jurusan Teknik Informatika atas motivasi dan fasilitas yang diberikan sehingga peneliti dapat menyelesaikan studi ini dengan tepat waktu.
4. Bapak I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs., selaku Koorprodi Ilmu Komputer dan 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 Resika Arthana, S.T., 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 staf 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 banyak 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. Made Kertanegara dan Ibu Ketut Puspa) 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 seluruh teman – teman prodi Ilmu Komputer Angkatan 2019 yang telah memberikan berbagai pengalaman dan kebersamaan selama peneliti menempuh pendidikan di Universitas Pendidikan Ganesha.

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, 18 September 2023

Peneliti

DAFTAR ISI

PRAKATA	i
ABSTRAK	Error! Bookmark not defined.
ABSTRACT	Error! Bookmark not defined.
DAFTAR ISI	iii
DAFTAR TABEL	vi
DAFTAR GAMBAR	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Batasan Masalah	5
1.5 Manfaat	6
BAB II KAJIAN TEORI	7
2.1 Penelitian Terkait	7
2.2 Landasan Teori	15
2.2.1 <i>Scraping Website</i>	15
2.2.2 <i>Text Preprocessing</i>	16
2.2.3 <i>Countvectorizer (Feature Extraction)</i>	18
2.2.4 <i>Feature Selection</i>	20
2.2.5 SVM	31
2.2.6 <i>Confusion Matrix</i>	36
2.2.7 <i>K-Fold Cross Validation</i>	37
2.2.8 <i>WorldCloud</i>	38
2.2.9 Software Pendukung	39
2.2.10 Analisis Sentimen	40

2.2.11	Pariwisata	40
2.2.12	Berita	41
2.2.13	Portal Berita Online	41
2.2.14	<i>Training set, Validation set dan Testing set</i>	42
2.2.15	<i>LDA (Latent Dirichlet Allocation)</i>	43
2.2.16	<i>Perplexity</i>	46
2.2.17	<i>Topik Coherence</i>	46
BAB III METODE PENELITIAN		48
3.1	Rancangan Penelitian	48
3.2	Studi Literatur	49
3.3	Subjek Penelitian/Populasi dan Sampel	50
3.3.1	Dataset	50
3.3.2	<i>Scraping Data Website</i>	51
3.3.3	Pelabelan	51
3.4	<i>Text Preprocessing</i>	53
3.4.1	<i>Cleaning</i>	53
3.4.2	<i>Case Folding</i>	54
3.4.3	<i>Tokenizing</i>	54
3.4.4	<i>Stopword removal</i>	54
3.4.5	<i>Stemming</i>	55
3.5	<i>Countvectorizer</i>	55
3.6	<i>Chi-Square dan Information Gain</i>	57
3.7	Pembagian dataset	62
3.8	Klasifikasi SVM	63
3.9	Evaluasi Model	64
3.9.1	<i>Confusion Matrix</i>	64
3.10	Pengujian model	64
3.11	Pemodelan topik menggunakan LDA	65
BAB IV HASIL DAN PEMBAHASAN		66

4.1	Pengambilan Data.....	66
4.2	Pelabelan dan Pembagian Data	68
4.3	<i>Text Preprocessing</i>	70
4.3.1	<i>Cleaning</i>	70
4.3.2	<i>Case Folding</i>	71
4.3.3	<i>Tokenizing</i>	71
4.3.4	<i>Stopword</i>	72
4.3.5	<i>Stemming</i>	73
4.4	Ekstraksi Fitur <i>Countvectorizer</i>	73
4.5	<i>Feature Selection Chi-Square dan Information Gain</i>	74
4.5.1	<i>Chi-Square</i>	74
4.5.2	<i>Information Gain</i>	77
4.6	Klasifikasi Sentimen.....	78
4.6.1	<i>K-Fold Cross Validation</i>	79
4.6.2	<i>Support Vector Machine</i>	81
4.6.3	Evaluasi <i>Confusion Matrix</i> Model Terbaik.....	86
4.7	Analisis Perbandingan Hasil Seleksi Fitur <i>Chi-Square</i> dengan <i>Information Gain</i>	92
4.8	Visualisasi <i>World Cloud</i>	94
4.8.1	Sentimen Judul Positif.....	95
4.8.2	Sentimen Judul Netral.....	96
4.8.3	Sentimen Judul Negatif.....	97
4.9	Hasil pemodelan topik LDA.....	98
4.10	Analisis Hasil dan Pembahasan Penelitian	104
BAB V PENUTUP.....		107
5.1	Kesimpulan.....	107
5.2	Saran.....	109
DAFTAR PUSTAKA		112
LAMPIRAN		148
RIWAYAT HIDUP.....		153

DAFTAR TABEL

Tabel 2.1 Penelitian terkait	14
Tabel 2.2 Penerapan <i>Case folding</i>	16
Tabel 2.3 Penerapan <i>Tokenization</i>	17
Tabel 2.4 Penerapan <i>Stemming</i>	17
Tabel 2.5 Penerapan <i>Stopword removal</i>	17
Tabel 2.6 Penerapan <i>Data Cleaning</i>	18
Tabel 2.7 Penerapan <i>N-Gram</i>	19
Tabel 2.8 Tabel <i>contingency</i>	23
Tabel 2.9 Hasil hitung <i>expected value</i>	24
Tabel 2.10 Tabel <i>chi-square</i>	25
Tabel 2.11 Jenis-jenis kernel SVM.....	36
Tabel 3.1 Contoh Judul Berita	50
Tabel 3.2 Kriteria Pelabelan.....	52
Tabel 3.3 Contoh judul dan labelnya	52
Tabel 3.4 Contoh hasil <i>Cleaning</i>	53
Tabel 3.5 Contoh hasil <i>Case Folding</i>	54
Tabel 3.6 Contoh hasil <i>Tokenizing</i>	54
Tabel 3.7 Contoh hasil <i>Stopword removal</i>	54
Tabel 3.8 Contoh hasil <i>Stemming</i>	55
Tabel 3.9 Contoh kalimat dalam dokumen	55
Tabel 3.10 Daftar kata dengan nomer indeks.....	56
Tabel 3.11 Frekuensi <i>term</i> dokumen 1.....	56
Tabel 3.12 Frekuensi <i>term</i> dokumen 2.....	57
Tabel 3.13 Hasil Ekstraksi fitur <i>Counvectorizer</i>	57
Tabel 3.14 Contoh Tabel Kontingensi Chi-2.....	58

Tabel 3.15 Contoh Tabel Kontingensi IG.....	60
Tabel 4.1 Hasil Data judul berita yang sudah discraping.....	67
Tabel 4.2 Hasil Pelabelan.....	68
Tabel 4.3 Perbandingan Jumlah Data pada kelas Sentimen.....	68
Tabel 4.4 Hasil dari <i>cleaning data</i>	70
Tabel 4.5 Hasil dari <i>Case Folding</i>	71
Tabel 4.6 Hasil <i>tokenizing</i>	72
Tabel 4.7 Hasil <i>Stopword</i>	72
Tabel 4.8 Hasil <i>Stemming</i>	73
Tabel 4.9 Nilai <i>Chi-Square</i> pada masing-masing fitur	75
Tabel 4.10 Memilih 2100 Fitur yang mempengaruhi label	76
Tabel 4.11 Fitur dengan nilai <i>information gain</i> nilai tertinggi hingga nilai terendah. 77	
Tabel 4.12 Memilih 2200 Fitur yang mempengaruhi label dengan nilai <i>information gain</i>	78
Tabel 4.13 Eksperimen jumlah fold	80
Tabel 4.14 Hasil training dan testing SVM tanpa menggunakan seleksi fitur	83
Tabel 4.15 Hasil training dan testing SVM menggunakan seleksi fitur <i>Chi-Square</i> ..84	
Tabel 4.16 Hasil training dan testing SVM menggunakan seleksi fitur <i>Information Gain</i>	85
Tabel 4.17 Perbandingan hasil testing model	86
Tabel 4.18 Confusion Matrix SVM dengan <i>Chi-Square</i>	87
Tabel 4.19 Confusion Matriks kelas label positif	88
Tabel 4.20 Confusion Matriks kelas label netral	88
Tabel 4.21 Confusion Matriks kelas label negatif	88
Tabel 4.22 Hasil perhitungan <i>coherence</i> dan <i>perplexity</i>	99
Tabel 4.23 Hasil ekstraksi topik judul berita sentimen positif	100
Tabel 4.24 Hasil ekstraksi topik judul berita sentimen netral	101
Tabel 4.25 Hasil ekstraksi topik judul berita sentimen negatif	102
Tabel 4.26 Contoh Kesalahan prediksi Klasifikasi	105

DAFTAR GAMBAR

Gambar 2.1 <i>Workflow Scraping Website</i>	16
Gambar 2.2 <i>Data Customer Bank</i>	22
Gambar 2.3 Hasil perhitungan nilai <i>chi-square</i>	24
Gambar 2.4 Contoh dataset prediksi beli Komputer	27
Gambar 2.5 Contoh hasil pemisahan atribut berdasarkan class	28
Gambar 2.6 Atribut umur.....	28
Gambar 2.7 Hasil perhitungan <i>entropy</i> masing-masing atribut	29
Gambar 2.8 Contoh hasil nilai <i>information gain</i> pada atribut umur	30
Gambar 2.9 Hasil perhitungan nilai <i>information gain</i> pada semua atribut	30
Gambar 2.10 Urutan atribut secara <i>descending</i>	31
Gambar 2.11 <i>Hyperplane</i> SVM memisahkan class positif dan negatif	34
Gambar 2.12 Pemetaan Input Space Berdimensi Dua dengan Pemetaan ke Dimensi Tinggi	34
Gambar 2.13 <i>Workflow K-Fold Cross-Validation, K = 5</i>	38
Gambar 2.14 Contoh <i>WordCloud</i>	39
Gambar 2.15 Ilustrasi LDA.....	43
Gambar 2.16 Plate notation LDA.....	44
Gambar 3.1 Rancangan Penelitian	49
Gambar 3.2 Rancangan proses <i>scraping</i>	51
Gambar 3.3 Alur perhitungan <i>chi-square</i> secara manual.....	58
Gambar 3.4 Alur perhitungan <i>information gain</i> secara manual.....	60
Gambar 3.5 Alur klasifikasi SVM	63
Gambar 4.1 Website Detik.com, Kompas.com dan Balipost.....	67
Gambar 4.2 Persentase Distribusi Sentimen	69
Gambar 4.3 Tampilan data sebelum proses <i>preprocessing</i>	70

Gambar 4.4 Hasil Ekstraksi fitur menggunakan <i>Countvectorizer</i> dengan <i>Binary Term Presence</i>	74
Gambar 4.5 Hasil seleksi fitur <i>Chi-Square</i> dengan jumlah fitur 2100.....	77
Gambar 4.6 Hasil seleksi fitur <i>Information Gain</i> dengan jumlah fitur 2200	78
Gambar 4.7 Code <i>K-Fold cross validation</i>	80
Gambar 4.8 Confusion Matriks SVM dengan seleksi fitur <i>Chi-Square</i>	87
Gambar 4.9 Perbandingan hasil akurasi <i>testing</i> model	93
Gambar 4.10 Perbandingan hasil akurasi <i>training</i> model.....	93
Gambar 4.11 <i>WorldCloud</i> Sentimen Positif.....	95
Gambar 4.12 <i>WorldCloud</i> Sentimen Netral.....	96
Gambar 4.13 <i>WorldCloud</i> Sentimen Negatif.....	97
Gambar 4.14 Hasil pemodelan topik judul berita sentimen positif.....	100
Gambar 4.15 <i>Intertopic Distance Map</i> pada judul berita positif	101
Gambar 4.16 Hasil pemodelan topik judul berita sentimen netral.....	102
Gambar 4.17 <i>Intertopic Distance Map</i> pada judul berita netral.....	102
Gambar 4.18 Hasil pemodelan topik judul berita sentimen negatif.....	103
Gambar 4.19 <i>Intertopic Distance Map</i> pada judul berita negatif.....	103

