

PERBANDINGAN KINERJA EKSTRAKSI FITUR TF-IDF DAN BOW DALAM SENTIMEN ANALISIS BPJS KESEHATAN DENGAN ALGORITMA XGBOOST

SKRIPSI



**PROGRAM STUDI SISTEM INFORMASI (SI) (S1)
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA**



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetak dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSR-E - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSR-E
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.



**Balai Besar
Sertifikasi
Elektronik**

SKRIPSI

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

Menyetujui

Pembimbing I	Ir. I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T. NIP.198907112020122004
Pembimbing II	I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs. NIP.198910262019031004



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

Skripsi oleh Made Donita Maharani ini
telah dipertahankan di depan dewan penguji
Pada tanggal 06 Januari 2026

Dewan Penguji

Ketua	Ir. Kadek Teguh Dermawan, S.Pd., M.Kom. NIP.199406132025061003
Anggota	Ir. Gede Surya Mahendra, S.Pd., M.Kom. NIP.199003132022031009
Anggota	Ir. I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T. NIP.198907112020122004
Anggota	I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs. NIP.198910262019031004



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSR - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSR
- Cetak dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

Diterima oleh Panitia Ujian Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha
guna memenuhi syarat-syarat untuk mencapai gelar Sarjana Komputer

Menyetujui

Ketua Ujian	Made Windu Antara Kesiman, S.T., M.Sc., Ph.D. NIP.198211112008121001
Sekretaris Ujian	Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng. NIP.199005152019031008

Mengesahkan,
Dekan Fakultas Teknik dan Kejuruan



Prof. Dr. Kadek Rihendra Dantes, S.T., M.T.
NIP.197912012006041001

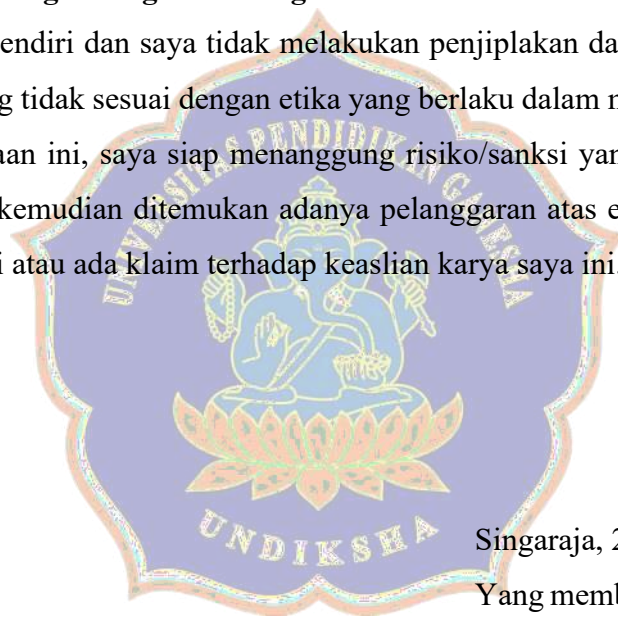


- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetak dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

LEMBAR PERNYATAAN KEASLIAN TULISAN

Nama : Made Donita Maharani
Tempat, Tanggal Lahir : Singaraja, 5 Desember 2002
NIM : 2115091065
Program Studi : S1 Sistem Informasi
Fakultas : Teknik dan Kejuruan

Dengan ini saya menyatakan bahwa karya tulis yang berjudul **Perbandingan Kinerja Ekstraksi Fitur Tf-Idf & Bow Dalam Sentimen Analisis Bpjs Kesehatan Dengan Algoritma Xgboost** beserta seluruh isinya adalah benar-benar karya sendiri dan saya tidak melakukan penjiplakan dan pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran atas etika keilmuan dalam karya saya ini atau ada klaim terhadap keaslian karya saya ini.



Singaraja, 20 Januari 2026

Yang membuat pernyataan,



Made Donita Maharani

NIM. 2115091065

PRAKATA

Dengan rasa syukur yang mendalam, penulis menyampaikan terimakasih dan rasa syukur yang sebesar-besarnya kepada Ida Shang Hyang Widhi Wasa atas segala berkah, kelancaran, petunjuk, dan kekuatan yang telah diberikan, sehingga penulis dapat menuntaskan tugas akhir yang berjudul **Perbandingan Kinerja Ekstraksi Fitur TF-IDF dan BOW Dalam Sentimen Analisis BPJS Kesehatan dengan Algoritma XGBoost** dengan baik. Penulisan karya akhir ini dilakukan untuk memenuhi persyaratan akademik untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Teknik dan Kejuruan, Universitas Pendidikan Ganesha.

Pada perjalanan dalam penulisan tugas akhir ini, penulis menghadapi begitu banyak tantangan dan rintangan. Namun, banyaknya doa dan semangat dari berbagai pihak, penulis dapat menyelesaikan penelitian ini dengan lancar. Pada kesempatan ini penulis ingin mengucapkan rasa terima kasih dan hormat kepada pihak-pihak yang terkait dalam pembuatan tugas akhir ini dengan tulus.

1. Bapak Prof. Dr. Kadek Rihendra Dantes, S.T., M.T., yang menjabat sebagai Dekan Fakultas Teknik dan Kejuruan, telah memberikan motivasi serta fasilitas yang memungkinkan penulis untuk menyelesaikan studi sesuai dengan rencana.
2. Bapak Dr. Putu Hendra Suputra, S.Kom., M.Cs., yang menjabat sebagai Ketua Jurusan Teknik Informatika, memberikan motivasi dan fasilitas yang sangat berharga dalam proses penyelesaian skripsi ini.
3. Bapak Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng., sebagai Koordinator Program Studi Sistem Informasi yang telah memberikan arahan, petunjuk,

dan motivasi kepada penulis dalam penyelesaian skripsi ini.

4. Ibu sebagai Pembimbing I yang telah memberikan bimbingan, arahan, petunjuk, dan motivasi kepada penulis dalam penyelesaian skripsi ini.
5. Bapak sebagai Pembimbing II yang telah memberikan bimbingan, arahan, petunjuk, dan motivasi kepada penulis dalam penyelesaian skripsi ini. 2
6. Seluruh Dosen jurusan Teknik Informatika / Program Studi Sistem Informasi yang dengan ikhlas berbagi ilmu dan pengalaman selama masa kuliah, diharapkan agar semua pengetahuan yang peneliti peroleh di kuliah, diharapkan agar semua pengetahuan yang luas bagi masyarakat.
7. Orang tua, kakak dan sahabat secara khusus yang telah memberikan doa dan dukungan yang sangat berarti kepada peneliti dalam menyelesaikan skripsi ini.
8. Rekan-rekan mahasiswa Program Studi Sistem Informasi yang telah memberikan dorongan semangat kepada penulis dalam menyelesaikan skripsi ini.
9. Semua pihak yang tidak dapat disebutkan satu per satu namun turut serta dalam membantu penyusunan skripsi ini.

Penulis menyadari sepenuhnya bahwa skripsi ini masih jauh dari kesempurnaan, mengingat keterbatasan pengetahuan, pengalaman, dan kemampuan yang dimiliki dalam menyusun tugas akhir ini. Meskipun telah berusaha dengan sungguh-sungguh dalam setiap tahap penyusunannya, penulis tidak menutup kemungkinan masih terdapat banyak kekurangan.

Oleh karena itu, penulis sangat menerima dengan baik kritik dan saran yang bersifat membangun dari berbagai pihak sebagai bahan untuk terus belajar. Besar

harapan penulis agar skripsi ini dapat memberikan manfaat, menjadi bahan rujukan yang bermanfaat bagi pihak yang berkepentingan, serta memberi kontribusi positif dalam pengembangan ilmu pengetahuan, khususnya di bidang sistem informasi dan analisis sentimen.

Singaraja, 20 Januari 2026

Penulis



DAFTAR ISI

	Halaman
PRAKATA.....	i
ABSTRAK.....	iv
ABSTRACT	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	x
DAFTAR LAMPIRAN.....	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	7
1.3 Tujuan Penelitian.....	8
1.4 Ruang Lingkup.....	8
1.5 Manfaat Penelitian.....	9
BAB II TINJAUAN PUSTAKA	10
2.1 Penelitian Terdahulu.....	10
2.1.1 Penelitian Terkait BPJS Kesehatan	10
2.1.2 Penelitian Terkait Implementasi XGBoost.....	12
2.1.3 Penelitian Terkait Perbandingan Ekstraksi Fitur TF-IDF dengan BoW.....	15
2.2 BPJS Kesehatan.....	18
2.3 Dasar Teori	20
2.3.1 Aplikasi X	20
2.3.2 <i>Text Mining</i>	21
2.3.3 Analisis Sentimen	21
2.3.4 Ekstraksi Fitur	22
2.3.5 TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>)	22
2.3.6 BoW (<i>Bag of Words</i>)	24
2.3.7 XGBoost.....	24
2.3.8 <i>Confusion Matrix</i>	27

BAB III METODOLOGI PENELITIAN	29
3.1 Alur Penelitian.....	29
3.1.1 Identifikasi Masalah	30
3.1.2 Studi Literatur.....	31
3.1.3 <i>Text Preprocessing</i>	35
3.1.4 Ekstraksi Fitur TF-IDF dan BoW	46
3.1.5 Model Algoritma XGBoost.....	49
3.1.6 Evaluasi Model	50
3.1.7 Pembuatan <i>Interface</i>	52
BAB IV HASIL DAN PEMBAHASAN	29
4.1 Hasil	55
4.1.1 <i>Scraping Data</i>	55
4.1.2 <i>Labeling</i>	55
4.1.3 <i>Text Preprocessing</i>	58
4.1.4 Ekstraksi Fitur TF-IDF dan BOW	60
4.1.5 Implementasi Model XGBoost.....	62
4.1.6 <i>User Interface</i>	84
4.2 Pembahasan.....	86
BAB V PENUTUP	84
5.1 Simpulan	84
5.2 Saran.....	85
DAFTAR PUSTAKA	87
LAMPIRAN	93

DAFTAR GAMBAR

Gambar	Halaman
Gambar 3.1 Alur Penelitian	30
Gambar 3. 2 Pipeline Akuisisi Data	30
Gambar 3.3 Halaman Input File	53
Gambar 3.4 Halaman Pemilihan Metode	53
Gambar 3.5 Halaman Proses Analisis & Hasil	54
Gambar 4. 1 Hasil Scraping.....	55
Gambar 4. 2 Diagram Batang Distribusi Sentimen Masyarakat Terhadap BPJS Kesehatan	57
Gambar 4. 3 Hasil Cleansing.....	58
Gambar 4. 4 Hasil Case Folding.....	59
Gambar 4. 5 Hasil Tokenizing.....	59
Gambar 4. 6 Hasil Stopword Removal.....	59
Gambar 4. 7 Hasil Normalisasi.....	60
Gambar 4. 8 Hasil Stemming	60
Gambar 4. 9 Hasil Ekstraksi dengan BOW	61
Gambar 4. 10 Hasil Ekstraksi dengan TF-IDF	61
Gambar 4. 11 Grafik Learning Curve.....	62
Gambar 4. 12 Confusion Matrix TF-IDF + XGBoost + scale post weight	68
Gambar 4. 13 Confusion Matrix TF-IDF + XGBoost + sample weight	69
Gambar 4. 14 Confusion Matrix TF-IDF + XGBoost + (SMOTE+Sample_Weight)	69
Gambar 4. 15 Confusion Matrix TF-IDF + XGBoost.....	70
Gambar 4. 16 Confusion Matrix TF-IDF + XGBoost + scale post weight (split data 80:20).....	71
Gambar 4. 17 Confusion Matrix TF-IDF + XGBoost + sample weight (split data 80:20).....	71
Gambar 4. 18 Confusion Matrix TF-IDF+ XGBoost + SMOTE + Sample_Weight (split data 80:20)	72
Gambar 4. 19 Confusion Matrix BOW + XGBoost + scale_post_weight.....	78
Gambar 4. 20 Confusion Matrix BOW + XGBoost + sample weight.....	78

Gambar 4. 21 Confusion Matrix BOW+ XGBoost + SMOTE+Sample_Weight.	79
Gambar 4. 22 Confusion Matrix BOW + XGBoost	80
Gambar 4. 23 Confusion Matrix BOW + XGBoost + scale post weight (split data 80:20)	80
Gambar 4. 24 Confusion Matrix BOW + XGBoost + sample weight (split data 80:20)	81
Gambar 4. 25 Confusion Matrix BOW + XGBoost + SMOTE + Sample Weight (split data 80:20)	82
Gambar 4. 26 Word Cloud Positif	83
Gambar 4. 27 Word Cloud Negatif	84
Gambar 4. 28 Tampilan Tab Upload Dataset	84
Gambar 4.29 Tampilan Hasil Preprocessing	85
Gambar 4.30 Tampilan Tab Ekstraksi Fitur	85
Gambar 4.31 Tampilan Tab Analisis Sentimen	86



DAFTAR TABEL

Tabel	Halaman
Tabel 2. 1 Confusion Matrix	27
Tabel 3. 1 Contoh Labeling Data	33
Tabel 3. 2 Hasil Cleansing.....	35
Tabel 3. 3 Hasil Case Folding	39
Tabel 3. 4 Hasil Tokenizing	40
Tabel 3. 5 Hasil Stopword Removal.....	43
Tabel 3. 6 Hasil Normalisasi	44
Tabel 3. 7 Hasil Stemming	45
Tabel 3. 8 Dokumen Bobot Kata.....	46
Tabel 3. 9 Contoh Perhitungan BOW.....	48
Tabel 3. 10 Contoh Hasil Vektorisasi BOW	48
Tabel 3. 11 Contoh Hasil Confusion Matrix dengan 10-Fold Cross Validation ...	51
Tabel 4. 1 Hasil Labeling Data.....	57
Tabel 4. 2 Skenario Pengujian XGBoost + TF-IDF	63
Tabel 4. 3 Performa Hasil Pengujian XGBoost + TF-IDF	67
Tabel 4. 4 Skenario Pengujian XGBoost + BOW	73
Tabel 4. 5 Performa Hasil Pengujian XGBoost + BOW	77
Tabel 4. 6 Perbandingan TF-IDF dan BoW Berdasarkan Performa Rata-Rata.....	90

DAFTAR LAMPIRAN

Lampiran 1. Riwayat Hidup	93
Lampiran 2. Surat Ketersediaan Pelabelan Data	94
Lampiran 3. Sertifikat Pendidik.....	95
Lampiran 4. Surat Keterangan Validasi	96
Lampiran 5. Dokumentasi bersama Validator	100
Lampiran 6. Proses Scraping Data di Aplikasi X.....	101
Lampiran 7. Proses Preprocessing	107
Lampiran 8. Proses TF-IDF (Term Frequency3Inverse Document Frequency)	112
Lampiran 9. Proses BoW (Bag of Words).....	113
Lampiran 10. Proses XGBoost + TF-IDF/BoW + scale_post_weight.....	114
Lampiran 11. Proses XGBoost + TF-IDF/BoW + sample_weight.....	117
Lampiran 12. Proses XGBoost + TF-IDF/BoW + sample_weight + SMOTE...	120
Lampiran 13. Proses XGBoost + TF-IDF/BoW	124
Lampiran 14. Proses XGBoost + TF-IDF/BoW + scale_post_weight (split data 80:20).....	127
Lampiran 15. XGBoost + TF-IDF/BoW + sample_weight (split data 80:20)....	130
Lampiran 16. XGBoost + TF-IDF/BoW+ sample_weight + SMOTE (split data 80:20).....	133
Lampiran 17. UI (User Interface)	137