

**ANALISIS PERBANDINGAN METODE PEMBOBOTAN
TF-RF DAN TF-ABS PADA KATEGORISASI DOKUMEN
BERITA DI BALAI DIKLAT INDUSTRI DENPASAR
MENGUNAKAN KLASIFIKASI K-NEAREST
NEIGHBORS**



OLEH

I KADEK WAHYU DANANJAYA

NIM 1815091039

SAMPUL

PROGRAM STUDI SISTEM INFORMASI

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK DAN KEJURUAN

UNIVERSITAS PENDIDIKAN GANESHA

SINGARAJA

2023



**ANALISIS PERBANDINGAN METODE PEMBOBOTAN
TF-RF DAN TF-ABS PADA KATEGORISASI DOKUMEN
BERITA DI BALAI DIKLAT INDUSTRI DENPASAR
MENGUNAKAN KLASIFIKASI K-NEAREST
NEIGHBORS**

SKRIPSI

Diajukan kepada

Universitas Pendidikan Ganesha

Untuk Memenuhi Salah Satu Persyaratan dalam Menyelesaikan

Program Sarjana Komputer

Oleh

I Kadek Wahyu Dananjaya

NIM 1815091039

PROGRAM STUDI SISTEM INFORMASI

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK DAN KEJURUAN

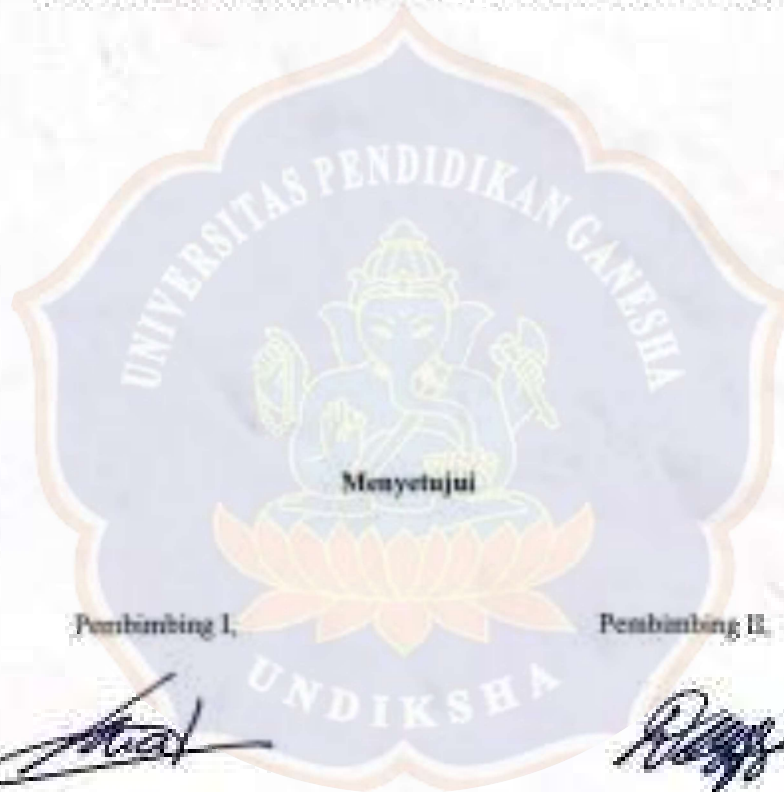
UNIVERSITAS PENDIDIKAN GANESHA

SINGARAJA

2023

SKRIPSI

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



Pembimbing I,

Pembimbing II,

I Ganti Ayu Agung Diani Indrikesi, S.Kom, M.T.
NIP. 198907112020122004

I Made Dendi Masruman, S.Pd, M.Eng.
NIP. 199005152019031003

Skripsi oleh I Kadek Wahyu Darmajaya ini
telah dipertahankan di depan dewan penguji
pada tanggal 24 November 2022

Dewan Penguji,



I Made Antri Pradnyana, S.T., M.T.

(Ketua)

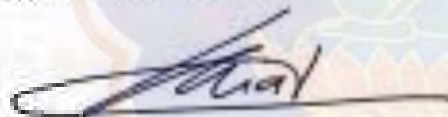
NIP. 198631182015041001



I Made Edy Listartha, S.Kom., M.Kom.

(Anggota)

NIP. 198608122019031005



I Gusti Ayu Agung Distri Indradewi, S.Kom., M.T.

(Anggota)

NIP. 198907112020122004



I Made Dendi Masyanjaya, S.Pd., M.Eng.

(Anggota)

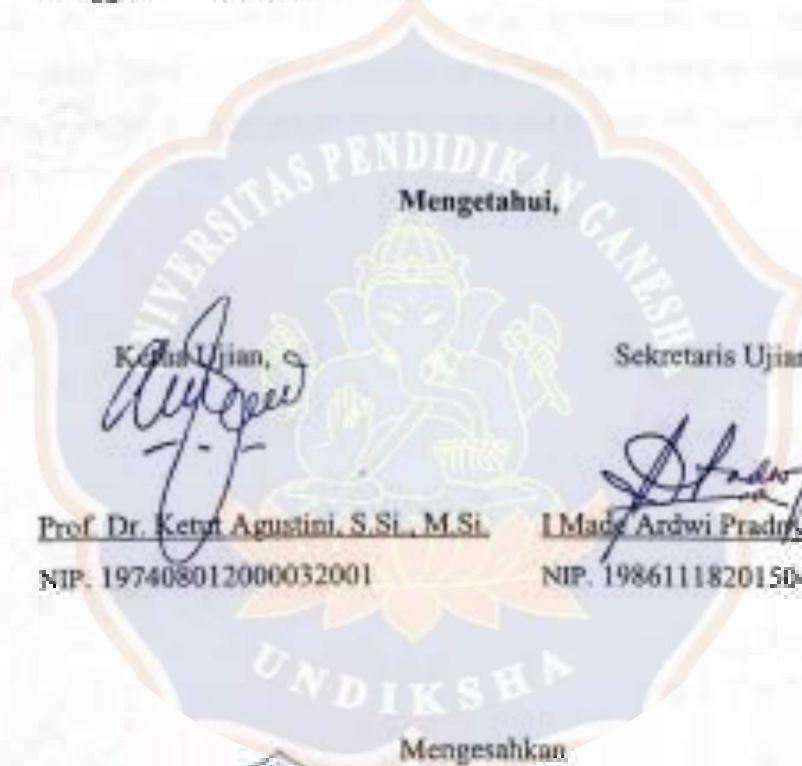
NIP. 199005152019031008

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

Pada:

Hari : Rabu

Tanggal : 25 Januari 2023



Ketua Ujian,

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

Sekretaris Ujian,

I Made Ardwi Pradnyana, S.T., M.T.
NIP. 198611182015041001

Mengesahkan

Dean Fakultas Teknik dan Kejuruan

Prof. Dr. Gede Sudirtha, S.Pd., M.Pd.

NIP. 197106161996021001

PERNYATAAN

Dengan ini saya menyatakan bahwa karya tulis yang berjudul "ANALISIS PERBANDINGAN METODE PEMBOBOTAN TF-RF DAN TF-ABS PADA KATEGORISASI DOKUMEN BERITA DI BALAI DIKLAT INDUSTRI DENPASAR MENGGUNAKAN KLASIFIKASI K-NEAREST NEIGHBORS" 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, 25 Januari 2023

Yang membuat pernyataan,



Kadek Wahyu Dananjaya

NIM. 1815091039

PRAKATA

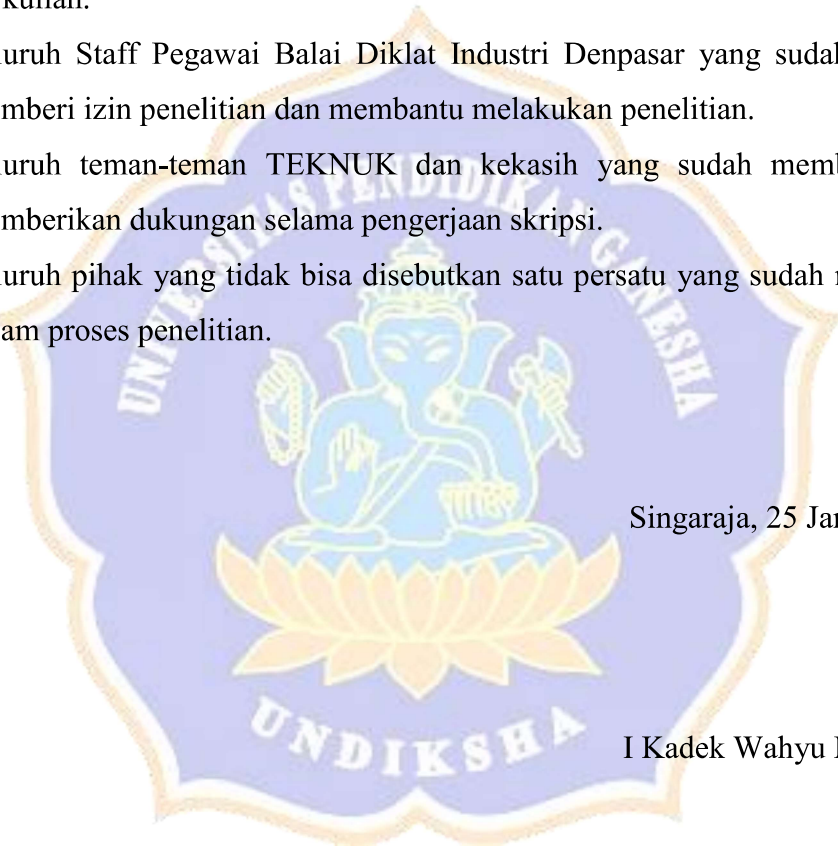
Segala puji dan syukur penulis panjatkan kepada Ida Sang Hyang Widhi Wasa, atas diberikan berkat dan rahmat-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “ANALISIS PERBANDINGAN METODE PEMBOBOTAN TF-RF DAN TF-ABS PADA KATEGORISASI DOKUMEN BERITA DI BALAI DIKLAT INDUSTRI DENPASAR MENGGUNAKAN KLASIFIKASI *K-NEAREST NEIGHBORS*” sebagai syarat dalam menyelesaikan Program Sarjana (S1) Program Studi Sistem Informasi. Penulis menyadari bahwa dalam penelitian ini masih memiliki banyak kekurangan dan masih jauh dari kata sempurna. Dalam penyusunan skripsi banyak hambatan dan kendala yang dihadapi penulis, namun bisa terselesaikan berkat adanya bantuan dan bimbingan dari berbagai pihak. Untuk itu, penulis menyampaikan ucapan terimakasih kepada:

1. Prof. Dr. I Wayan Lasmawan, M.Pd. Selaku Rektor Universitas Pendidikan Ganesha (Undiksha)
2. Dr. I Gede Sudirtha, S.Pd., M.Pd. Selaku Dekan Fakultas Teknik dan Kejuruan Universitas Pendidikan Ganesha.
3. Dr. Luh Joni Erawati Dewi, S.T., M.Pd. Selaku Ketua Jurusan Teknik Informatika.
4. I Made Ardwi Pradnyana, S.T., M.T. Selaku Koordinator Program Studi Sistem Informasi dan juga sebagai Dosen Penguji I yang telah memberikan saran dan arahan penyusunan skripsi.
5. I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T. Selaku Dosen Pembimbing I yang sudah berkenan memberikan bimbingan dan arahan penyusunan skripsi.
6. I Made Dendi Maysanjaya, S.Pd., M.Eng.. Selaku Dosen Pembimbing II yang sudah meluangkan waktu untuk memberikan bimbingan penyusunan skripsi.
7. I Made Edy Listartha, S.Kom., M.Kom. Selaku Dosen Penguji II yang sudah memberikan saran dan arahan penelitian skripsi.

8. Gede Aditra Pradnyana, S.Kom., M.Kom. Selaku Dosen Pembimbing yang sudah memberikan saran dan arahan terbaik dalam pembuatan proposal skripsi.
9. Ibu Ade Asih Susiari Tantri, S.Pd., M.Pd., selaku Dosen Pendidikan Bahasa dan Sastra Indonesia yang telah bersedia meluangkan waktu membantu kegiatan penelitian.
10. Seluruh Dosen dan Staff Pegawai Fakultas Teknik dan Kejuruan.
11. Kedua Orang Tua yang selalu membantu dan memberi dukungan selama berkuliah.
12. Seluruh Staff Pegawai Balai Diklat Industri Denpasar yang sudah bersedia memberi izin penelitian dan membantu melakukan penelitian.
13. Seluruh teman-teman TEKNUK dan kekasih yang sudah membantu dan memberikan dukungan selama pengerjaan skripsi.
14. Seluruh pihak yang tidak bisa disebutkan satu persatu yang sudah membantu dalam proses penelitian.

Singaraja, 25 Januari 2023

I Kadek Wahyu Dananjaya



DAFTAR ISI

SAMPUL	i
HALAMAN JUDUL.....	iii
LEMBAR PERSETUJUAN PEMBIMBING	iv
LEMBAR PERSETUJUAN DOSEN PENGUJI.....	v
LEMBAR PERSETUJUAN DAN PENGESAHAN PANITIA UJIAN.....	vi
PERNYATAAN KEASLIAN TULISAN	vii
PRAKATA.....	ix
ABSTRAK	xi
ABSTRACT.....	xii
DAFTAR ISI.....	xiii
DAFTAR TABEL.....	xvi
DAFTAR GAMBAR.....	xvii
DAFTAR LAMPIRAN.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	6
1.4 Ruang Lingkup Penelitian.....	6
1.5 Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA.....	8
2.1 Studi Sebelumnya	8
2.2 Dasar Teori.....	16
2.2.1 Balai Diklat Industri Denpasar.....	16
2.2.2 Dokumen	20
2.2.3 Berita	21
2.2.4 Klasifikasi	21
2.2.5 Klasifikasi Dokumen.....	22
2.2.6 <i>Data Mining</i>	23
2.2.7 <i>Machine Learning</i>	27

2.2.8 <i>Text Mining</i>	31
2.2.9 <i>Text Preprocessing</i>	31
2.2.10 <i>K-Nearest Neighbors</i>	35
2.2.11 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	38
2.2.12 <i>Term Frequency-Relevance Frequency (TF-RF)</i>	39
2.2.13 <i>Term Frequency-Absolute (TF-ABS)</i>	39
2.2.14 <i>Confussion Matrix</i>	40
BAB III METODOLOGI PENELITIAN	43
3.1 Waktu dan Tempat Penelitian	43
3.2 Metode Penelitian.....	43
3.1.1 <i>Data Collection</i>	44
3.1.2 <i>Text Preprocessing</i>	47
3.1.3 <i>Term Weighting</i>	50
3.1.4 <i>Klasifikasi K-Nearest Neighbors</i>	55
3.1.5 <i>Validasi Model</i>	59
3.1.6 <i>Pengujian Confussion Matrix</i>	61
3.1.7 <i>Gambaran Umum Modul Klasifikasi Berita</i>	62
BAB IV HASIL DAN PEMBAHASAN	64
4.1 Implementasi	64
4.1.1 <i>Import Dataset</i>	64
4.1.2 <i>Proses Text Preprocessing</i>	65
4.1.2 Implementasi Metode <i>Term Weighting</i>	71
4.1.2.1 <i>Term Frequency-Relevance Frequency (TF-RF)</i>	71
4.1.2.2. <i>Term Frequency-Absolute (TF-ABS)</i>	75
4.1.3 <i>Implementasi Metode K-Nearest Neighbors</i>	78
4.1.4 <i>Pengujian Confussion Matrix</i>	80
4.2 Hasil dan Pembahasan.....	81
4.2.1 <i>Metode Pembobotan Kata TF-RF</i>	82
4.2.2 <i>Metode Pembobotan Kata TF-ABS</i>	93
4.2.3 <i>Perbandingan TF-RF dan TF-ABS</i>	103
4.2.4 <i>Modul Klasifikasi Berita</i>	108

BAB V PENUTUP	109
5.1 Kesimpulan	109
5.2 Saran.....	110
DAFTAR PUSTAKA	111
DAFTAR LAMPIRAN.....	116



DAFTAR TABEL

Tabel 2. 1 Rumus <i>Confussion Matrix</i>	40
Tabel 3. 1 Jumlah Dataset Penelitian	45
Tabel 3. 2 Contoh Dataset Berita	46
Tabel 3. 3 Contoh <i>Term Frequency</i>	50
Tabel 3. 4 Contoh Hasil Perhitungan RF	51
Tabel 3. 5 Contoh Hasil Pembobotan TF-RF.....	51
Tabel 3. 6 Contoh <i>Term Frequency</i> Untuk ABS.....	52
Tabel 3. 7 Contoh Hasil Perhitungan Kriteria Kelas A.....	53
Tabel 3. 8 Contoh Hasil Perhitungan Kriteria Kelas B.....	53
Tabel 3. 9 Contoh Hasil Bobot ABS.....	54
Tabel 3. 10 Contoh Hasil Nilai Mutlak Bobot ABS	54
Tabel 3. 11 Contoh Hasil TF-ABS.....	55
Tabel 3. 12 Contoh Hasil Perkalian Skalar	56
Tabel 3. 13 Contoh Hasil Perhitungan Panjang Vektor	57
Tabel 3. 14 Contoh Hasil <i>Cosine Similarity</i>	57
Tabel 3. 15 Contoh Nilai $k = 3$	58
Tabel 3. 16 Contoh Hasil Klasifikasi K-NN	59
Tabel 3. 17 Contoh Pengukuran <i>Confussion Matrix</i>	61
Tabel 4. 1. Hasil TF-RF dan TF-ABS.....	104

DAFTAR GAMBAR

Gambar 2. 1 Logo BDI Denpasar	16
Gambar 2. 2 Struktur Organisasi BDI Denpasar.....	17
Gambar 2. 3 Struktur Organisasi Divisi TI BDI Denpasar	18
Gambar 2. 4 <i>Knowledge Discovery in Database (KDD)</i>	24
Gambar 2. 5 <i>Semi-Supervised Learning (Han et al., 2012)</i>	29
Gambar 2. 6 <i>Active Learning (Han et al., 2012)</i>	30
Gambar 2. 7 Contoh Hasil <i>Text Preprocessing</i>	33
Gambar 2. 8 Nilai K Kecil, Sedang dan Besar.....	36
Gambar 3. 1 Prosedur Kerja Penelitian.....	44
Gambar 3. 2 Contoh Hasil Proses <i>Case Folding</i>	48
Gambar 3. 3 Contoh Hasil Proses <i>Tokenizing</i>	48
Gambar 3. 4 Contoh Hasil Proses <i>Filtering</i>	49
Gambar 3. 5 Contoh Hasil Proses <i>Stemming</i>	50
Gambar 3. 6 Ilustrasi <i>K-Fold Cross Validation</i>	60
Gambar 3. 7 Nilai Aktual vs Nilai Prediksi	61
Gambar 3. 8 Contoh Tampilan Modul	63
Gambar 4. 1. <i>Source Code Import Dataset</i>	64
Gambar 4. 2. Dataset.....	65
Gambar 4. 3. <i>Source Code Case Folding</i>	65
Gambar 4. 4. Hasil <i>Case Folding</i>	66
Gambar 4. 5. <i>Source Code Tokenizing</i>	67
Gambar 4. 6. Hasil <i>Tokenizing</i>	68
Gambar 4. 7. <i>Source Code Filtering</i>	68
Gambar 4. 8. Hasil <i>Filtering</i>	69
Gambar 4. 9. <i>Source Code Stemming</i>	70
Gambar 4. 10. Hasil <i>Stemming</i>	71
Gambar 4. 11. <i>Source Code Term Frequency</i>	72

Gambar 4. 12. Hasil <i>Term Frequency</i>	72
Gambar 4. 13. <i>Source Code Relevance Frequency</i>	73
Gambar 4. 14. Hasil <i>Relevance Frequency</i>	73
Gambar 4. 15. <i>Source Code Term Frequency - Relevance Frequency</i>	74
Gambar 4. 16. Hasil <i>Term Frequency-Relevance Frequency</i>	74
Gambar 4. 17 <i>Source Code Absolute</i>	76
Gambar 4. 18 Hasil <i>Absolute</i>	77
Gambar 4. 19 <i>Source Code Term Frequency-Absolute</i>	77
Gambar 4. 20 Hasil <i>Term Frequency-Absolute</i>	78
Gambar 4. 21 <i>Source Code K-NN</i> Pada TF-RF dan TF-ABS	79
Gambar 4. 22 <i>Source Code Confussion Matrix</i> Pada TF-RF dan TF-ABS	81
Gambar 4. 23. Hasil <i>Confussion Matrix</i> TF-RF K=3	82
Gambar 4. 24. Hasil Klasifikasi TF-RF K=3	84
Gambar 4. 25. Hasil <i>Confussion Matrix</i> TF-RF K=5	85
Gambar 4. 26. Hasil Klasifikasi TF-RF K=5	87
Gambar 4. 27. Hasil <i>Confussion Matrix</i> TF-RF K=7	88
Gambar 4. 28. Hasil Klasifikasi TF-RF K=7	89
Gambar 4. 29. Hasil <i>Confussion Matrix</i> TF-RF K=9	90
Gambar 4. 30. Hasil Klasifikasi TF-RF K=9	92
Gambar 4. 31. Hasil <i>Confussion Matrix</i> TF-ABS K=3.....	93
Gambar 4. 32. Hasil Klasifikasi TF-ABS K=3	95
Gambar 4. 33. Hasil <i>Confussion Matrix</i> TF-ABS K=5.....	96
Gambar 4. 34. Hasil Klasifikasi TF-ABS K=5	97
Gambar 4. 35. Hasil <i>Confussion Matrix</i> TF-ABS K=7.....	98
Gambar 4. 36. Hasil Klasifikasi TF-ABS K=7	100
Gambar 4. 37. Hasil <i>Confussion Matrix</i> TF-ABS K=9.....	101
Gambar 4. 38. Hasil Klasifikasi TF-ABS K=9	103
Gambar 4. 39. Hasil <i>K-Fold Cross Validation</i>	107
Gambar 4. 40. Modul Klasifikasi Berita	108

DAFTAR LAMPIRAN

Lampiran 1. Riwayat Hidup.....	116
Lampiran 2. Dokumentasi Mengumpulkan Data	117
Lampiran 3. Tampilan Website Layanan Kabar Insan Oke.....	118
Lampiran 4. Hasil Validasi Dataset	119
Lampiran 5. Hasil Teks <i>Preprocessing</i>	120

