

**SISTEM PEMODELAN TOPIK PADA PENCARIAN
DOKUMEN KARYA AKHIR MENGGUNAKAN
METODE *LATENT DIRICHLET ALLOCATION* (LDA)
(STUDI KASUS: SISTEM REPOSITORY UNDIKSHA)**



**OLEH
I MADE SURYASA
NIM. 1615051088**

**PRODI PENDIDIKAN TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA**

2023

**SISTEM PEMODELAN TOPIK PADA PENCARIAN
DOKUMEN KARYA AKHIR MENGGUNAKAN
METODE *LATENT DIRICHLET ALLOCATION* (LDA)
(STUDI KASUS: SISTEM REPOSITORY UNDIKSHA)**

SKRIPSI

**Diajukan kepada
Universitas Pendidikan Ganesha
Untuk Memenuhi Salah Satu Syarat Dalam Menyelesaikan Program
Sarjana Pendidikan Teknik Informatika**

**Oleh
I Made Suryasa
Nim. 1615051088**

**PRODI PENDIDIKAN TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA
2023**

SKRIPSI

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

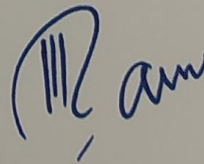
Menyetujui,

Pembimbing I,



Gede Aditra Pradnyana, S.Kom., M.Kom.
NIP. 198901192015041004

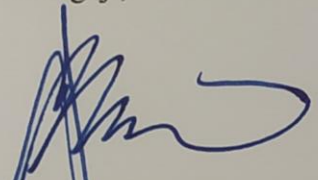
Pembimbing II,



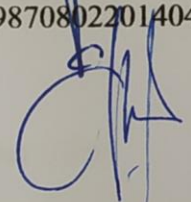
Ida Bagus Nyoman Pascima, S.Pd., M.Cs.
NIP. 198907132019031017

Skripsi oleh I Made Suryasa ini
Telah dipertahankan di depan penguji
Pada tanggal: 11 Oktober 2022

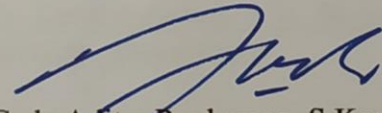
Dewan Penguji,


Gede Saindra Santyadiputra, S.T., M.Cs.
NIP. 198708022014041001

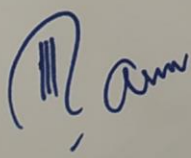
(Ketua)


I Nengah Eka Mertayasa, S.Pd., M.Pd.
NIP. 199503022019031006

(Anggota)


Gede Aditra Pradnyana, S.Kom., M.Kom.
NIP. 198901192015041004

(Anggota)


Ida Bagus Nyoman Pascima, S.Pd., M.Cs.
NIP. 198907132019031017

(Anggota)

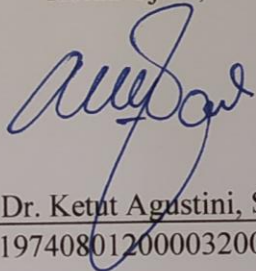
Diterima oleh Panitia Ujian Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha Singaraja
Guna memenuhi syarat-syarat untuk mencapai Gelar Sarjana Pendidikan

Pada:

Hari : Selasa
Tanggal : 21 Pebruari 2023

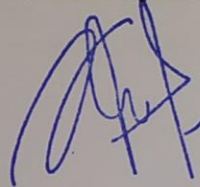
Mengetahui,

Ketua Ujian,



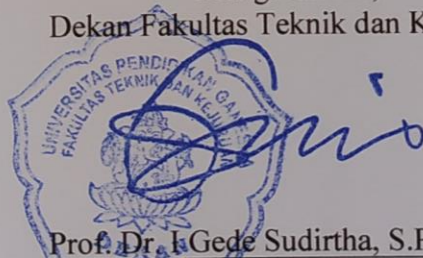
Prof. Dr. Ketut Agustini, S.Si., M.Si.
NIP. 197408012000032001

Sekretaris Ujian,



Dr.phil., Dessy Seri Wahyuni S.Kom., M.Eng.
NIP. 198502152008122007

Mengesahkan,
Dekan Fakultas Teknik dan Kejuruan



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

LEMBAR PERNYATAAN

Dengan ini penulis menyatakan bahwa karya tulis yang berjudul “**Sistem Pemodelan Topik pada Pencarian Dokumen Karya Akhir Menggunakan Metode *Latent Dirichlet Allocation* (LDA) (Studi Kasus: Sistem Repositori Undiksha)**” beserta seluruh isinya adalah 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 di kemudian ditemukan adanya pelanggaran atas etika keilmuan dalam karya saya ini, atau ada klaim terhadap keaslian karya saya ini.

Singaraja, 11 Oktober 2022

Yang membuat pernyataan,



1 Made Suryasa
NIM. 1615051088

KATA PENGANTAR

PENULISAN SKRIPSI INI KUPERSEMBAHKAN UNTUK

IDA SANG HYANG WIDHI WASA

Atas Berkat dan Restu-Nya, Penulis Bisa Menyelesaikan Skripsi Ini. Serta Atas Segala Kemudahan dan Kelancaran di Tiap Jalan yang Telah Diberikan-Nya.

ORANG TUAKU TERCINTA

(I Ketut Susila & Ni Wayan Suasti)

Yang Sudah Membimbing dan Mendidik Penulis dengan Penuh Kasih Sayang dan Keikhlasan serta Selalu Memberikan Penulis Semangat, Motivasi, Dukungan dan Doa dalam Setiap Langkah yang Penulis Tempuh pada Jenjang Pendidikan. Terima kasih Atas Segala Kebutuhan yang Dipenuhi Baik Berupa Materi dan Kebahagiaan serta Kenikmatan Hidup Lahir Batin yang Begitu Melimpah.

SAUDARIKU

(Ni Luh Putu Superyadi & Ni Luh Komang Suryani)

Yang selalu menghibur, memotivasi, mendorong penulis untuk menyelesaikan skripsi dan menyemangati penulis ketika mengalami kesusahan saat mengerjakan skripsi. Terima kasih atas segala doa yang telah terpanjatkan dan telah menyelipkan namaku di antara doa-doa yang indah itu.

**DAN JUGA TERIMA KASIH PENULIS UCAPKAN KEPADA
SELURUH STAF DOSEN PTI**

Yang telah membimbing penulis dengan sabar dalam menyelesaikan skripsi ini

REKAN-REKAN SEPERJUANGAN

Yang tidak dapat penulis sebutkan satu persatu yang telah memotivasi dan mendukung penulis menyelesaikan skripsi ini.

MOTTO

*“Hidup Layaknya Gambling Yang Wajib Dilakukan, Hanya Saja
Kebijaksanaan Dalam Bertaruh Yang Membedakannya”*



PRAKATA

Puji Syukur penulis panjatkan kepada Ida Sang Hyang Widhi Wasa penulis dapat menyelesaikan skripsi yang berjudul “**Sistem Pemodelan Topik pada Pencarian Dokumen Karya Akhir Menggunakan Metode *Latent Dirichlet Allocation (LDA)* (Studi Kasus: Sistem Repositori Undiksha)**”. Skripsi ini disusun guna memenuhi persyaratan mencapai gelar sarjana pendidikan dalam bidang Pendidikan Teknik Informatika di Universitas Pendidikan Ganesha.

Dalam menyelesaikan skripsi ini, penulis banyak mendapat bantuan baik berupa moral maupun material dari berbagai pihak. Untuk itu, dalam kesempatan ini, penulis mengucapkan terima kasih kepada:

1. Ibu Dr. phil. Dessy Seri Wahyuni, S.Kom., M.Eng. selaku Koordinator Program Studi Pendidikan Teknik Informatika.
2. Bapak Gede Aditra Pradnyana, S.Kom., M.Kom. selaku Pembimbing I yang telah banyak memberikan bimbingan, dan waktu yang telah diluangkan di tengah-tengah kesibukan beliau kepada penulis sehingga skripsi ini dapat terselesaikan tepat pada waktunya.
3. Bapak Ida Bagus Nyoman Pascima, S.Pd., M.Cs. selaku Pembimbing II telah banyak memberikan motivasi, bimbingan, dan waktu yang telah diluangkan di tengah-tengah kesibukan beliau kepada penulis sehingga skripsi ini dapat terselesaikan tepat pada waktunya.
4. Bapak Gede Saindra Santyadiputra, S.T., M.Cs. selaku Penguji I sekaligus sebagai Ahli/Validator data yang telah banyak meluangkan waktu, tenaga, dan dengan penuh kesabaran dalam memberikan bimbingan kepada penulis sehingga skripsi ini dapat diselesaikan.
5. Bapak I Nengah Eka Mertayasa, S.Pd.,M.Pd. selaku Penguji II telah banyak meluangkan waktu, tenaga, dan dengan penuh kesabaran dalam memberikan bimbingan kepada penulis sehingga skripsi ini dapat diselesaikan.
6. Semua pihak yang tidak dapat penulis sebutkan satu per-satu yang telah memberikan bantuan dan motivasi demi kelancaran proses penelitian ini.

Penulis menyadari sepenuhnya bahwa apa yang tersaji dalam skripsi ini masih jauh dari sempurna karena keterbatasan kemampuan yang penulis miliki. Untuk itu demi kesempurnaan skripsi ini, penulis mengharapkan segala kritik maupun saran yang sifatnya membangun dari berbagai pihak. Penulis berharap skripsi ini dapat bermanfaat dan berguna bagi kita semua khususnya bagi pengembangan dunia pendidikan.

Singaraja, 11 Oktober 2022



Penulis

DAFTAR ISI

	HALAMAN
LEMBAR PERNYATAAN	vi
KATA PENGANTAR	vii
PRAKATA	ix
ABSTRAK	xi
ABSTRACT	xii
DAFTAR ISI	xiii
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN	xviii
JUDUL PENELITIAN	xix
IDENTITAS PENELITI	xix
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 IDENTIFIKASI MASALAH	7
1.3 RUMUSAN MASALAH	8
1.4 TUJUAN PENELITIAN	8
1.5 BATASAN MASALAH	8
1.6 MANFAAT PENELITIAN	9
BAB II KAJIAN PUSTAKA DAN LANDASAN TEORI	11
2.1 KAJIAN PUSTAKA	11
2.2 LANDASAN TEORI	18
2.2.1 Karya Akhir	18
2.2.2 <i>Institutional Repositories</i>	19
2.2.3 <i>Text Mining</i>	23
2.2.4 <i>Latent Dirichlet Allocation</i>	30
2.2.5 <i>Information Retrieval</i>	36
2.2.6 <i>Jensen-Shannon Distance</i>	42
2.2.7 <i>Relevance Evaluation</i>	43
BAB III METODE PENELITIAN	47

3.1	JENIS PENELITIAN	47
3.2	MODEL PENELITIAN DAN PENGEMBANGAN.....	47
3.2.1	Fase Analisis	48
3.2.2	Fase Desain	52
3.2.3	Fase Implementasi.....	74
3.2.4	Fase Pengujian	74
3.2.5	Fase Perawatan.....	76
BAB IV HASIL DAN PEMBAHASAN		77
4.1	HASIL PENELITIAN.....	77
4.1.1	Hasil Implementasi.....	77
4.1.2	Hasil Pengujian	101
4.2	PEMBAHASAN	108
4.2.1	<i>Data Gathering</i>	108
4.2.2	<i>Preprocessing</i>	109
4.2.3	<i>Build Model</i>	111
4.2.4	Menghitung <i>Similarity, Ranking</i> , dan Kinerja IR	113
BAB V PENUTUP.....		121
5.1	KESIMPULAN.....	121
5.2	SARAN	123
DAFTAR PUSTAKA		124
RIWAYAT HIDUP.....		130
LAMPIRAN.....		131

DAFTAR TABEL

Tabel	Halaman
Tabel 2.1 Rangkuman Kajian Pustaka yang Relevan	16
Tabel 2.2 Confusion Matrix	44
Tabel 2.3 Kategori Penilaian Efektivitas Precision.....	45
Tabel 3.1 Kebutuhan Fungsional Sistem Pemodelan Topik Karya Akhir.....	50
Tabel 3.2 Kebutuhan Non-Fungsional Sistem Pemodelan Topik Karya Akhir....	51
Tabel 3.3 Kamus Data.....	53
Tabel 3.4 Penjelasan DFD Level 1	55
Tabel 3.5 Penjelasan DFD Level 2 (Proses 1.0)	57
Tabel 3.6 Dataset Contoh Repositori	63
Tabel 3.7 Dataset Hasil Preprocessing	66
Tabel 3.8 Output Distribusi Kata per Topik dari LDA	69
Tabel 3.9 Output Distribusi Topik per Dokumen dari LDA.....	70
Tabel 4.1 File Implementasi Arsitektur Perangkat Lunak	79
Tabel 4.2 Karakter Data yang Tidak Digunakan	82
Tabel 4.3 Nilai Inisiasi Masing-Masing Parameter	87
Tabel 4.4 Label Topik Beserta Jumlah Dokumen yang Membahasnya.....	91
Tabel 4.5 Daftar Kueri yang Diujikan.....	96
Tabel 4.6 Label Topik Pada Model LDA.....	103
Tabel 4.7 Percobaan Pertama Mencari Dokumen Topik Pertama	103
Tabel 4.8 Percobaan Kedua Mencari Dokumen Topik Kedua	104
Tabel 4.9 Percobaan Ketiga Mencari Dokumen Topik Ketiga	104
Tabel 4.10 Percobaan Keempat Mencari Dokumen Topik Keempat	104
Tabel 4.11 Percobaan Kelima Mencari Dokumen Topik Kelima.....	105
Tabel 4.12 Percobaan Keenam Mencari Dokumen Topik Keenam.....	105
Tabel 4.13 Screenshot Presentase Token Topik	116

DAFTAR GAMBAR

Gambar	Halaman
Gambar 2.1 Proses Text Mining	24
Gambar 2.2 Proses Metode Filtering	27
Gambar 2.3 Proses Metode Wrapping	27
Gambar 2.4 Peta Representasi Grafis Topic Modeling.....	30
Gambar 2.5 Penggambaran dibalik LDA.....	33
Gambar 2.6 Model Grafis Representasi LDA.....	34
Gambar 2.7 Proses Information Retrieval.....	38
Gambar 2.8 Kombinasi Boolean Dalam Diagram Venn.....	40
Gambar 3.1 Model Waterfall	48
Gambar 3.2 Rancangan Basis Data.....	52
Gambar 3.3 Entity Relationship Diagram.....	53
Gambar 3.4 DFD Level 0.....	54
Gambar 3.5 DFD Level 1	55
Gambar 3.6 DFD Level 2 (Proses 1.0).....	56
Gambar 3.7 Flowchart Building Model LDA	58
Gambar 3.8 Flowchart Information Retrieval	60
Gambar 3.9 Representasi Dokumen.....	70
Gambar 3.10 Desain Halaman Dashboard.....	72
Gambar 3.11 Desain Halaman Search	73
Gambar 3.12 Desain Halaman Hasil Search	73
Gambar 4.1 Data Repositori yang Disimpan dalam Table document.....	81
Gambar 4.2 Data repositori setelah preprocessing.....	85
Gambar 4.3 Chart Coherence Score Hasil Tuning Parameter.....	89
Gambar 4.4 <i>Zoom-In Chart Coherence Score Hasil Tuning Parameter</i>	90
Gambar 4.5 Klaster Topik Model Hasil Build dengan pyLDAvis	91
Gambar 4.6 Implementasi Halaman Dashboard	100
Gambar 4.7 Implementasi Halaman Search.....	100
Gambar 4.8 Implementasi Halaman Hasil <i>Search</i>	101
Gambar 4.9 <i>Wordcloud 200 Term</i> Frekuensi Tertinggi.....	111

Gambar 4.10 Hasil Searching “Sistem Laravel” Pada Information Retrieval	114
Gambar 4.11 Hasil Searching "Sistem Laravel" Pada Repo Undiksha	115
Gambar 4.12 Hasil Searching Menggunakan Kueri "game"	118
Gambar 4.13 Hasil Searching Menggunakan Kueri "game game game"	119



DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1 Pengujian <i>Black Box</i>	132
Lampiran 2 Pengujian <i>White Box</i>	133
Lampiran 3 <i>Screenshot</i> Observasi Pencarian Repositori	134
Lampiran 4 Daftar Hasil <i>Tuning</i> Parameter	138
Lampiran 5 Distribusi <i>Term</i> Masing-Masing Topik	146
Lampiran 6 Distribusi Topik Masing-Masing Dokumen.....	151
Lampiran 7 Daftar <i>Stopword (Stoplist)</i> Tambahan	161
Lampiran 8 Metadata yang Disediakan oleh Situs OAI Undiksha	196

