

**PENCARIAN TOPIK PENELITIAN PADA STUDI KASUS JURNAL JIFTI
MENGUNAKAN TEKNIK *HIEARCHICAL DIRICHLET PROCESSES***

SKRIPSI



Oleh:

Moch Erreza

19081010166

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2024**

LEMBAR PENGESAHAN

SKRIPSI

**Judul : PENCARIAN TOPIK PENELITIAN PADA STUDI KASUS JURNAL
JIFTI MENGGUNAKAN TEKNIK *HIEARCHICAL DIRICHLET
PROCESSES***

Oleh : Moch Erreza

NPM : 19081010166

Telah Diseminarkan Dalam Ujian Skripsi Pada :

Hari Rabu, Tanggal 21 Februari 2024

Mengetahui

Dosen Pembimbing

1.

Dr. Ir. Kartini, S.Kom, MT
NIP : 19611110 19903 2 001

Dosen Penguji

1.

Andreas Nugroho Sihananto, S.Kom, M.Kom
NPT : 211199 00 412271

2.

Agung Mustika Rizki, S.Kom, M.Kom
NIP : 19930725 202203 1 008

2.

M. Muharrom Al Haromainy, S.Kom, M.Kom
NIP : 19950601 202203 1 006

Menyetujui

**Dekan
Fakultas Ilmu Komputer**

Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP : 19681126 199403 2 001

**Koordinator Program Studi
Informatika**

Fetty Tri Anggraeny, S.Kom, M.Kom
NIP : 19820211 2021212 005

SURAT PERNYATAAN ANTI PLAGIAT

Saya Mahasiswa Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini :

Nama : Moch Erreza

NPM : 19081010166

Dengan ini menyatakan bahwa judul Skripsi yang saya ajukan dan kerjakan, yang berjudul :

**“PENCARIAN TOPIK PENELITIAN PADA STUDI KASUS JURNAL JIFTI
MENGUNAKAN TEKNIK HIEARCHICAL DIRICHLET PROCESSES”**

Bukan merupakan plagiat dari skripsi atau tugas akhir maupun penelitian orang lain dan juga bukan merupakan produk atau software yang saya beli dari pihak lain. Saya juga menyatakan bahwa skripsi ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun institusi Pendidikan lainnya.

Jika Ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka saya siap menerima konsekuensinya.

Surabaya, 21 Februari 2024

Hormat saya,



Moch Erreza

NPM. 19081010166

PENCARIAN TOPIK PENELITIAN PADA STUDI KASUS JURNAL JIFTI MENGUNAKAN TEKNIK *HIEARCHICAL DIRICHLET PROCESSES*

Nama Mahasiswa : Moch Erreza

NPM : 19081010166

Program Studi : Informatika

Dosen Pembimbing : Dr. Ir. Kartini, S.Kom, MT

Agung Mustika Rizki, S.Kom, M.Kom

ABSTRAK

Studi ini mengajukan pendekatan menggunakan proses Dirichlet hierarkis (HDP), sebuah model Bayesian non-parametrik yang dirancang untuk menangani permasalahan pengelompokan data dengan banyak kelompok. Model ini memandang setiap kelompok data sebagai gabungan dari sejumlah komponen terbuka yang disimpulkan secara otomatis. Pendekatan ini sering digunakan dalam situasi di mana kita mencari hubungan antara cluster, seperti ketika kita mencoba menemukan topik-topik dalam sebuah koleksi dokumen. Eksperimen yang dilakukan pada sejumlah dokumen menunjukkan bahwa penggunaan metode HDP menghasilkan kinerja yang lebih baik dan lebih efisien daripada model yang digunakan sebelumnya.

Fokus dari penelitian ini adalah pada pengembangan model untuk menganalisis topik-topik yang muncul dalam jurnal-jurnal yang berkaitan dengan Sains Teknologi Informasi dan Robotika. Data yang digunakan dalam penelitian ini disusun dari ringkasan jurnal-jurnal Teknologi Informasi dan Robotika yang diterbitkan antara tahun 2019 hingga 2022, yang diperoleh dari situs web resmi jurnal tersebut di <https://jifti.upnjatim.ac.id/index.php/jifti/issue/archive>. Data diambil secara manual menggunakan tool ekstensi Google Chrome.

Metode analisis yang diterapkan dalam penelitian ini meliputi analisis tren, yang mencakup beberapa tahap. Langkah awal melibatkan pengenalan dan pengelompokan topik-topik berdasarkan jenis penelitian yang dilakukan dalam Program Studi Informatika. Kemudian, dilakukan estimasi jumlah dokumen yang terkategori sesuai dengan setiap topik yang telah terbentuk. Dari hasil klasterisasi ini, topik yang menunjukkan tren paling tinggi kemudian diprioritaskan untuk divisualisasikan. Informasi statistik terkait dengan topik-topik yang dibahas dalam jurnal ilmiah di bidang Analisa dan Implementasi.

Penelitian ini dimulai dengan tahap preprocessing, diikuti dengan penggunaan model topik Hierarchical Dirichlet Process (HDP) untuk menganalisis 20 topik dengan memeriksa kata-kata yang paling umum muncul dalam abstrak Jurnal Teknologi Informasi, Teknologi Informasi dan Robotika (JIFTI). Jumlah kemunculan setiap kata menjadi indikator dari metode Hierarchical Dirichlet Process (HDP) yang diterapkan. Temuan utama meliputi sejumlah kata yang sering muncul, seperti “*practicum, test, manage, major, develop, aim, feed, bitching, technology, people, apply, tourism, student, user, feed, learn, digit, laboratory, pusvetma, product*”. Dari 20 kata yang sering muncul dalam setiap topik, dapat

dilihat bahwa mayoritas abstrak jurnal JIFTI menyoroti penelitian dalam implementasi.

Kata Kunci: *Topic modeling, HDP dan Text Preprocessing*

UCAPAN TERIMA KASIH

Dengan segala rasa puji Syukur saya panjatkan kepada Allah SWT, Tuhan Yang Maha Pengasih dan Maha Penyayang, yang dengan limpahan rahmat-Nya, memberikan petunjuk dan kekuatan sehingga saya dapat menyelesaikan skripsi ini. Terima kasih atas anugerah-Nya yang tak terhingga, yang senantiasa mengiringi setiap langkah perjalanan penulisan ini. Pada kesempatan ini, penulis mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT. selaku Rektor Universitas Pembangunan Nasional Veteran Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT, selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom, selaku Koordinator Program Studi Informatika.
4. Ibu Dr. Ir. Kartini, S.Kom, MT selaku Dosen Pembimbing 1 atas kontribusi dan saran yang membangun selama penyusunan skripsi ini.
5. Bapak Agung Mustika Rizki, S.Kom, M.Kom selaku pembimbing 2 yang juga sudah meluangkan waktu dan memberi arahan kepada penulis dalam menyelesaikan skripsi ini.
6. Orang Tua dan Keluarga yang selalu memberikan dukungan moral, doa restu dan motivasi di setiap proses penulisan ini.
7. Tiara Vabella Anggraini karena selalu ada di samping saya, mendengarkan cerita dan turut berkontribusi dalam perjalanan ini. Semua canda, tawa, dan dukungan moral dari Anda telah membuat perjalanan penulis menjadi lebih ringan.
8. Semua Teman-teman yang turut serta dalam mendukung penyelesaian skripsi ini, baik secara langsung maupun tidak langsung.

Akhir kata, semoga dengan adanya laporan ini dapat bermanfaat bagi penulis, pembaca serta memberikan ilmu dan pemikiran yang baru bagi pihak yang membutuhkan.

DAFTAR ISI

LEMBAR PENGESAHAN	i
SURAT PERNYATAAN ANTI PLAGIAT	ii
ABSTRAK	iii
UCAPAN TERIMA KASIH	v
DAFTAR ISI	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Penelitian	5
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA	7
2.1 Tinjauan Penelitian Terdahulu	7
2.2 Bahasa <i>Phyton</i>	14
2.3 <i>Text mining</i>	15
2.4 <i>Topic modeling</i> /Pemodelan Topik	25
2.5 <i>Hierarchical Dirichlet Process</i> (HDP)	26
BAB III METODOLOGI PENELITIAN	31
3.1 Metode Penelitian	31
3.2 Objek Penelitian	31
3.3 Bentuk Data	31
3.4 Sumber Data	32
3.5 Instrumen Penelitian	32
3.6 Desain Penelitian	35
BAB IV HASIL DAN PEMBAHASAN	52
4.1 Pengumpulan Data	52
4.2 <i>Pre-Procesing</i> Data	54
4.2.1 <i>Casefolding</i>	54
4.2.2 Remove Stopwords	58

4.2.3 <i>Filtering Frequent Words</i>	61
4.2.4 <i>Filtering Rare Words</i>	62
4.2.5 <i>Stemming</i>	63
4.2.6 Normalisasi	65
4.2.7 <i>Tokenizing</i>	66
4.2.8 Simpan <i>Dataset Bersih</i>	67
4.3 Hasil <i>Topic modeling Hierarchical Dirichlet Process (HDP)</i>	68
4.4 Pembahasan	81
BAB V PENUTUP	85
5.1 Kesimpulan	85
5.2 Saran	86
DAFTAR PUSTAKA	87

DAFTAR GAMBAR

Gambar 2.1.	Graphical model for HDP.....	28
Gambar 3.1.	Pencampuran Topik.....	50
Gambar 4.1.	Halaman Utama JIFTI UPN Jatim	52
Gambar 4.2.	Tampilan Data yang Telah Dikumpulkan	53
Gambar 4.3	Tahap <i>Casefolding</i>	57
Gambar 4.4	Remove Stopwords.....	60
Gambar 4.5	<i>Filtering Frequent Words</i>	61
Gambar 4.6	<i>Filtering Rare Words</i>	62
Gambar 4.7	<i>Stemming</i>	64
Gambar 4.8	Normalisasi.....	65
Gambar 4.9	<i>Tokenizing</i>	66
Gambar 4.10	Simpan <i>Dataset</i> Bersih.....	67
Gambar 4.11.	Hasil <i>Topic Modeling</i>	70
Gambar 4.12.	Representasi Grafis HDP.....	72
Gambar 4.13.	Topik HDP.....	74
Gambar 4.14.	Word Cloud	79
Gambar 4.15.	Persentase Perbandingan Jenis Topik.....	80

DAFTAR TABEL

Tabel 2.1.	Tinjauan Penelitian Tedahulu	7
Tabel 3.1.	Contoh <i>Casefolding</i>	37
Tabel 3.2.	Contoh <i>Remove Stopword</i>	38
Tabel 3.3.	Contoh <i>Filtering Frequent Words</i>	39
Tabel 3.4.	Contoh <i>Filtering Rare Words</i>	40
Tabel 3.5.	Contoh <i>Stemming</i>	42
Tabel 3.6.	Contoh Normalisasi	43
Tabel 3.7.	Contoh Tokenisasi	44
Tabel 3.8.	Contoh Simpanan <i>Dataset</i> Ber	45