

**PEMODELAN TOPIK DOKUMEN SKRIPSI MAHASISWA S1 TEKNIK
INFORMATIKA UIN SUNAN KALIJAGA YOGYAKARTA
MENGUNAKAN ANALISIS LATENT DIRICHLET ALLOCATION**

Skripsi

untuk memenuhi sebagian persyaratan

mencapai derajat Sarjana S-1

Program Studi Teknik Informatika



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Disusun oleh:

Achmad Ibrahim Humam

17106050029

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2022

HALAMAN PEGESAHAN



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-579/Un.02/DST/PP.00.9/03/2022

Tugas Akhir dengan judul : PEMODELAN TOPIK DOKUMEN SKRIPSI MAHASISWA S1 TEKNIK INFORMATIKA UIN SUNAN KALIJAGA YOGYAKARTA MENGGUNAKAN ANALISIS LATENT DIRICHLET ALLOCATION

yang dipersiapkan dan disusun oleh:

Nama : ACHMAD IBRAHIM HUMAM
Nomor Induk Mahasiswa : 17106050029
Telah diujikan pada : Selasa, 22 Februari 2022
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Dr. Ir. Bambang Sugiantoro, S.Si., M.T.
SIGNED

Valid ID: 6231a3f582415



Penguji I

Muhammad Didik Rohmad Wahyudi, S.T.,
MT.
SIGNED

Valid ID: 6230481da2968



Penguji II

Agus Mulyanto, S.Si., M.Kom.
SIGNED

Valid ID: 62318dc2bd1f7



Yogyakarta, 22 Februari 2022
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6231dad617132

SURAT PERSETUJUAN SKRIPSI



Universitas Islam Negeri Sunan Kalijaga



FM-UINSK-BM-05-03/R0

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi

Lamp :

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Achmad Ibrahim Humam
NIM : 17106050029
Judul Skripsi : Analisis Latent Dirichlet Allocation Untuk Pemodelan Topik
Dokumen Skripsi (Studi Kasus: Skripsi Mahasiswa S1 Teknik
Informatika UIN Sunan Kalijaga Yogyakarta)

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika.

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 6 Februari 2022
Pembimbing

Dr. Ir. Bambang Sugiantoro, S.Si., MT
NIP. 19751024 200912 1 002

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini:

Nama : Achmad Ibrahim Humam

NIM : 17106050029

Program Studi : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul “**Analisis Latent Dirichlet Allocation Untuk Pemodelan Topik Dokumen Skripsi (Studi Kasus: Skripsi Mahasiswa S1 Teknik Informatika UIN Sunan Kalijaga Yogyakarta)**” merupakan hasil penelitian saya sendiri, tidak terdapat pada karya yang pernah di ajukan untuk memperoleh gelar kesarjana di suatu perguruan tinggi, dan bukan plagiasi karya orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 6 Februari 2022

Yang menyatakan



Achmad Ibrahim Humam

NIM. 17106050029

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Alhamdulillahirabbil'alamin, segala puji dan syukur kehadiran Allah SWT yang selalu memberikan berkah, rahmat dan hidayah-Nya berupa kesabaran, kekuatan, serta kelancaran dalam penulisan skripsi dengan judul “Analisis Latent Dirichlet Allocation Untuk Pemodelan Topik Dokumen Skripsi (Studi Kasus: Skripsi Mahasiswa S1 Teknik Informatika UIN Sunan Kalijaga Yogyakarta)” dari awal hingga terselesaikan penulisan ini. Sholawat serta salam tak lupa dihaturkan kepada Nabi Muhammad SAW, beserta keluarga, sahabat dan umatnya.

Penulis juga mengucapkan terimakasih kepada semua pihak yang telah membantu dalam proses pelaksanaan tugas akhir ini sehingga laporan tugas akhir ini dapat terselesaikan dengan baik. Penulis menyadari bahwa penulisan skripsi ini tidak bisa terwujud tanpa adanya bantuan, bimbingan dan dorongan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati pada kesempatan ini penulis mengucapkan banyak terima kasih kepada:

1. Bapak Prof. Dr. Phil. Al Makin, S.Ag., M.A., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Dr. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Ibu Maria Ulfah Siregar, S.Kom. MIT., Ph.D, selaku Ketua Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta serta sebagai Dosen Pembimbing Akademik.

4. Bapak Dr. Ir. Bambang Sugiantoro, S.Si., M.T., selaku Dosen Pembimbing Skripsi yang telah sabar membimbing dalam menyusun dan menyelesaikan skripsi ini.
5. Seluruh Dosen Program Studi Teknik Informatika UIN Sunan Kalijaga yang telah membagikan banyak ilmunya kepada penulis.
6. Seluruh pegawai kantor Admisi UIN Sunan Kalijaga Yogyakarta yang telah memberikan izin dan dukungan kepada penulis dalam melaksanakan penelitian.
7. Orang tua penulis yang tak henti-hentinya memberikan dukungan, do'a, dan kasih sayangnya kepada penulis.
8. Kakak dan keluarga penulis yang selalu memberi dorongan, semangat, dan motivasi kepada penulis.
9. Seluruh anggota Cheese Cake yang telah memberikan bantuan dan dukungan moral kepada penulis.
10. Seluruh pasukan Kost Biru Pengkolan Sapen dan Kost Nogopuro yang telah memberikan bantuan dan dukungan moral kepada penulis.
11. Seluruh pasukan ngopi malam yang telah memberikan bantuan dan dukungan moral kepada penulis.
12. Teman-teman Teknik Informatika 2017 yang telah memberikan dukungan dan semangat kepada penulis.
13. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu dalam pelaksanaan penyusunan skripsi.

Penulis menyadari dalam penulisan skripsi ini masih banyak kekurangan dan jauh dari sempurna, oleh karena itu penulis memohon maaf serta menerima

segala saran dan kritik yang membangun dari para pembaca. Akhir kata, semoga penulisan skripsi ini dapat menjadi panduan serta referensi yang berguna bagi pembaca dan dapat dimanfaatkan sebaik-baiknya.

Wassalamualaikum Wr.Wb.

Yogyakarta, Februari 2022

Penulis



Achmad Ibrahim Humam
17106050029



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan untuk kedua orang tua saya yang senantiasa menjaga, menyayangi, membimbing, dan mendoakan saya sampai sejauh ini.

Semoga Bapak bahagia disana dan selalu dilimpahkan rahmat, ampunan, syafa'at, dan dilancarkan segala amal jariahnya. Semoga Ibu selalu dilimpahkan kesehatan, kebahagiaan, kemudahan rezeki, istiqomah dalam ibadah kepada Allah hingga akhir usia diberikan khusnul khotimah.



HALAMAN MOTO

“Hidup ini seperti pensil yang lama-lama akan habis, tetapi akan meninggalkan tulisan-tulisan yang indah dalam kehidupan”

-Nami-



DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PEGESAHAN.....	ii
SURAT PERSETUJUAN SKRIPSI	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
KATA PENGANTAR	v
HALAMAN PERSEMBAHAN	viii
HALAMAN MOTO	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xv
DAFTAR PERSAMAAN	xvi
DAFTAR KODE.....	xvii
INTISARI.....	xviii
ABSTRACT.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Keaslian Penelitian	4
1.7 Sistematika Penulisan.....	5
BAB II TINJUAN PUSTAKA DAN LANDASAN TEORI	6

2.1	Tinjauan Pustaka	6
2.2	Landasan Teori	13
2.2.1	Skripsi	13
2.2.2	Text Mining.....	13
2.2.3	Topic Modeling.....	13
2.2.4	Preprocessing	15
2.2.4.1	Case Folding	15
2.2.4.2	Punctuation Removal.....	15
2.2.4.3	Tokenization	16
2.2.4.4	Lemmatization	16
2.2.4.5	Stopword Removal	16
2.2.5	N-Gram Model	16
2.2.6	Latent Dirichlet Allocation (LDA).....	16
2.2.7	Topic Coherence	19
2.2.8	Python	22
2.2.9	Gensim	22
2.2.10	pyLDAvis.....	23
BAB III METODE PENELITIAN.....		24
3.1	Studi Pustaka	24
3.2	Pengumpulan Data	24
3.3	Preprocessing.....	25
3.4	Membentuk Model N-Gram	25
3.5	Membentuk Input Model.....	25
3.6	Implementasi Pemodelan Topik.....	26
3.7	Hyperparameter Tuning	26

3.8	Membentuk Model Optimal	27
3.9	Visualisasi Hasil	27
3.10	Kebutuhan Sistem	27
BAB IV HASIL DAN PEMBAHASAN		28
4.1	Persiapan Data	28
4.2	Preprocessing.....	31
4.2.1	Case Folding	31
4.2.2	Punctuation Removal	34
4.2.3	Lemmatization.....	38
4.2.4	Stopword Removal.....	41
4.3	Membentuk Model N-Gram.....	45
4.4	Analisis Pemodelan Topik.....	48
4.4.1	Membentuk Input Model.....	48
4.4.2	Implementasi Pemodelan Topik.....	50
4.4.3	Hyperparameter Tuning	53
4.4.3.1	Mencari Model Topik Optimal	53
4.4.3.2	Memilih Model Topik Optimal	57
4.5	Analisis Model Topik Optimal.....	62
4.5.1	Visualisasi	63
4.6	Pengujian Model.....	72
BAB V PENUTUP.....		75
5.1	Kesimpulan.....	75
5.2	Saran	76
DAFTAR PUSTAKA		77
LAMPIRAN.....		79

CURRICULUM VITAE	108
------------------------	-----



DAFTAR TABEL

Tabel 2. 1 Perbandingan Tinjauan Pustaka	10
Tabel 4. 1 Contoh Teks Abstrak	29
Tabel 4. 2 Contoh Penerapan <i>Case Folding</i>	31
Tabel 4. 3 Contoh Penerapan <i>Punctuation Removal</i>	35
Tabel 4. 4 Contoh Penerapan <i>Lemmatization</i>	39
Tabel 4. 5 Contoh Penerapan <i>Stopword Removal</i>	43
Tabel 4. 6 Contoh Hasil Pembentukan <i>N-Gram</i>	46
Tabel 4. 7 Contoh Hasil <i>Corpus</i>	49
Tabel 4. 8 Model dengan jumlah topik 6	59
Tabel 4. 9 Model dengan jumlah topik 11	59
Tabel 4. 10 Model dengan jumlah topik 16	60

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Topic Modeling	14
Gambar 2. 2 Model Representasi LDA.....	17
Gambar 2. 3 Proses Perhitungan <i>topic coherence</i>	21
Gambar 2. 4 Perhitungan <i>coherence score</i> terkuat	21
Gambar 3. 1 Tahapan Penelitian	24
Gambar 4. 1 <i>Flowchart</i> Proses Penelitian.....	28
Gambar 4. 2 Model Awal.....	51
Gambar 4. 3 Gambaran Distribusi Topik dan Kata dalam Dokumen	52
Gambar 4. 4 Gambaran Segmentasi Kata dalam Perhitungan <i>Coherence C_v</i>	52
Gambar 4. 5 Data Model Optimal Per- <i>K</i>	56
Gambar 4. 6 Perbandingan <i>Coherence Score</i> dari Model Optimal Per- <i>K</i>	57
Gambar 4. 7 Perbandingan <i>Coherence Score</i> Berdasarkan Parameter Distribusi	62
Gambar 4. 8 Model Optimal	63
Gambar 4. 9 Topik-1	64
Gambar 4. 10 <i>Wordcloud</i> Topik-1	65
Gambar 4. 11 Topik-2	65
Gambar 4. 12 <i>Wordcloud</i> Topik-2	66
Gambar 4. 13 Topik-3.....	66
Gambar 4. 14 <i>Wordcloud</i> Topik-3	67
Gambar 4. 15 Topik-4	67
Gambar 4. 16 <i>Wordcloud</i> Topik-4	68
Gambar 4. 17 Topik-5	68
Gambar 4. 18 <i>Wordcloud</i> Topik-5	69
Gambar 4. 19 Topik-6	69
Gambar 4. 20 <i>Wordcloud</i> Topik-6	70
Gambar 4. 21 Perbandingan Jumlah Dokumen Berdasarkan Topik	72
Gambar 4. 22 <i>Confusion Matrix</i>	74
Gambar 4. 23 Pengujian Model	74

DAFTAR PERSAMAAN

Persamaan 2. 1 Inisialisasi Parameter	18
Persamaan 2. 2 Probabilitas Distribusi Topik Sebuah Dokumen	19
Persamaan 2. 3 Probabilitas Distribusi Topik	19



DAFTAR KODE

Kode 4. 1 Memuat Data	29
Kode 4. 2 <i>Case Folding</i>	31
Kode 4. 3 <i>Punctuation Removal</i>	35
Kode 4. 4 <i>Lemmatization</i>	38
Kode 4. 5 Inisialisasi Daftar <i>Stopword</i>	42
Kode 4. 6 <i>Stopword Removal</i>	43
Kode 4. 7 Membentuk Model <i>N-Gram</i>	46
Kode 4. 8 Membentuk <i>Input Model</i>	49
Kode 4. 9 Membentuk Model Awal.....	50
Kode 4. 10 Mencari Model Topik Optimal.....	55
Kode 4. 11 Model Optimal.....	63
Kode 4. 12 Visualisasi Hasil	64

**PEMODELAN TOPIK DOKUMEN SKRIPSI MAHASISWA S1 TEKNIK
INFORMATIKA UIN SUNAN KALIJAGA YOGYAKARTA
MENGUNAKAN ANALISIS LATENT DIRICHLET ALLOCATION**

Achmad Ibrahim Humam
17106050029

INTISARI

Setiap tahun, pendidikan tinggi meluluskan banyak mahasiswa dengan penelitian dan tugas akhir yang beragam sebagai syarat atas kelulusannya. Hal ini menyebabkan jumlah laporan tugas akhir selalu bertambah. Bertambahnya laporan tugas akhir ini menumbuhkan peluang semakin banyak mahasiswa yang mengambil penelitian dengan tema, objek, dan metode penelitian yang mirip atau hampir sama. Meski demikian, banyaknya laporan tugas akhir terdahulu juga menciptakan banyak sumber referensi untuk penelitian yang akan datang, tetapi tidak jarang mahasiswa masih kebingungan dalam menentukan topik penelitian yang akan dibahas. Oleh sebab itu, perlu adanya penelitian untuk mengetahui topik apa saja yang sudah sering dibahas di tugas akhir terdahulu untuk memudahkan mahasiswa dalam memilih topik tugas akhir yang akan dibahas dan membantu dosen pembimbing skripsi dalam membimbing dan mengarahkan proses penyusunan skripsi oleh mahasiswa serta membantu upaya terciptanya topik pembahasan skripsi yang semakin beragam.

Penelitian ini bertujuan untuk melakukan pemodelan topik dokumen skripsi mahasiswa dengan melakukan analisis *Latent Dirichlet Allocation* (LDA) untuk mencari tahu topik apa saja yang sering dibahas dalam skripsi mahasiswa S1 Teknik Informatika UIN Sunan Kalijaga Yogyakarta dari tahun 2010-2021 dengan data sebanyak 666 dokumen.

Proses evaluasi menghasilkan model optimal dengan jumlah topik sebanyak 6. Enam topik tersebut adalah “Analisis dan Perancangan Sistem”, “Data Mining”, “Jaringan Komputer”, “Sistem Pendukung Keputusan”, “Pengujian Perangkat Lunak”, dan “Keamanan Komputer”. Dari hasil perbandingan persentase kata-kata pada setiap topik, topik “Analisis dan Perancangan Sistem” mendominasi pembahasan skripsi mahasiswa S1 Teknik Informatika UIN Sunan Kalijaga Yogyakarta.

Kata Kunci: *text mining*, *topic modeling*, LDA, skripsi

**TOPIC MODELING OF THE UNDERGRADUATE THESIS DOCUMENT
OF INFORMATICS ENGINEERING STUDENTS UIN SUNAN
KALIJAGA YOGYAKARTA USING DIRICHLET LATENT
ALLOCATION ANALYSIS**

Achmad Ibrahim Humam
17106050029

ABSTRACT

Every year, many universities graduate students with various researches and final projects as graduation requirements. This causes the number of final project reports to always increase. This increase in final project reports increases opportunities for more students to take up research with similar or almost the same themes, objects, and research methods. However, many previous final project reports also create many sources of reference for further research, but it is not uncommon for students to still be confused in determining the research topic to be discussed. Therefore, there is a need for research to find out what topics are often discussed in previous final projects to make it easier for students to choose the final project topics to be discussed and assist thesis supervisors in guiding and directing the thesis writing process by students and helping efforts to create increasingly diverse thesis discussion topics.

This research aims to do topic modeling to the undergraduate thesis documents by analyzing Latent Dirichlet Allocation (LDA) to find out what topics are often discussed in the undergraduate thesis of Informatics Engineering UIN Sunan Kalijaga Yogyakarta in 2010-2021 with data of 666 documents.

The six topics are "System Analysis and Design", "Data Mining", "Computer Networks", "Decision Support Systems", "Software Testing", and "Computer Security". From the results of the comparison of the percentage of words in each topic, the topic "System Analysis and Design" dominates the undergraduate thesis discussion of Informatics Engineering students at UIN Sunan Kalijaga Yogyakarta.

Keywords: text mining, topic modeling, LDA, thesis

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan merupakan salah satu bagian penting dan tidak bisa dipisahkan dalam perkembangan kehidupan manusia. Pendidikan merupakan upaya terencana dalam proses pembimbingan dan pembelajaran bagi individu agar berkembang dan tumbuh menjadi manusia yang mandiri, bertanggungjawab, kreatif, berilmu, sehat, dan berakhlak mulia baik dilihat dari aspek jasmani maupun rohani. Pendidikan terdiri atas pendidikan informal, non formal, dan formal. Semua model pendidikan ini penting dalam membentuk kepribadian seseorang. Salah satunya pendidikan formal, yaitu jalur pendidikan yang terstruktur dan berjenjang yang terdiri atas pendidikan dasar, pendidikan menengah, dan pendidikan tinggi (Ilma, 2015).

Setiap tahun, jenjang pendidikan tinggi meluluskan banyak mahasiswa dengan penelitian dan tugas akhir yang beragam sebagai syarat atas kelulusannya. Hal ini menyebabkan jumlah laporan skripsi selalu bertambah. Bertambahnya penelitian skripsi ini menumbuhkan peluang semakin banyaknya mahasiswa yang mengambil penelitian dengan tema, objek, dan metode penelitian yang mirip atau hampir sama (Sholehhudin et al., 2018). Meski demikian, banyaknya penelitian skripsi terdahulu juga menciptakan banyak sumber referensi untuk penelitian yang akan datang, tetapi tidak jarang mahasiswa masih kebingungan dalam menentukan topik penelitian yang akan dibahas. Untuk mengatasi hal ini dan membantu dosen pembimbing skripsi dalam membimbing dan mengarahkan proses penyusunan

skripsi oleh mahasiswa serta membantu upaya terciptanya topik pembahasan skripsi yang semakin beragam, perlu adanya identifikasi topik apa saja yang sudah sering dibahas dalam skripsi terdahulu.

Dengan melakukan *text mining* menggunakan metode pemodelan topik (*topic modeling*) pada kumpulan dokumen penelitian terdahulu, bisa didapatkan kumpulan topik apa saja yang biasa digunakan dalam penelitian terdahulu. *Text mining* merupakan variasi dari bidang *data mining*, perbedaannya adalah sumber data *text mining* merupakan teks dengan bentuk yang tidak terstruktur berbeda dengan *data mining* yang mengolah data terstruktur seperti dalam bentuk database (Hearst, 1999). Salah satu metode dalam *text mining* adalah pemodelan topik, metode ini adalah kumpulan algoritma yang digunakan untuk menemukan struktur tersembunyi dari topik yang terdapat dalam setiap dokumen (Blei et al., 2003).

Dalam mengidentifikasi isi dokumen penelitian secara cepat dan akurat, abstrak menjadi bagian paling penting karena abstrak merupakan sintesis keseluruhan isi dokumen penelitian. Abstrak mengandung konsep, pernyataan masalah, pendekatan, dan kesimpulan yang dirangkai sedemikian rupa saling berkaitan dan memiliki makna utuh sehingga menggambarkan keseluruhan isi tulisan (Nasution, 2017).

Data yang diidentifikasi adalah abstrak dokumen skripsi mahasiswa S1 Teknik Informatika UIN Sunan Kalijaga Yogyakarta. Dokumen berasal dari *Digital Library* UIN Sunan Kalijaga yang bisa diakses melalui <https://digilib.uin-suka.ac.id/>. Dengan melakukan analisis *Latent Dirichlet Allocation* (LDA) untuk melakukan pemodelan topik dokumen skripsi, diharapkan dapat menghasilkan

kumpulan topik yang biasa digunakan dalam skripsi mahasiswa S1 Teknik Informatika UIN Sunan Kalijaga Yogyakarta. Dengan hasil analisis tersebut diharapkan dapat menjadi bahan pertimbangan bagi mahasiswa dan dosen pembimbing skripsi dalam menentukan topik apa yang akan dibahas dalam proses penyusunan skripsi yang akan datang.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, maka rumusan masalah dalam penelitian ini adalah belum adanya pemodelan topik dokumen skripsi mahasiswa dengan melakukan analisis *Latent Dirichlet Allocation* (LDA) untuk mencari tahu topik apa saja yang sering dibahas dalam skripsi mahasiswa S1 Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta.

1.3 Batasan Masalah

Agar penyusunan dan pembahasan penelitian lebih terarah, diperlukan batasan masalah dalam penelitian ini. Berikut adalah batasan masalah yang digunakan dalam penelitian ini:

1. Pemodelan topik menggunakan algoritma *Latent Dirichlet Allocation* (LDA).
2. Data yang digunakan merupakan data yang diperoleh dari website *Digital Library* UIN Sunan Kalijaga.
3. Data yang digunakan merupakan dokumen skripsi mahasiswa S1 Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta dari tahun 2010-2021 sebanyak 666 data.
4. Data yang digunakan berupa data teks abstrak skripsi.

5. Data diolah menggunakan bahasa pemrograman *Python* di *environment Anaconda* versi 1.9.7.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah melakukan pemodelan topik dokumen skripsi mahasiswa dengan melakukan analisis *Latent Dirichlet Allocation* (LDA) untuk mencari tahu topik apa saja yang sering dibahas dalam skripsi mahasiswa S1 Program Studi Teknik Informatika UIN Sunan Kalijaga.

1.5 Manfaat Penelitian

Berdasarkan dari penjelasan di atas, manfaat dari penelitian ini adalah sebagai berikut:

1. Memberikan hasil analisis dari pemodelan topik berupa topik apa yang sering dibahas oleh mahasiswa S1 Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta dalam menyusun skripsi.
2. Menjadi bahan pertimbangan bagi mahasiswa maupun dosen pembimbing skripsi dalam menentukan topik yang akan digunakan dalam penyusunan skripsi di masa yang akan datang, baik itu sebagai sumber referensi penelitian skripsi selanjutnya, ataupun hanya sebagai pembanding untuk mencari topik pembahasan yang jarang atau belum pernah dilakukan sebelumnya.

1.6 Keaslian Penelitian

Penelitian tentang *text mining* sudah banyak dilakukan sebelumnya. Namun, penelitian mengenai pemodelan topik yang diajukan sebagai Skripsi masih jarang dilakukan dan penerapan pemodelan topik dengan algoritma *Latent Dirichlet Allocation* (LDA) pada data abstrak skripsi mahasiswa Program Studi Teknik

Informatika UIN Sunan Kalijaga belum pernah dilakukan. Hal ini diketahui berdasarkan dari referensi dan tinjauan pustaka yang telah dilakukan penulis.

1.7 Sistematika Penulisan

Penyusunan penelitian tugas akhir ini disusun secara sistematis dalam 5 bab, dimulai dari BAB I sampai BAB V. Berikut penjelasan pada masing-masing bab:

BAB I PENDAHULUAN

Bab ini berisi tentang penjelasan latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian, dan sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab ini berisi penjelasan tentang tinjauan pustaka dan landasan teori yang berkaitan dengan penelitian ini.

BAB III METODE PENELITIAN

Bab ini berisi penjelasan tentang tahap-tahap atau metode yang digunakan untuk mencapai tujuan dan memperoleh hasil penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil penerapan metode yang digunakan dan analisis hasil penelitian.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan dari hasil penelitian dan saran berupa kekurangan yang perlu diperbaiki dalam penelitian ini.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa pemodelan topik dokumen skripsi mahasiswa dengan melakukan analisis *Latent Dirichlet Allocation (LDA)* dapat mencari tahu topik apa saja yang sering dibahas dalam skripsi mahasiswa S1 Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta. Pada penelitian ini dilakukan analisis *Latent Dirichlet Allocation (LDA)* dengan rentang topik sebanyak 2-21 serta nilai α dan β dengan rentang 0.01-1 lalu dicari model paling optimal dengan *coherence score* paling tinggi, tetapi dengan pertimbangan bahwa jumlah topik tidak boleh terlalu banyak untuk menghindari *over-clustering* dan hasil pemodelan topik berupa kumpulan kata-kata dalam suatu topik bisa saling koheren.

Percobaan yang telah dilakukan menghasilkan model optimal dengan jumlah topik sebanyak 6. Dari enam topik yang dihasilkan dapat disimpulkan bahwa topik-1 membahas “Analisis dan Perancangan Sistem”, topik-2 membahas “Data Mining”, topik-3 membahas “Jaringan Komputer”, topik-4 membahas “Sistem Pendukung Keputusan”, topik-5 membahas “Pengujian Perangkat Lunak”, dan topik-6 membahas “Keamanan Komputer”. Hasil topik tersebut diambil berdasarkan data statistik kata-kata yang sering muncul pada topik-topik tersebut. Dari hasil perbandingan jumlah dokumen pada setiap topik, topik “Analisis dan

Perancangan Sistem” mendominasi pembahasan skripsi mahasiswa S1 Teknik Informatika UIN Sunan Kalijaga Yogyakarta.

5.2 Saran

Pada penelitian ini masih banyak sekali kekurangan. Maka dari itu penulis menyarankan beberapa hal untuk penelitian selanjutnya, diantaranya:

1. Jumlah data yang digunakan pada penelitian berikutnya bisa lebih banyak lagi.
2. Penelitian selanjutnya dapat melakukan penyempurnaan proses *preprocessing*, yaitu diantaranya proses perbaikan *slangword* atau kata yang penulisannya tidak tepat, sehingga penulisan kata bisa menjadi seragam dan hasil analisis bisa menjadi lebih baik.
3. Penelitian pemodelan topik dokumen skripsi selanjutnya bisa menggunakan sumber data berbahasa Indonesia.
4. Penelitian pemodelan topik dokumen skripsi selanjutnya sebaiknya tidak hanya menggunakan bagian abstrak skripsi tetapi menggunakan keseluruhan dokumen agar mendapatkan penjelasan yang lebih lengkap mengenai isi dokumen tersebut.

DAFTAR PUSTAKA

- Agustina, A. (2017). *Analisis dan Visualisasi Suara Pelanggan Pada Pusat Layanan Pelanggan Dengan Pemodelan Topik Menggunakan Latent Dirichlet Allocation (LDA) Studi Kasus: Pt. Petrokimia Gresik*. Institut Teknologi Sepuluh Nopember.
- Anggraini, E. (2020). *Latent Dirichlet Allocation untuk Pemodelan Topik Abstrak Dokumen Skripsi (Studi Kasus: Abstrak Dokumen Skripsi Mahasiswa Statistika UII Tahun Angkatan 2011-2015)*. Universitas Islam Indonesia.
- Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the Acm*, 55(4), 77–84.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Machine Learning Research*, 3, 993–1022.
- Campbell, J. C., Hindle, A., & Stroulia, E. (2014). *Latent Dirichlet Allocation : Extracting Topics from Software Engineering Data*. 1–21.
- Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Neural Information Processing Systems*, 32, 288–296.
- Dicoding. (2021). *Memulai Pemrograman Dengan Python*. Dicoding. <https://www.dicoding.com/academies/86>
- Filcha, A., & Hayaty, M. (2019). *Implementasi Algoritma Rabin-Karp untuk Pendeteksi Plagiarisme pada Dokumen Tugas Mahasiswa (Rabin-Karp Algorithm Implementation to Detect Plagiarism o n Student ' s Assignment Document)*. VII(1), 25–32.
- gensim 4.1.2. (2021). PyPI. <https://pypi.org/project/gensim/>
- Greene, D., O’Callaghan, D., & Cunningham, P. (2014). How many topics? Stability analysis for topic models. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8724 LNAI(PART 1), 498–513. https://doi.org/10.1007/978-3-662-44848-9_32
- Hearst, M. A. (1999). What Is Text Mining? In *Proceedings of ACL’99: the 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland*.
- Ilma, N. (2015). Peran Pendidikan Sebagai Modal Utama Membangun Karakter Bangsa. *TADBIR Jurnal Manajemen Pendidikan Islam*, 3(1), 82–87.
- Jaka, A. T. (2015). Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining. *Jurnal Informatika Upgris*, 1(1), 1–9.
- Kapadia, S. (2019). *Language Models: N-Gram*. Towards Data Science. <https://towardsdatascience.com/introduction-to-language-models-n-gram-e323081503d9>
- Miftahuddin, Y. (2018). *Penerapan Algoritma Lemmatization pada Dokumen Bahasa Indonesia*. 3(2), 47–56.
- Nasution, M. K. M. (2017). *Abstrak - Suatu Karya Ilmiah*. 3, 2–5. <https://doi.org/10.13140/RG.2.2.14333.26086>

- Putra, I. M. K. B. (2017). *Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA)*. Institut Teknologi Sepuluh Nopember.
- pyLDAvis 3.3.1. (2021). PyPI. <https://pypi.org/project/pyLDAvis/>
- Python – The new generation Language. (2021). GeeksforGeeks. <https://www.geeksforgeeks.org/python-the-new-generation-language/>
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, 399–408. <https://doi.org/10.1145/2684822.2685324>
- Ruaika, D. (2019). *K-Means Clustering Dan Naïve Bayes Classifier Untuk Kategorisasi Dokumen Tugas Akhir (Studi Kasus: Skripsi SI Teknik Informatika UIN Sunan Kalijaga Yogyakarta)*. Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
- Sanjaya, N. A. (2021). *Implementasi Latent Dirichlet Allocation (LDA) Untuk Klasterisasi Cerita Berbahasa Bali*. 8(1), 127–134. <https://doi.org/10.25126/jtiik.202183556>
- Septiawan, A. (2020). *Analisis Metode Latent Dirichlet Allocation (LDA) untuk Clustering Terjemahan Al-Qur'an Bahasa Indonesia*. Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
- Sholehuddin, M., Fauzi, M. A., & Adinugroho, S. (2018). *Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi (Studi Kasus : Universitas Brawijaya)*. 2(11), 5518–5524.
- Tan, A., Ridge, K., Labs, D., & Terrace, H. (2000). *Text Mining : The state of the art and the challenges*.