

**PENENTUAN TREN TOPIK TUGAS AKHIR
MENGUNAKAN METODE *LATENT DIRICHLET ALLOCATION*
(LDA) DENGAN VISUALISASI *STREAMGRAPH***



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
pada Departemen Ilmu Komputer/ Informatika**

Disusun oleh :

MUHAMMAD FIKRI MUKHLISHIN

24010312120017

**DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO**

2019

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Muhammad Fikri Mukhlishin
NIM : 24010312120017
Judul : Penentuan Tren Topik Tugas Akhir Menggunakan Metode *Latent Dirichlet Allocation (LDA)* dengan Visualisasi *Streamgraph*

Dengan ini saya menyatakan bahwa dalam tugas akhir/ skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.



Semarang, 28 Februari 2019

Muhammad Fikri Mukhlishin
NIM. 24010312120017

HALAMAN PENGESAHAN

Judul : Penentuan Tren Topik Tugas Akhir Menggunakan Metode *Latent Dirichlet Allocation* (LDA) dengan Visualisasi *Streamgraph*

Nama : Muhammad Fikri Mukhlishin

NIM : 24010312120017

Telah diajukan pada sidang tugas akhir pada tanggal 28 Februari 2019 dan dinyatakan lulus pada tanggal 28 Februari 2019.

Mengetahui,
Ketua Departemen Ilmu Komputer/ Informatika



Dr. Retno Kusumaningrum, S.Si, M.Kom
NIP. 198104202005012001

Semarang, 28 Februari 2019

Panitia Penguji Tugas Akhir
Ketua,

A handwritten signature in black ink, appearing to be 'Priyo Sidik S.', written over a faint background.

Priyo Sidik S, S.Si.,M.Kom.
NIP. 197007051997021001

HALAMAN PENGESAHAN

Judul : Penentuan Tren Topik Tugas Akhir Menggunakan Metode *Latent Dirichlet Allocation* (LDA) dengan Visualisasi *Streamgraph*

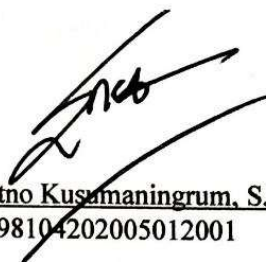
Nama : Muhammad Fikri Mukhlisin

NIM : 24010312120017

Telah diujikan pada sidang tugas akhir pada tanggal 28 Februari 2019.

Semarang, 28 Februari 2019

Dosen Pembimbing



Dr. Retno Kusumaningrum, S.Si, M.Kom
NIP. 198104202005012001

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Fikri Mukhlisin
NIM : 24010312120017
Program Studi : Informatika
Departemen : Ilmu Komputer/ Informatika
Fakultas : Sains dan Matematika
Jenis Karya : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan **Hak Bebas Royalti Noneksklusif** (*Non-exclusive royalty Free Right*) kepada Universitas Diponegoro atas karya ilmiah saya yang berjudul:

Penentuan Tren Topik Tugas Akhir Menggunakan Metode Latent Dirichlet Allocation (LDA) dengan Visualisasi Streamgraph

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini Universitas Diponegoro berhak menyimpan, mengalih media/ formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 28 Februari 2019



Yang menyatakan

Muhammad Fikri Mukhlisin

24010312120017

ABSTRAK

Tugas akhir adalah salah satu karya ilmiah yang ditulis oleh mahasiswa pada akhir masa studinya yang merupakan salah satu syarat kelulusan untuk seorang mahasiswa. Karya ilmiah tersebut disusun menurut kaidah keilmuan dan ditulis berdasarkan kaidah Bahasa Indonesia di bawah bimbingan dan arahan dosen pembimbing. Mahasiswa sering kebingungan dalam pemilihan topik yang akan dikerjakan sehingga menambah waktu pengerjaan tugas akhirnya. Mengetahui tren topik akan memudahkan untuk mencari topik yang memiliki banyak referensi. Pada penelitian ini, diusulkan menggunakan metode *Latent Dirichlet Allocation* (LDA) untuk melakukan pengelompokan (*clustering*) terhadap abstrak dari tugas akhir dalam bahasa Indonesia. LDA akan melakukan *clustering* terhadap abstrak tugas akhir berdasarkan tingkat kemiripan kata-kata yang ada pada abstrak tersebut ke dalam sejumlah topik. LDA akan menghasilkan dua buah keluaran yaitu, *topic proportion* dan probabilitas kata-topik. Visualisasi diperlukan untuk dapat melihat lebih jelas hasil *clustering* abstrak tugas akhir. Salah satu visualisasi yang dapat menampilkan hasil *clustering* adalah dalam bentuk *Streamgraph*. Tujuan dari penelitian ini adalah untuk menghasilkan *Streamgraph* yang menampilkan hasil *clustering* dari abstrak tugas akhir dengan menggunakan kombinasi parameter LDA yang terbaik. Kombinasi parameter terbaik dalam kasus ini adalah dengan menggunakan nilai *alpha* 0,1, *beta* 0,1, dan *threshold* 0,001, sedangkan jumlah topik ditentukan berdasarkan kata unik di dalam kumpulan abstrak, sehingga mendapatkan nilai sebesar 13. Kombinasi parameter tersebut menghasilkan nilai *perplexity* sebesar 12,69. Hasil pengujian *usability* pada visualisasi memperoleh nilai untuk indikator *learnability* 92 %, *efficiency* 95,38 %, *memorability* 90 %, *errors* 2 %, dan *satisfaction* 90 %.

Kata Kunci : *Latent Dirichlet Allocation, clustering, streamgraph, pengujian usability*

ABSTRACT

The thesis is one of the scientific papers written by students at the end of their study period which is one of the graduation requirements. The scientific paper is arranged according to scientific rules and written based on Indonesian language rules under the guidance and direction of the supervisor. However, in the final project, students are often finding difficulty choosing their thesis topics and that will cause their thesis postponed. Knowing topic trends will make it easy to find topics that have many references. In this research, we use the Latent Dirichlet Allocation (LDA) method to cluster the abstract of the final assignment with a proper Indonesian language. Based on the similarity of the thesis's abstract, LDA will cluster the words into topics. The LDA would produce two outputs, topic proportion, and the probability of word-topic. To see the results of the final clustering thesis's abstract, we needed a visualization. One of the visualizations that could display the results of clustering was Streamgraph. The purpose of this research was to produce a Streamgraph that displays the results of clustering from the final assignment abstract using the best combination of LDA parameters. The best combination of parameters in this case, was to use alpha 0,1, beta 0,1, and threshold 0,001, while the topics were determined based on the unique words of abstracts resulting in the value is 13. The combination of parameters yielded a perplexity value of 12,69. The usability testing resulted in visualization for learnability was 92%, efficiency was 95,38%, memorability was 90%, errors was 2%, and satisfaction was 90%.

Keywords: *Latent Dirichlet Allocation, clustering, streamgraph, usability testing.*

KATA PENGANTAR

Segala puji syukur bagi Allah SWT atas karunia-Nya yang diberikan kepada penulis sehingga penulis dapat menyelesaikan tugas akhir ini. Tugas akhir yang berjudul “Penentuan Tren Topik Tugas Akhir Menggunakan Metode *Latent Dirichlet Allocation* (LDA) dengan Visualisasi *Streamgraph*”. Tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Dalam penyusunan laporan tugas akhir ini tentulah telah banyak mendapat bantuan dan dukungan dari berbagai pihak. Untuk itu, pada kesempatan ini penulis mengucapkan rasa hormat dan terima kasih kepada:

1. Dr. Retno Kusumaningrum, S.Si, M.Kom, selaku Ketua Departemen Ilmu Komputer / Informatika FSM Universitas Diponegoro Semarang dan selaku dosen Pembimbing yang telah meluangkan waktu dan berkenan memberikan bimbingan, arahan, masukan, serta motivasi yang sangat berharga dan fokus akan tujuan bagi penulis.
2. Panji Wisnu Wirawan, ST, MT, selaku Koordinator Tugas Akhir Departemen Ilmu Komputer/Informatika FSM Universitas Diponegoro Semarang
3. Orangtua, keluarga, teman dekat dan sahabat yang telah mendukung, membantu dan memberikan semangat kepada penulis dalam menyelesaikan tugas akhir ini.
4. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu kelancaran penelitian ini, semoga Tuhan yang memberikan balasan yang lebih baik.

Penulis menyadari bahwa dalam laporan ini masih banyak kekurangan baik dari segi materi ataupun dalam penyajiannya karena keterbatasan kemampuan dan pengetahuan penulis. Oleh karena itu, kritik dan saran sangat penulis harapkan. Semoga laporan ini dapat bermanfaat bagi pembaca pada umumnya dan penulis pada khususnya.

Semarang, 18 Februari 2019

Penulis

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xvi
1. BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan dan Manfaat.....	3
1.4 Ruang Lingkup	3
1.5 Sistematika Penulisan.....	4
2. BAB II TINJAUAN PUSTAKA	6
2.1 Penentuan Tren Topik Tugas Akhir	6
2.2 Preprocessing.....	7
2.3 Latent Dirichlet Allocation (LDA).....	16
2.4 Perplexity.....	23
2.5 Streamgraph.....	24
2.6 Usability Testing	24
2.7 Pengembangan Perangkat Lunak	28
3. BAB III METODOLOGI PENELITIAN	30

3.1	<i>Preprocessing</i>	32
3.1.1	Tokenisasi	33
3.1.2	<i>Stopword Removal</i>	34
3.1.3	<i>Stemming</i>	36
3.2	LDA Collapsed Gibbs Sampling	43
3.3	Evaluasi Kinerja Model LDA	46
3.4	Visualisasi Streamgraph	47
3.5	Evaluasi Visualisasi Streamgraph	49
3.6	Analisa dan Desain Sistem	53
3.6.1	Analisis Sistem	53
3.6.1.1	Deskripsi Sistem	53
3.6.1.2	Kebutuhan Fungsional Sistem	53
3.6.1.3	Kebutuhan Non-Fungsional Sistem	54
3.6.2	Perancangan Sistem	54
3.6.2.1	Pemodelan Data	54
3.6.2.2	Pemodelan Fungsi	55
3.6.2.3	Perancangan Antarmuka	57
4.	BAB IV HASIL EKSPERIMEN DAN ANALISA	62
6.1	Hasil Pengembangan Sistem	62
4.1.1	Lingkungan Implementasi	62
4.1.2	Implementasi Antarmuka	63
4.2	Skenario Pengujian Sistem	67
4.2.1	Pengujian Fungsional Sistem	68
4.2.2	Pengujian Kinerja Sistem	68
4.2.2.1	Data	68
4.2.2.2	Eksperimen	68
4.3	Hasil dan Analisis Sistem	69
4.3.1	Pengujian Fungsional Sistem	69
4.3.2	Hasil Eksperimen 1	70

4.3.3	Hasil Eksperimen 2.....	74
5.	BAB V PENUTUP	80
5.1	Kesimpulan.....	80
5.2	Saran	80
	DAFTAR PUSTAKA.....	81
	LAMPIRAN-LAMPIRAN	84

DAFTAR GAMBAR

Gambar 2.1 Proses Pengindeksan (Yates & Neto, 1999).....	7
Gambar 2.2 Alur Pembangkitan Informasi pada Dokumen dengan Topic Models (Liu, 2013).....	17
Gambar 2.3 <i>Graphical Model</i> PLSA dan LDA.....	18
Gambar 2.4 Topologi jaringan LDA (Liu, 2013)	19
Gambar 2.5 Contoh Implementasi LDA sebagai Inferensi (Blei, Ng & Jordan, 2003)	20
Gambar 2.6 Visualisasi LDA sebagai Inferensi (Kusumaningrum, 2014).....	20
Gambar 2.7 LDA Model (Kusumaningrum et al., 2014)	22
Gambar 2.8 Contoh LDA sebagai Generatif Model (Blei, Ng & Jordan, 2003).....	23
Gambar 2.9 <i>Streamgraph</i>	24
Gambar 2.10 <i>Waterfall Model</i> (Sommerville, 2011).....	28
Gambar 3.1 Gambaran Umum Penelitian	30
Gambar 3.2 <i>Flowchart Preprocessing</i>	32
Gambar 3.3 <i>Flowchart</i> Tokenisasi	33
Gambar 3.4 <i>Flowchart Stopword Removal</i>	35
Gambar 3.5 <i>Flowchart Stemming</i> Sastrawi (Bashri, 2017).....	36
Gambar 3.6 <i>Flowchart Subproses Stemming</i> Sastrawi (Bashri, 2017)	37
Gambar 3.7 <i>Flowchart Subproses Stemming Plural</i> (Bashri, 2017).....	38
Gambar 3.8 <i>Flowchart Subproses Stemming Singular</i> (Bashri, 2017).....	39
Gambar 3.9 <i>Flowchart</i> loopPengembalianAkhiran (Bashri, 2017).....	41
Gambar 3.10 <i>Flowchart</i> Proses LDA <i>Collapsed Gibbs Sampling</i>	45
Gambar 3.11 <i>Flowchart</i> Pembentukan <i>file csv</i>	48
Gambar 3.12 <i>Entity Relationship Diagram (ERD)</i>) Penentuan Tren Topik Tugas Akhir.....	54
Gambar 3.13 <i>Data Context Diagram (DCD)</i> Penentuan Tren Topik Tugas Akhir	56
Gambar 3.14 <i>Data Flow Diagram Level 1</i> Penentuan Tren Topik Tugas Akhir	57
Gambar 3.15 Rancangan Antarmuka Halaman Beranda.....	58
Gambar 3.16 Rancangan Antarmuka Halaman Login Admin	58
Gambar 3.17 Rancangan Antarmuka Halaman Input Data	59
Gambar 3.18 Rancangan Antarmuka Halaman Data.....	59

Gambar 3.19 Rancangan Antarmuka Halaman <i>Preprocessing</i>	60
Gambar 3.20 Rancangan Antarmuka Halaman LDA	60
Gambar 3.21 Antarmuka Halaman Visualisasi <i>Streamgraph</i>	61
Gambar 4.1 Implementasi Antarmuka Beranda	63
Gambar 4.2 Implementasi Halaman <i>Login Admin</i>	63
Gambar 4.3 Implementasi Halaman Input Data	64
Gambar 4.4 Implementasi Antarmuka Data	65
Gambar 4.5 Implementasi Antarmuka Halaman <i>Preprocessing</i>	66
Gambar 4.6 Implementasi Antarmuka Halaman LDA	66
Gambar 4.7 Implementasi Antarmuka Halaman Visualisasi <i>Streamgraph</i>	67
Gambar 4.8 Grafik Nilai <i>Perplexity</i> untuk Seluruh Kombinasi Parameter	71
Gambar 4.9 Contoh Grafik Probabilitas Kata dari Parameter K-3	72
Gambar 4.10 Grafik Grafik Probabilitas Kata dari Parameter K-7	72
Gambar 4.11 Visualisasi <i>Streamgraph</i> untuk Kombinasi K-03	73
Gambar 4.12 Hasil <i>Post-test</i> Kuesioner	75
Gambar 4.13 Persentase Indikator <i>Usability</i>	79

DAFTAR TABEL

Tabel 2.1 <i>State of The Art</i>	6
Tabel 2.2 Kombinasi Awalan-Akhiran yang Tidak Diperbolehkan.....	9
Tabel 2.3 Aturan Pemenggalan Awalan Algoritma <i>Stemming</i> Nazief dan Adriani	11
Tabel 2.4 Modifikasi dan Penambahan Aturan Pemenggalan Awalan oleh Algoritma <i>Stemming Confix Stripping</i>	13
Tabel 2.5 Daftar Aturan <i>Rule Precedence</i>	13
Tabel 2.6 Modifikasi Aturan Pemenggalan Awalan oleh Algoritma <i>Stemming Enhanced Confix Stripping</i> (Tahitoe & Purwitasari, 2010).....	14
Tabel 2.7 Modifikasi Aturan Pemenggalan Awalan dan Penambahan Aturan Pemenggalan Sisipan oleh Algoritma <i>Stemming modified Enhanced Confix Stripping</i>	16
Tabel 2.8 Penambahan dan Modifikasi Aturan Pemenggalan Awalan <i>Stemmer Sastrawi</i>	16
Tabel 2.9 Daftar Variabel LDA.....	18
Tabel 2.10 Skor Jawaban Responden.....	26
Tabel 2.11 Rekapitulasi Hasil Kuesioner	26
Tabel 2.12 Keterangan Persentase Tanggapan Responden	27
Tabel 2.13 Contoh Rekapitulasi Hasil Kuesioner	27
Tabel 3.1 Proses Tokenisasi	34
Tabel 3.2 Daftar <i>Stopword</i>	34
Tabel 3.3 Proses Penghilangan <i>Stopword</i>	35
Tabel 3.4 Proses <i>Stemming</i>	42
Tabel 3.5 Bentuk <i>Bag of Words</i>	42
Tabel 3.6 <i>Vocabulary</i>	43
Tabel 3.7 Contoh Probabilitas Kata-Topik.....	45
Tabel 3.8 Contoh <i>Topic Proportion</i>	46
Tabel 3.9 Contoh Struktur <i>file csv</i>	49
Tabel 3.10 Responden Pengujian <i>Usability</i>	50
Tabel 3.11 Aspek Penilaian Indikator <i>Usability</i>	50
Tabel 3.12 Contoh Hasil <i>Post-Test</i> Kuesioner	51
Tabel 3.13 Contoh Penyelesaian Pengerjaan <i>Task</i>	51

Tabel 3.14 Contoh Waktu Eksekusi Pengerjaan <i>Task</i>	52
Tabel 3.15 Kebutuhan Fungsional Sistem.....	54
Tabel 3.16 Kebutuhan Non Fungsional Sistem.....	54
Tabel 4.1 Rencana Pengujian Fungsional Sistem.....	69
Tabel 4.2 Daftar Kombinasi Parameter	70
Tabel 4.3 Hasil <i>Post-Test</i> Kuesioner	74
Tabel 4.4 Penyelesaian Pengerjaan <i>Task</i>	77
Tabel 4.5 Waktu Eksekusi Pengerjaan <i>Task</i>	77

DAFTAR LAMPIRAN

Lampiran 1. Perhitungan LDA <i>Collapsed Gibbs Sampling</i>	85
Lampiran 2. Lembar Pengujian <i>Usability</i>	98
Lampiran 3. Deskripsi dan Hasil Pengujian Fungsional Sistem	101

BAB I

PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat, serta ruang lingkup penelitian dan sistematika penulisan tugas akhir mengenai Penentuan Tren Topik Tugas Akhir Menggunakan Metode *Latent Dirichlet Allocation* dengan Visualisasi *Streamgraph* studi kasus di Departemen Ilmu Komputer/ Informatika Universitas Diponegoro.

1.1 Latar Belakang

Tugas akhir adalah salah satu karya ilmiah yang ditulis oleh mahasiswa pada akhir masa studinya yang merupakan salah satu syarat kelulusan untuk seorang mahasiswa. Karya ilmiah tersebut disusun menurut kaidah keilmuan dan ditulis berdasarkan kaidah Bahasa Indonesia di bawah bimbingan dan arahan dosen pembimbing, untuk memenuhi kriteria-kriteria kualitas yang telah ditetapkan sesuai kelimuan masing-masing. Dalam pengerjaan Tugas Akhir, sering kali mahasiswa kebingungan dalam pemilihan topik yang akan dikerjakan sehingga menambah waktu pengerjaan tugas akhirnya. Pengetahuan tentang tren topik tugas akhir mahasiswa di sebuah Universitas pada umumnya maupun di program studi tertentu pada khususnya dapat membawa manfaat yang sangat positif bagi pengembangan kurikulum (Prilianti dan Wijaya, 2014). Dengan mengetahui tren topik yang ada, akan lebih mudah untuk mencari topik yang memiliki banyak referensi. Dengan demikian pemilihan topik yang akan digunakan dalam tugas akhir akan lebih cepat. Semakin cepat pemilihan topik maka akan mempercepat mahasiswa untuk menyelesaikan tugas akhirnya.

Tugas akhir memiliki abstrak yang merupakan intisari dari pembahasan secara keseluruhan. Cara untuk mengetahui isi dari abstrak dapat menggunakan *topic modelling*. *Topic Modeling* adalah kumpulan algoritma yang digunakan untuk menemukan struktur tersembunyi dari tema yang terdapat dalam setiap dokumen (Blei, 2012). *Topic modeling* memiliki beberapa model, yaitu *linear topic modeling* dan *probabilistic topic modeling* (Pleplé, 2013). *Probabilistic topic modeling* merupakan pengembangan dari metode *linear topic modeling*, yaitu *Latent Semantic Indexing* (LSI) (Liu, 2013). LSI mengasumsikan bahwa kata-kata dalam suatu dokumen akan mempengaruhi topik yang dapat digunakan sebagai rujukan ke dokumen itu sendiri

(Deerwester et al., 1990). Sedangkan *probabilistic topic modeling* mendefinisikan topik sebagai distribusi berdasarkan kata-kata dan mendefinisikan dokumen sebagai campuran topik (Blei, 2012). Oleh karena itu, pada penelitian ini akan ditentukan topik tugas akhir dengan menggunakan *probabilistic topic modeling* sebagai salah satu bentuk *topic modeling*. *Probabilistic topic modeling* merupakan sebuah algoritma yang berfungsi untuk menemukan topik utama dalam suatu dokumen tersebut (Blei, 2012).

Terdapat beberapa metode pengembangan dari *probabilistic topic modeling* diantaranya, yaitu *Probabilistic Latent Semantic Analysis* (PLSA) dan *Latent Dirichlet Allocation* (LDA). LDA merupakan pengembangan dari PLSA (Blei, Ng dan Jordan, 2003) yang lebih stabil untuk mengolah data dalam jumlah besar (Liu, 2013). Metode LDA mengasumsikan bahwa pada satu dokumen terdapat lebih dari satu topik, yang masing-masing merupakan distribusi melalui kosakata. Metode LDA dapat diterapkan sebagai proses generatif maupun proses inferensi. LDA sebagai proses generatif akan menghasilkan sebuah korpus dari probabilitas kata-topik dan *topic proportion* yang telah diketahui sebelumnya. Sedangkan LDA sebagai proses inferensi akan menghasilkan probabilitas kata-topik dan *topic proportion* dari korpus yang digunakan sebagai masukan.

Penelitian ini akan diaplikasikan LDA dengan proses inferensi. Metode LDA dengan proses inferensi, akan menghasilkan *topic proportion* dan probabilitas kata-topik berupa *matrix*. Penelitian ini memerlukan memvisualisasikan agar lebih mudah membaca hasil dari analisis *topic modeling* menggunakan LDA ini. Penelitiain menggunakan LDA sebelumnya, memvisualisasikan hasilnya dalam bentuk grafik, seperti pada penelitian yang dilakukan Zulhanif pada tahun 2016 yang memvisualisasikan hasil LDA menggunakan grafik batang dan *wordcloud*, lalu pada penelitian oleh Prihatini dan kawan-kawannya pada tahun 2017 yang berjudul *Fuzzy-Gibbs Latent Dirichlet Allocation Model for Feature Extraction on Indonesian Documents* juga menggunakan grafik batang, lalu pada analisis dan visualisasi suara pelanggan pada pusat pelayanan dan pelanggan dengan pemodelan topik menggunakan LDA studi kasus : PT. Petrokimia Gresik oleh Agustina pada tahun 2017, yang menggunakan visualisasi grafik batang dan garis yang menampilkan topik dan probabilitas yang diperoleh dari LDA. Pada ketiga penelitian tersebut, digunakan visualisasi dengan grafik untuk menunjukkan probabilitas dari setiap kata, sedangkan

pada penelitian ini ingin menunjukkan tren dari probabilitas topik dari tiap waktu ke waktu. Sehingga kurang cocok jika menggunakan visualisasi dalam bentuk grafik.

Penelitian ini menggunakan data berupa dokumen abstrak dari laporan tugas akhir mahasiswa informatika Undip pada tahun 2011-2016 yang berjumlah 294 dokumen. Tren topik tidak dapat dibaca dengan mudah, oleh karena itu diusulan untuk menampilkan visualisasi dari hasil proses LDA dalam bentuk *Streamgraph*. Visualisasi dengan bentuk *Streamgraph* dapat menampilkan distribusi topik dalam interval waktu, selain itu *Streamgraph* akan menyajikan grafik dengan bentuk lapisan-lapisan, dimana dalam lapisan-lapisan tersebut akan mewakili sebuah topik dalam waktu tertentu (Stojanovski, Dimitrovski & Madjarov, 2014).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka dapat dirumuskan permasalahan yaitu bagaimana memetakan tren topik tugas akhir mahasiswa Departemen Ilmu Komputer/ Informatika Universitas Diponegoro dengan menggunakan metode LDA dan bagaimana memvisualisasikan hasil dengan menggunakan visualisasi *Streamgraph*.

1.3 Tujuan dan Manfaat

Tujuan yang ingin dicapai dalam tugas akhir ini antara lain :

1. Menentukan kombinasi parameter LDA terbaik untuk mengoptimalkan hasil dari aplikasi.
2. Mengukur seberapa mudah menggunakan visualisasi *Streamgraph* bagi Pengguna dengan melakukan pengujian *usability*.

Adapun manfaat yang diharapkan dari penelitian tugas akhir ini adalah terciptanya aplikasi yang dapat membantu memetakan tren topik tugas akhir dalam bentuk yang mudah dibaca sehingga dapat memberikan referensi dalam pemilihan topik tugas akhir mereka.

1.4 Ruang Lingkup

Ruang lingkup yang akan dibahas dalam penelitian ini adalah:

1. *Input* dari sistem ini berupa data abstrak dari dari tugas akhir di departemen Ilmu Komputer/ Informatika Undip dari tahun 2011-2016.

2. Abstrak yang digunakan dalam korpus menggunakan bahasa Indonesia.
3. Kinerja dari *clustering* menggunakan LDA hanya dilihat dari nilai *perplexity*.
4. *Output* dari sistem ini berupa hasil pemetaan tren topik tugas akhir yang divisualisasikan dengan *Streamgraph*.
5. Pengujian *usability* dilakukan untuk menilai visualisasi *Streamgraph* berdasarkan indikator *usability* menurut Nielsen.
6. Pengembangan sistem ini menggunakan bahasa pemrograman Python, HTML, CSS, PHP, Javascript dan Sistem manajemen basis data MySQL.

1.5 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi dalam beberapa pokok bahasan, yaitu :

BAB I PENDAHULUAN

Bab ini memberikan gambaran mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup serta sistematika penulisan tugas akhir mengenai Penentuan Tren Topik Tugas Akhir menggunakan LDA dengan Visualisasi *Streamgraph*.

BAB II TINJAUAN PUSTAKA

Bab ini memberikan kajian pustaka yang berhubungan dengan tema tugas akhir sebagai landasan untuk perumusan dan analisis permasalahan pada tugas akhir. Bab ini menyajikan penentuan tren topik, tahap *preprocessing*, penjelasan tentang LDA dengan visualisasi *Streamgraph*.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan garis besar penyelesaian masalah tugas akhir yang diawali dengan penyajian garis besar penyelesaian masalah dalam bentuk blok proses. Garis besar penyelesaian masalah diawali dengan melakukan *preprocessing* untuk pembentukan *bag of words*, proses LDA, proses evaluasi dengan visualisasi *Streamgraph*.

BAB IV HASIL EKSPERIMEN DAN ANALISA

Bab ini menguraikan hasil skenario eksperimen dan analisa pada penelitian yang dimulai dari teknis pengumpulan data, penjelasan

mengenai semua skenario eksperimen dan analisa dari setiap hasil eksperimen yang telah dilakukan.

BAB V

PENUTUP

Bab ini menjabarkan kesimpulan dari uraian yang telah diulas pada bab-bab sebelumnya dan saran untuk pengembangan penelitian lebih lanjut.