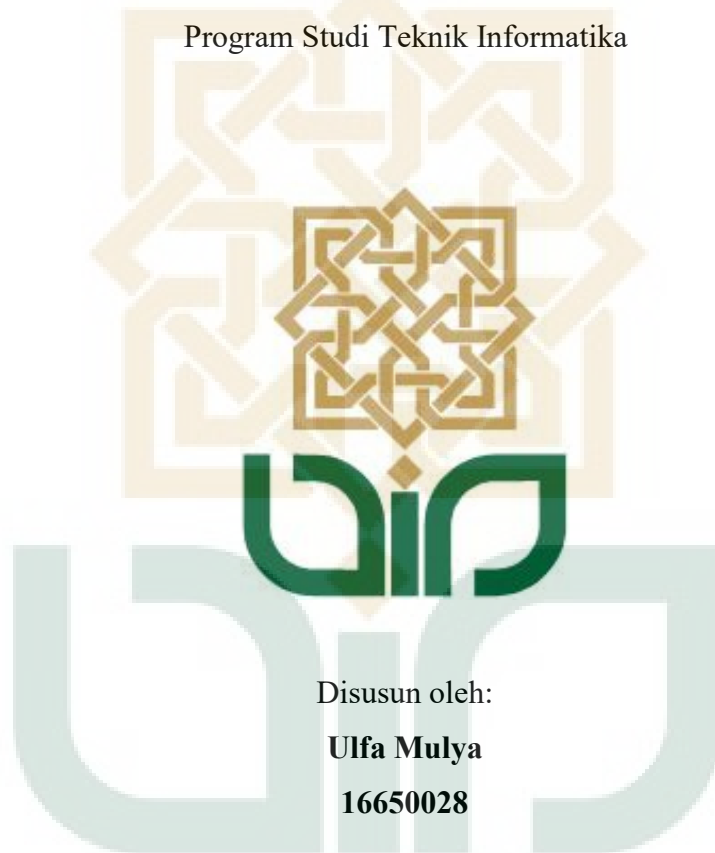


**ANALISIS KOMPARASI PEMODELAN TOPIK *LATENT*  
*DIRICHLET ALLOCATION* DAN *LATENT SEMANTIC*  
*ANALYSIS* PADA ULASAN RESTORAN DI YOGYAKARTA**

Skripsi  
untuk memenuhi sebagai persyaratan  
mencapai derajat Sarjana S-1  
Program Studi Teknik Informatika



Disusun oleh:  
**Ulfa Mulya**  
**16650028**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
YOGYAKARTA  
2020**

## HALAMAN PENGESAHAN



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-995/Un.02/DST/PP.00.9/04/2020

Tugas Akhir dengan judul : ANALISIS KOMPARASI PRMODELAN TOPIK LATENT DIRICHLET  
ALLOCATION DAN LATENT SEMANTIC ANALYSIS PADA ULASAN  
RESTORAN DI YOGYAKARTA

yang dipersiapkan dan disusun oleh:

Nama : ULFA MULYA  
Nomor Induk Mahasiswa : 16650028  
Telah diujikan pada : Kamis, 09 April 2020  
Nilai ujian Tugas Akhir : A

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

### TIM UJIAN TUGAS AKHIR

Ketua Sidang/Penguji I

Muhammad Didik Rohmad Wahyudi, S.T., MT.  
NIP. 19760812 200901 1 015

Penguji II

Agus Mulyanto, S.Si., M.Kom.  
NIP. 19710823 199903 1 003

Penguji III

Muhammad Taufiq Nuruzzaman, S.T. M.Eng.  
NIP. 19791118 200501 1 003

Yogyakarta, 09 April 2020  
UIN Sunan Kalijaga  
Fakultas Sains dan Teknologi  
Dekan



Dr. Murtono, M.Si.  
NIP. 19691212 200003 1 001

## 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 : Ulfa Mulya  
NIM : 16650028  
Judul Skripsi : "Analisis Komparasi Pemodelan Topik *Latent Dirichlet Allocation* dan *Latent Semantic Analysis* pada Ulasan Restoran di 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, 07 April 2020  
Pembimbing

M. Didik Rohmad Wahyudi, S.T., MT.  
NIP. 19760812 200901 1 015

## PERNYATAAN KEASLIAN SKRIPSI

### PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini :

Nama : Ulfa Mulya

NIM : 16650028

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul “**Analisis Komparasi Pemodelan Topik *Latent Dirichlet Allocation* dan *Latent Semantic Analysis* pada Ulasan Restoran di 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, 08 April 2020



Ulfa Mulya  
NIM.16650028

## KATA PENGANTAR

Alhamdulillahirobbil'alamin. Puji syukur kehadirat Allah SWT karena rahmat dan izin-Nya sehingga penulis dapat menyelesaikan penelitian yang berjudul **Analisis Komparasi Pemodelan Topik *Latent Dirichlet Allocation* dan *Latent Semantic Analysis* pada Ulasan Restoran di Yogyakarta** sebagai salah satu syarat untuk mencapai gelar sarjana program studi Teknik Informatika Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Sholawat serta salam selalu tercurahkan kepada junjungan kita Nabi Muhammad SAW. beserta seluruh keluarga dan sahabat beliau.

Penulis menyadari bahwa apa yang dilakukan dalam penyusunan laporan penelitian ini masih terlalu jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran yang berguna dalam penyempurnaan analisis ini di masa yang akan datang. Semoga apa yang telah penulis lakukan dapat bermanfaat bagi pembaca.

Selanjutnya, penulis ingin mengucapkan terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian skripsi ini, baik secara langsung maupun tidak langsung. Ucapan terimakasih penulis sampaikan kepada:

1. Bapak Prof. Drs. K.H. Yudian Wahyudi, M.A., Ph.D., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Bapak Dr. Murtono, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.

3. Bapak Sumarsono, S.T., M.Kom., selaku Ketua Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta.
4. Bapak Muhammad Didik Rohmad Wahyudi, S.T., MT., selaku dosen pembimbing skripsi yang telah sabar dan meluangkan waktunya untuk memberikan koreksi dan kritik saran kepada penulis sehingga skripsi ini dapat diselesaikan.
5. Bapak Dr. Agung Fatwanto, S.Si., M.Kom., selaku Dosen Pembimbing Akademik.
6. Agus Mulyanto, S.Si., M.Kom., Ph.D., Aulia Faqih Rifa'i, M.Kom., M. Taufiq Nuruzzaman, S.T., Maria Ulfah Siregar, S.Kom. MIT., Ph.D., Nurochman, S.Kom., M.Kom., Rahmat Hidayat, S.Kom., M.Cs., Dr. Shofwatul 'Uyun, S.T., M.Kom., selaku dosen pengampu mata kuliah program studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta yang telah banyak membantu sehingga penulis dapat menyusun tugas akhir.
7. Pak Wahdan, dan seluruh staf karyawan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
8. Manusia terbaik, Ibu Suryani dan Bapak Salman. Segala yang baik baik berawal dan berasal dari mereka.
9. Guru ngaji penulis, Ibu Siti-Pak Hasan yang sudah menjadi orang tua kedua penulis.
10. Seluruh Keluarga Besar Pakpuh Amat Marsudi dan Nenek Noerbaini yang senantiasa mendoakan dan memberikan dukungan kepada penulis.

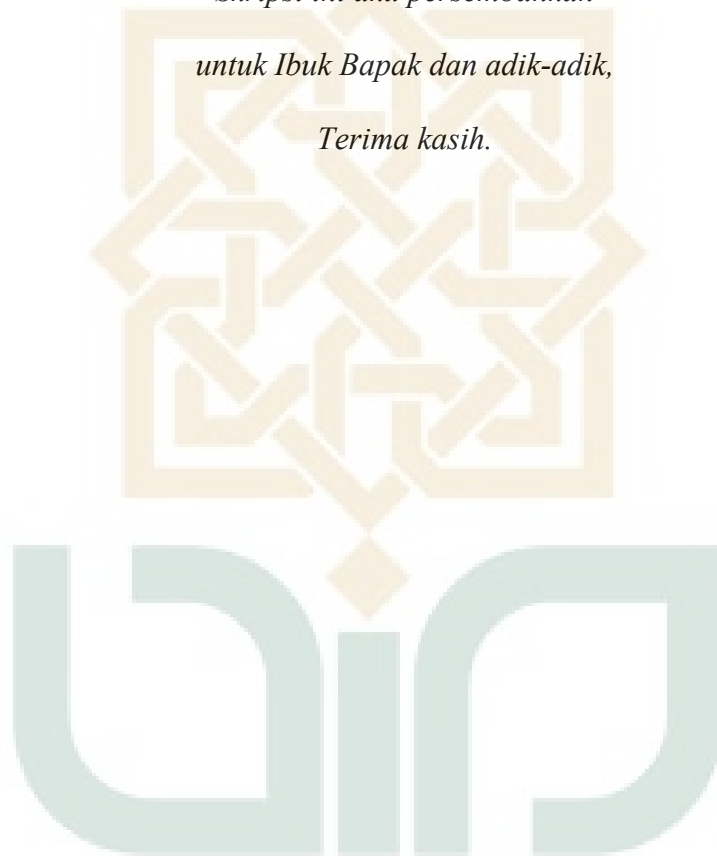
11. Teman-teman diskusi dalam banyak hal anggota WAG gathel, wacana, trio, trio n dad, ghibah, new, anu, kopril, lost in Thai.
12. Terkhusus Sekar, Nadia, Lina, Yayang, Lia, Raffi, Mamad, Ari, Hendra, Nur, Alvri, Meri, Rizky, Mutiah, Irma, Mbak Nafi, Mbak Mardiyah.
13. Teman-teman yang pernah menjadi satu kelompok selama kuliah dan seluruh teman-teman Teknik Informatika 2016 yang tidak dapat penulis sebutkan satu per satu.
14. Keluarga Sirarom Mansion. Teman seperjuangan di Negeri Gajah Putih, dan tetap menjadi mentor-mentor terbaik sampai saat ini.
15. Anggota PERMITHA, mas Arul, mbak Uli, mas Mirza, dkk; Ajarn-Ajarn ICT PSU; teman-teman ICT PSU, p'Aseeyah, p'Kan, p'Ven, p'Ko, p'Coco, p'Sentanut, Chuluq, Sa, Na, Ya.
16. Guru-guru virtual penulis di twitter, Dokdes, dr.Gia, dr.Jiemi, Mbak Amalia Paravoti, Pak Ayang, Pak Budhiman, Kak Cita, Teh Fen, Pinotski, Lantip, akun NUGarislucu, dan HinduGL.
17. Teman-teman KKN Kelompok 104 Angkatan 99.
18. Serta semua pihak yang memberi dukungan *lillah*, sehingga penelitian ini dapat terselesaikan.

Yogyakarta, 20 Maret 2020

Penulis

## HALAMAN PERSEMBAHAN

*Skripsi ini aku persembahkan  
untuk Ibuk Bapak dan adik-adik,  
Terima kasih.*





## HALAMAN MOTTO

*“We have four magic words: maaf, tolong, terima kasih, permisi,  
and I have two encouraging words: bismillah dan alhamdulillah ”*

*“Tersesat boleh diam jangan ”*

-Penulis.

## DAFTAR ISI

<b>HALAMAN PENGESAHAN.....</b>	<b>ii</b>
<b>SURAT PERSETUJUAN SKRIPSI.....</b>	<b>iii</b>
<b>PERNYATAAN KEASLIAN SKRIPSI.....</b>	<b>iv</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>HALAMAN PERSEMBAHAN.....</b>	<b>viii</b>
<b>HALAMAN MOTTO.....</b>	<b>ix</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR TABEL.....</b>	<b>xiv</b>
<b>INTISARI.....</b>	<b>xv</b>
<b>ABSTRACT.....</b>	<b>xvi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	4
1.6. Keaslian Penelitian.....	4
1.7. Sistematika Penulisan.....	4
<b>BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....</b>	<b>7</b>
2.1. Tinjauan Pustaka.....	7
2.2. Landasan Teori.....	13
2.2.1. <i>Machine Learning</i> .....	13
2.2.2. <i>Text Mining</i> .....	13
2.2.3. <i>Topic Modeling</i> .....	14

2.2.4.	<i>Latent Dirichlet Allocation (LDA)</i> .....	15
2.2.5.	<i>Latent Semantic Analysis (LSA)</i> .....	17
2.2.6.	<i>Topic Coherence</i> .....	20
2.2.7.	<i>Google Maps</i> .....	23
2.2.8.	<i>Python</i> .....	23
<b>BAB III METODE PENELITIAN</b> .....		<b>25</b>
3.1.	Metode Penelitian.....	25
3.2.	Alur Penelitian.....	25
3.2.1.	Studi Pendahuluan.....	26
3.2.2.	Pengumpulan Data.....	26
3.2.3.	<i>Data Preprocessing</i> .....	26
3.2.4.	Membentuk Input Model.....	27
3.2.5.	<i>Topic Modeling</i> .....	28
3.2.6.	Analisis Hasil.....	28
3.2.7.	Luaran Model Topik.....	28
3.3.	Kebutuhan Sistem.....	29
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....		<b>30</b>
4.1.	Pengumpulan Data.....	30
4.2.	<i>Data Preprocessing</i> .....	32
4.2.1.	<i>Case Folding</i> .....	32
4.2.2.	<i>Stopword Removing</i> .....	34
4.3.	Pembentukan Frasa.....	35
4.4.	Pembentukan Input Model.....	35
4.5.	Membangun Topik Model.....	37
4.5.1.	<i>Topic Modeling</i> .....	37
4.5.1.1.	<i>Latent Dirichlet Allocation</i> .....	37
4.5.1.2.	<i>Latent Semantic Analysis</i> .....	38
4.5.2.	<i>Topic Coherence Measurement</i> .....	39

4.6.	Analisis Hasil.....	41
4.6.1.	Analisis Hasil Jumlah Topik.....	41
4.6.1.1.	Analisis Hasil Jumlah Topik LDA.....	41
4.6.1.2.	Analisis Hasil Jumlah Topik LSA.....	43
4.6.2.	Analisis Hasil Luaran Model.....	44
4.6.2.1.	Analisis Perbandingan Pemodelan Topik.....	48
4.6.3.	Analisis Hasil Luaran Model Topik.....	49
<b>BAB V PENUTUP.....</b>		<b>52</b>
5.1.	Kesimpulan.....	52
5.2.	Saran.....	52
<b>DAFTAR PUSTAKA.....</b>		<b>54</b>
<b>LAMPIRAN.....</b>		<b>57</b>
<b>CURRICULUM VITAE.....</b>		<b>74</b>

## DAFTAR GAMBAR

<b>Gambar 2.1</b> Ilustrasi Topic Modeling (D. Blei, Carin, & Dunson, 2010).....	15
<b>Gambar 2.2</b> Representasi Graphical LDA Model (D. M. Blei et al., 2003).....	16
<b>Gambar 2.3</b> Proses Perhitungan Topic Coherence (Röder, Both, & Hinneburg, 2015).....	21
<b>Gambar 2.4</b> Coherence Measurement Score (Mimno et al., 2011).....	21
<b>Gambar 3.1</b> Skema Alur Penelitian.....	25
<b>Gambar 4.1</b> Flowchart Proses Analisis.....	30
<b>Gambar 4.2</b> Alur Topik Model.....	37
<b>Gambar 4.3</b> Gambaran Proses Distribusi Topik dan Kata dalam Dokumen.....	38
<b>Gambar 4.4</b> Matriks Kerja LSA.....	39
<b>Gambar 4.5</b> Segmentasi Cv Coherence Measurement.....	40
<b>Gambar 4.6</b> Grafik LDA Coherence Measurement.....	42
<b>Gambar 4.7</b> Grafik LSA Coherence Measurement.....	43
<b>Gambar 4.8</b> Plot Top 3 Topics LDA.....	47
<b>Gambar 4.9</b> Plot Top 3 Topics LSA.....	47
<b>Gambar 4.10</b> Visualisasi Model Interaktif LDA.....	49

## DAFTAR TABEL

<b>Tabel 2.1</b> Tinjauan Pustaka.....	10
<b>Tabel 2.2</b> Tinjauan Pustaka (Lanjutan).....	11
<b>Tabel 2.3</b> Tinjauan Pustaka (Lanjutan).....	12
<b>Tabel 2.4</b> Kualitas Model Topik (“ilmu data - Sarah,” n.d.).....	22
<b>Tabel 4.1</b> Data Ulasan.....	31
<b>Tabel 4.2</b> Data Indeks.....	32
<b>Tabel 4.3</b> Contoh Penerapan Case Folding.....	33
<b>Tabel 4.4</b> Contoh Penerapan Stopword Removing.....	34
<b>Tabel 4.5</b> Contoh Hasil Bigram Model.....	35
<b>Tabel 4.6</b> Contoh Dokumen untuk Pembobotan TF-IDF.....	36
<b>Tabel 4.7</b> Contoh Proses Pembobotan TF-IDF.....	36
<b>Tabel 4.8</b> Nilai Koherensi Topik LDA Model.....	42
<b>Tabel 4.9</b> Nilai Koherensi Topik LSA Model.....	44
<b>Tabel 4. 10</b> Ulasan Khusus.....	48
<b>Tabel 4. 11</b> Analisis Komparasi LDA dan LSA.....	48
<b>Tabel 4.12</b> Koordinat Topik Luaran.....	50
<b>Tabel 4.13</b> Informasi Topik.....	50
<b>Tabel 4.14</b> Tabel Token.....	51

**Analisis Komparasi Pemodelan Topik *Latent Dirichlet Allocation* dan *Latent Semantic Analysis* pada Ulasan Restoran di Yogyakarta**

**Ulfa Mulya**

**16650028**

**INTISARI**

Restoran merupakan prospek usaha yang paling menjanjikan di Yogyakarta. Untuk mengetahui hal-hal yang menjadi perhatian masyarakat perlu dilakukannya sebuah survey. Adapun salah satu cara yang dapat dilakukan yaitu dengan memanfaatkan ulasan konten dari pengguna *google maps*. Maka dari itu, dilakukan penelitian guna menghemat biaya, waktu, dan tenaga.

Penelitian ini bertujuan untuk mencari tren topik tersembunyi pada sebuah kumpulan ulasan restoran secara otomatis guna mempermudah pembacaan dan pemahaman terhadap data dengan menggunakan metode *topic modeling* yaitu *Latent Dirichlet Allocation (LDA)* dan *Latent Semantic Analysis (LSA)*. Kedua metode tersebut diuji dengan 1010 data ulasan untuk menemukan luaran model topik terbaik yang paling mewakili interpretasi masyarakat.

Hasil evaluasi dengan *Cv coherence measurement* memberikan nilai 0.552396 dengan luaran model sebanyak 7 topik pada algoritma LDA dan nilai 0.55132 dengan luaran model sebanyak 4 topik pada algoritma LSA. Dalam penelitian ini algoritma LDA terbukti bekerja lebih baik, dan 7 topik menjadi luaran topik terbaik, yaitu “harga”, “suasana”, “menu”, “unik”, “lesehan”, “instagramable”, “nongkrong”.

**Kata kunci:** *text mining, topic modeling, LDA, LSA, ulasan google maps.*

**Comparative Analysis of Latent Dirichlet Allocation and Latent Semantic  
Analysis Topic Modeling on Restaurant Reviews in Yogyakarta**

**Ulfa Mulya**

**16650028**

**ABSTRACT**

Restaurants are the most promising business prospects in Yogyakarta. To find out things that are of concern to the community, a survey is needed. As for one way that can be done is to take advantage of content reviews from Google Maps users. Therefore, research is carried out to save costs, time and energy.

This research aims to find hidden topic trends in a collection of restaurant reviews automatically to facilitate reading and understanding data using topic modeling methods namely Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA). Both methods were tested with 1010 review data to find the output of the best topic model that represents human interpretation.

The results of the evaluation by measuring Cv coherence produce a value of 0.552396 with an output model of 7 topics on the LDA algorithm and produce a value of 0.55132 with an output model of 4 topics on the LSA algorithm. In this research the LDA algorithm is proven to work better, and 7 topics become the best topics, i.e. *“harga”*, *“suasana”*, *“menu”*, *“unik”*, *“lesehan”*, *“instagramable”*, *“nongkrong”*.

**Keywords:** text mining, topic modeling, LDA, LSA, google maps reviews.



# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Data Badan Pusat Statistik menunjukkan adanya kenaikan kunjungan turis asing sekitar 2,85% pada tahun 2019. Di sisi lain, berdasarkan data TripAdvisor, Yogyakarta adalah kota destinasi wisata terpopuler tahun 2019, nomor 3 setelah Bali dan Lombok. Untuk objek wisatanya sendiri, Kota Gudeg ini terkenal dengan wisata alam dan budayanya, sehingga masyarakat dan pemerintah tinggal memaksimalkannya dengan perawatan yang baik. Sebab hal ini, salah satu prospek bisnis yang menjanjikan dan sangat potensial adalah kuliner. Untuk jangkauan pasar yang lebih luas, restoran menjadi pilihan terbaik.

Restoran menjadi tempat yang paling dicari wisatawan untuk istirahat dan *recharge energy* setelah seharian lelah berjalan-jalan. Tempat yang fleksibel untuk jumlah masa yang besar maupun kecil. Tidak hanya itu, restoran juga menjadi tempat berkumpulnya keluarga. Untuk benar-benar mengetahui hal yang menjadi perhatian masyarakat perlu dilakukannya sebuah survey. Ada satu *platform* dimana pengguna dapat menikmati layanan berupa petunjuk arah dan letak lokasi suatu tempat, juga dapat memberikan penilaian terhadap tempat tersebut, yaitu *gmaps*.

*Google Maps Platform* adalah peta digital dunia yang paling populer saat ini. Manfaat utamanya yaitu mengarahkan siapapun ke lokasi manapun yang dituju. Selain itu, *gmaps* juga menyediakan tempat untuk berbagi pendapat

tentang suatu tempat yang dikunjungi. Media digital menyebutnya dengan *content-mining*. Dengan memanfaatkan ulasan konten dari pengguna *gmaps*, akan didapatkan sebuah survey yang sah dan hasil yang produktif. Dimana selanjutnya, hasil akhir yang didapat bisa dijadikan referensi untuk mendirikan suatu usaha kuliner ataupun dapat dijadikan sebagai acuan pengembangan.

Mengingat banyaknya ulasan restoran yang terdapat di dalam *google maps* dengan kriteria yang bermacam-macam, maka diperlukan banyak percobaan untuk mendapatkan data dengan *filter* tertentu. Perlu adanya pembelajaran terhadap mesin untuk membantu proses pengambilan sejumlah ulasan yang diperlukan. Pada proses selanjutnya, mesin diberikan pengetahuan dengan komputasi pada bidang *text mining* untuk melakukan eksplorasi pada data. Perlu dicarinya tren topik tersembunyi pada sebuah kumpulan ulasan restoran secara otomatis guna mempermudah pembacaan dan pemahaman terhadap data. Hal ini disebut dengan *topic modeling* (Megan R. Brett, 2012).

Pada penelitian kali ini, penulis akan melakukan komparasi algoritma *topic modeling*, yaitu *Latent Dirichlet Allocation* dan *Latent Semantic Analysis*. Hasil komparasi diharapkan dapat ditemukannya algoritma terbaik dengan nilai koherensi yang tinggi sehingga menghasilkan luaran topik model yang produktif, efektif dan memiliki makna jelas.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang yang telah dijelaskan, maka permasalahan yang ada adalah belum adanya perbandingan penerapan algoritma *Latent Dirichlet*

*Allocation* dan *Latent Semantic Analysis* dalam melakukan analisis pemodelan topik ulasan restoran di Yogyakarta, dan perlunya menemukan luaran model topik yang terbaik.

### 1.3. Batasan Masalah

Dalam penelitian ini terdapat beberapa batasan masalah yang dibahas agar penyusunan dan pembahasan penelitian dapat dilakukan secara terarah dan tercapai sesuai dengan yang diharapkan. Antara lain sebagai berikut:

1. Penerapan *text mining* menggunakan algoritma *Latent Dirichlet Allocation* dan *Latent Semantic Analysis* dalam melakukan pemodelan topik ulasan restoran.
2. Data yang akan digunakan adalah ulasan restoran berbahasa Indonesia dari aplikasi *google maps* dengan batasan lokasi di Yogyakarta dan memiliki bintang *rating* diatas empat.
3. Bahasa pemrograman yang akan digunakan adalah bahasa pemrograman *Python* dengan berbagai macam *library* di dalamnya.

### 1.4. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Melakukan perbandingan algoritma *Latent Dirichlet Allocation* dan *Latent Semantic Analysis* untuk melakukan analisis pemodelan topik ulasan restoran di Yogyakarta.
2. Menemukan luaran model topik terbaik.

### 1.5. Manfaat Penelitian

Berdasarkan latar belakang dan tujuan di atas, adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Penelitian ini dapat menambah wawasan dan pengetahuan khususnya di bidang *text mining*.
2. Mengetahui pola topik abstrak pada kumpulan ulasan restoran
3. Mengetahui hal-hal yang menjadi perhatian pelanggan.

### 1.6. Keaslian Penelitian

Penelitian menggunakan *text mining* khususnya pemodelan topik sudah banyak dilakukan. Namun, penelitian pada data ulasan restoran di *google maps* dengan metode pemodelan topik *Latent Dirichlet Allocation* dan *Latent Semantic Analysis*, yang diajukan sebagai Tugas Akhir strata S1 pada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta ini belum pernah dilakukan. Hal ini diketahui berdasarkan dari referensi dan tinjauan pustaka yang dilakukan oleh peneliti sebelumnya.

### 1.7. Sistematika Penulisan

Sebagai gambaran dan kerangka yang jelas mengenai pokok bahasan setiap bab dalam penelitian ini, maka diperlukan sistematika penulisan. Penyusunan laporan tugas akhir ini memiliki sistematika penulisan yang diawali dari BAB I dan diakhiri BAB V. Berikut adalah penjelasan pada tiap-tiap bab dalam laporan penelitian ini:

## BAB I PENDAHULUAN

Bab pendahuluan berisikan penjelasan mengenai latar belakang dilakukannya penelitian, rumusan masalah penelitian, batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian, dan sistematika penulisan penelitian.

## BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab tinjauan pustaka dan landasan teori berisikan mengenai penelitian terdahulu dan teori-teori dasar yang terkait dengan penelitian ini. Teori yang digunakan terdiri dari *text mining*, *topic modeling*, metode *Latent Dirichlet Allocation* dan *Latent Semantic Analysis*, *text preprocessing*, *N Gram*, *Dictionary*, *Corpus* dan *Python*.

## BAB III METODE PENELITIAN

Bab metode penelitian berisi tentang penjelasan mengenai metode ataupun algoritma yang digunakan serta tahapan-tahapan yang dilakukan untuk mencapai tujuan dan kesimpulan tugas akhir.

## BAB IV HASIL DAN PEMBAHASAN

Bab hasil dan pembahasan membahas analisis data dan hasil dari penelitian yang telah dilakukan.

## BAB V PENUTUP

Bab penutup berisi tentang kesimpulan dari hasil penelitian yang telah dilakukan. Selanjutnya, kekurangan yang ada pada

penelitian dituliskan pada saran untuk pengembangan penelitian di masa yang akan datang.



## **BAB V**

### **PENUTUP**

#### **5.1. Kesimpulan**

Berdasarkan penelitian yang telah dilakukan tentang Analisis Komparasi Pemodelan Topik *Latent Dirichlet Allocation* dan *Latent Semantic Analysis*, dapat disimpulkan sebagai berikut.

1. *Latent Dirichlet Allocation* (LDA) dan *Latent Semantic Analysis* (LSA) sama-sama menghasilkan model topik yang baik, namun luaran model terbaik dihasilkan oleh LDA. Dalam studi kasus ini, LDA terbukti bekerja lebih baik dari LSA.
2. Tujuh topik menjadi topik luaran terbaik, yaitu topik 1 membahas tentang “harga”, topik 2 membahas tentang “suasana”, topik 3 membahas tentang “menu”, topik 4 membahas tentang “keunikan”, topik 5 membahas tentang “lesehan”, topik 6 membahas tentang “*instagramable*”, dan topik 7 membahas tentang “nongkrong”.

#### **5.2. Saran**

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

1. Penelitian selanjutnya dapat melakukan percobaan dengan menggunakan data corpus yang lebih besar, dan sasaran data dari semua rating.

2. Membandingkan kembali dengan metode pembelajaran mesin lainnya untuk menemukan topik-topik yang lebih bermakna.
3. Memberi nama topik dari luar persepsi peneliti, hal ini bisa dicoba dengan mencari vektor centroid, seperti pada cara kerja *clustering*.





## DAFTAR PUSTAKA

- Aletras, N., & Stevenson, M. (2013). Evaluating topic coherence using distributional semantics. *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)--Long Papers*, 13–22.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- Blei, D., Carin, L., & Dunson, D. (2010). Probabilistic topic models. *IEEE Signal Processing Magazine*, 27(6), 55–65.  
<https://doi.org/10.1109/MSP.2010.938079>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(4–5), 993–1022.  
<https://doi.org/10.1016/b978-0-12-411519-4.00006-9>
- Chusen, R. (2018). Topic Modeling with LSA and LDA | Forest for The Tree. Retrieved April 18, 2020, from <https://forestforthetree.com/statistics/2018/01/28/topic-modelling-with-lsa-and-lda.html>
- Douven, I., & Meijs, W. (2007). Measuring coherence. *Synthese*, 156, 405–425.  
<https://doi.org/10.1007/s11229-006-9131-z>
- Google Maps - Wikipedia bahasa Indonesia, ensiklopedia bebas. (n.d.). Retrieved March 9, 2020, from [https://id.wikipedia.org/wiki/Google\\_Maps#Referensi](https://id.wikipedia.org/wiki/Google_Maps#Referensi)
- ilmu data - Sarah. (n.d.). Retrieved April 7, 2020, from <https://stackoverflow.com/questions/54762690/coherence-score-0-4-is-good-or-bad/55816086>

- Kuhlman, D. (2009). *A python book: Beginning python, advanced python, and python exercises*. Dave Kuhlman Lutz.
- Landauer, T. K., & Dumais, S. T. (1997). *A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge*.
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262–272.
- Mohammed, S., & Al-augby, S. (2020). *LSA & LDA Topic Modeling Classification: Comparison study on E-books*. 2502–4752. <https://doi.org/10.11591/ijeecs.v19.i1.pp%25p>
- Putra, K. B., & Kusumawardani, R. P. (2017). Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA). *Jurnal Teknik ITS*, 6(2). <https://doi.org/10.12962/j23373539.v6i2.23205>
- Python – The new generation Language - GeeksforGeeks. (n.d.). Retrieved March 9, 2020, from <https://www.geeksforgeeks.org/python-the-new-generation-language/>
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408.
- Rosen, G. (n.d.). *Improving Restaurant Recommendations on Yelp*. 1–8.

- Tong, Z., & Zhang, H. (2016). A Text Mining Research Based on LDA Topic Modelling. *Computer Science & Information Technology*, 6, 201–210. <https://doi.org/10.5121/csit.2016.60616>
- Tri Fadhila, A. (2017). KLASIFIKASI KATEGORI BAHASA INDONESIA MENGGUNAKAN ALGORITMA LSA. *Simki-Techsin*, 01(01), 1–7.
- Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16.
- Zulhanif. (2016). Pemodelan Topik Dengan Latent Dirichlet Allocation. *Seminar Nasional Pendidikan Matematika*, 1–8.