

**ANALISIS TOPIC MODELLING PARIWISATA YOGYAKARTA  
MENGUNAKAN LATENT DIRICHLET ALLOCATION (LDA)**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**URAY NUR KHADIJAH**

**22.21.1571**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2024**

**ANALISIS TOPIC MODELLING PARIWISATA YOGYAKARTA  
MENGUNAKAN LATENT DIRICHLET ALLOCATION (LDA)**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**URAY NUR KHADIJAH**

**22.21.1571**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2024**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**ANALISIS TOPIC MODELLING PARIWISATA YOGYAKARTA  
MENGUNAKAN LATENT DIRICHLET ALLOCATION (LDA)**

yang disusun dan diajukan oleh

**Uray Nur Khadijah**

**22.21.1571**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 29 Februari 2024

**Dosen Pembimbing,**



**Nurf Cahyono, M.Kom**  
**NIK. 190402278**

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS TOPIC MODELLING PARIWISATA YOGYAKARTA  
MENGUNAKAN LATENT DIRICHLET ALLOCATION (LDA)  
yang disusun dan diajukan oleh

Uray Nur Khadijah

22.21.1571

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 29 Februari 2024

Nama Penguji	Susunan Dewan Penguji	Tanda Tangan
<u>Muhammad Ainul Fikri, S.T., M.Eng</u> NIK. 190302680		
<u>Anna Baiqa, M.Kom</u> NIK. 190302290		
<u>Nuri Cahyono, M.Kom</u> NIK. 190302278		

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 29 Februari 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.  
NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Uray Nur Khadijah  
NIM : 22.21.1571

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS TOPIC MODELLING PARIWISATA YOGYAKARTA  
MENGUNAKAN LATENT DIRICHLET ALLOCATION (LDA)**

Dosen Pembimbing : Nuri Cahyono, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 29 Februari 2024

Yang Menyatakan,



Uray Nur Khadijah

## HALAMAN PERSEMBAHAN

Alhamdulillahirobbil'alamiin, segala puji bagi Allah Subhanahu Wa Ta'ala yang telah mencurahkan rahmat dan karunia-Nya, sehingga Peneliti mampu menyelesaikan Skripsi dengan judul "Analisis Topic Modelling Pariwisata Yogyakarta Menggunakan *Latent Dirichlet Allocation (LDA)*" ini dengan baik. Shalawat serta Salam turunkan kepada kekasih Allah yakni baginda Rasulullah Muhammad Shalallahu Alaihi Wassalam yang telah membawa kita dari jaman yang penuh dengan kebodohan menuju jaman yang penuh dengan santapan ilmu seperti yang kita rasakan sekarang ini. Semoga doa dan shalawat dapat menggiring kita semua menuju jalan yang lebih baik dan benar, baik di dunia maupun akhirat. Aamiin Ya Rabbal Alamin.

Karya ini saya persembahkan untuk:

1. Ayah dan Ibu serta nenek tercinta, yang sampai saat ini masih terus memberikan doa, dukungan, serta motivasi
2. Seluruh keluarga saya yang selalu memberikan segala dukungan, doa, serta semangat.
3. Kepada bapak Nuri Cahyono, M.Kom yang telah memberikan bimbingan dalam skripsi ini serta memberikan ilmu baru dalam pengerjaan skripsi.
4. Kepada teman-teman transfer Angkatan 22 dan adek-adek kelas yang lainnya, terima kasih banyak sudah menemani serta membantu saya mulai dari awal perkuliahan hingga Skripsi ini selesai disusun.
5. Terima kasih juga teruntuk sahabat-sahabat tercinta yang sudah support system, memberikan semangat serta doa terbaik kalian.
6. Kepada keluarga besar munzalan terutama pondok tan isleman terimakasih telah memberikan doa terbaik kalian dan support dalam doa.

## KATA PENGANTAR

Alhamdulillah, berkat Rahmat, Hidayah dan Ridho Allah SWT penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini merupakan tugas akhir dari penulis dalam memenuhi persyaratan untuk mencapai gelar Sarjana Komputer (S.Kom).

Oleh karena itu, penulis ingin menyampaikan rasa hormat serta rasa terimakasih kepada banyak pihak yang telah membantu penulis dalam dukungan, bimbingan serta motivasinya terkhususnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta.
2. Bu Windha Mega Pradnya D, M.Kom selaku ketua program studi SI Informatika.
3. Pak Nuri Cahyono, M.Kom sebagai dosen pembimbing yang baik, Ikhlas, ramah dan humble serta memberikan solusi sehingga permasalahan atas kesulitan dalam penulisan skripsi ini dapat terselesaikan.
4. Kedua orang tua saya, nenek dan kakek, bibi dan paman saya serta semua keluarga yang senantiasa turut mendoakan, memberikan dukungan dan semangat dalam menjalani kuliah dan menyelesaikan skripsi ini.
5. Seluruh sahabat-sahabat dan calon saya yang ikut menemani jalan proses kehidupan perkuliahanku selama di Amikom Yogyakarta, terimakasih berkat kalian penulis semakin terdorong untuk menyelesaikan skripsi.
6. Penulis menyadari sepenuhnya, bahwa penulisan skripsi ini jauh dari sempurna, oleh sebab itu penulis mengharapkan kritik dan saran yang membangun untuk membantu pembuatan skripsi yang lebih baik.

Yogyakarta, 2024

Uray Nur Khadijah

## DAFTAR ISI

HALAMAN PERSETUJUAN .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iv
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI .....	vii
DAFTAR TABEL .....	ix
DAFTAR GAMBAR .....	x
DAFTAR LAMPIRAN .....	xi
INTISARI .....	xii
<i>ABSTRACT</i> .....	xiii
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan Penelitian .....	2
1.5 Manfaat Penelitian .....	3
1.6 Sistematika Penulisan .....	3
<b>BAB II TINJAUAN PUSTAKA</b> .....	4
2.1 Studi Literatur .....	4
2.2 Dasar Teori .....	15
<b>BAB III METODE PENELITIAN</b> .....	22
3.1 Alur Penelitian .....	22
3.2 Tahapan penelitian .....	23
3.2.1 Pengambilan Data .....	23
3.2.2 Text Preprocessing .....	23
3.2.3 Latent Dirichlet Allocation (LDA) .....	24



3.2.4	Perplexity .....	25
3.2.5	Coherence .....	26
3.3	Alat dan Bahan.....	27
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>28</b>
4.1	Pengumpulan Data.....	28
4.2	Text Processing.....	29
4.3	Hasil pengolahan.....	34
4.4	Evaluasi Topic Coherence dan Perplexity .....	39
4.5	Visualisasi data teks.....	41
<b>BAB V PENUTUP .....</b>		<b>42</b>
5.1	Kesimpulan.....	42
5.2	Saran .....	42
<b>REFERENSI .....</b>		<b>43</b>
<b>LAMPIRAN HASIL PREPROCESSING .....</b>		<b>46</b>
<b>LAMPIRAN PERPLEXITY .....</b>		<b>47</b>
<b>LAMPIRAN TOPIC COHERENCE .....</b>		<b>48</b>

## DAFTAR TABEL

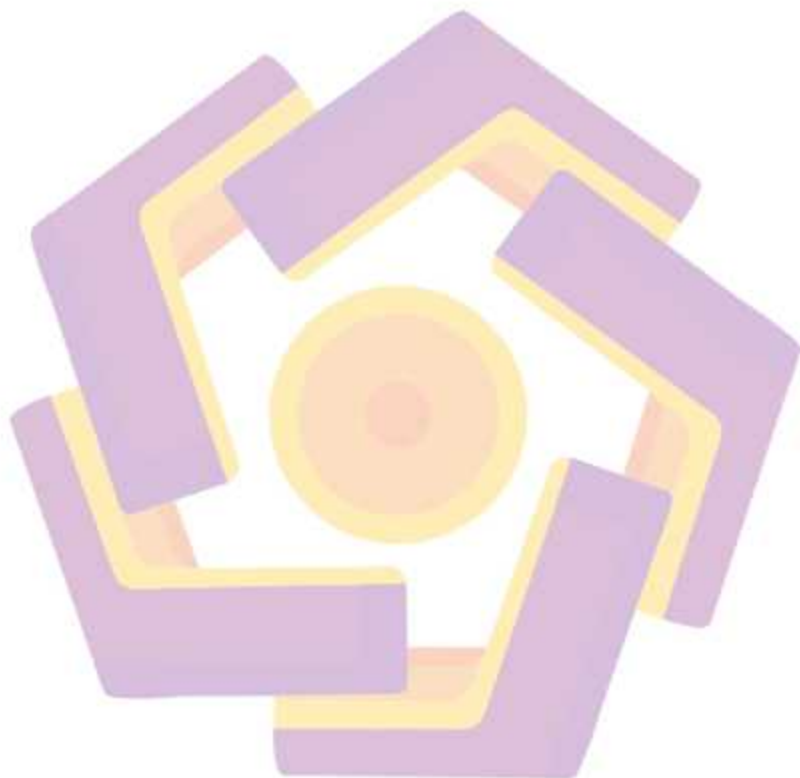
Tabel 2. 1 Keaslian Penelitian .....	8
Tabel 4. 1 Data Penulisan .....	29
Tabel 4. 2 <i>Casefolding</i> .....	30
Tabel 4. 4 Tokenaizing .....	31
Tabel 4. 5 Normalization .....	32
Tabel 4. 6 Stopword Removal .....	33
Tabel 4. 7 Stemming .....	33
Tabel 4. 8 Model LDA Topic 1 .....	35
Tabel 4. 9 Model LDA Topic 2 .....	36
Tabel 4. 10 Model LDA Topic 3 .....	37
Tabel 4. 11 Nilai Topic Coherence .....	39
Tabel 4. 12 Nilai Perplexity .....	40

## DAFTAR GAMBAR

Gambar 2. 1 Alur Kerja Preprocessing .....	16
Gambar 2. 2 Plate Notation Metode LDA .....	19
Gambar 3. 1 Alur Penelitian .....	22
Gambar 4. 1 File Pariwisatabaru.csv .....	28
Gambar 4. 2 Hasil Preprocessing csv .....	34
Gambar 4. 3 Diagram Topik 1 .....	35
Gambar 4. 4 Wordcloud Topik 1 .....	36
Gambar 4. 5 Diagram Topik 2 .....	37
Gambar 4. 6 Wordcloud Topik 2 .....	37
Gambar 4. 7 Diagram Topik 3 .....	38
Gambar 4. 8 Wordcloud Topik 3 .....	38
Gambar 4. 9 Topic Coherence .....	39
Gambar 4. 10 Perplexity .....	40
Gambar 4. 11 Visualisasi data teks word cloud.....	41

## DAFTAR LAMPIRAN

Lampiran 1 Tabel Hasil Preprocessing.....	46
Lampiran 2 Lampiran Hasil Nilai Perplexity.....	47
Lampiran 3 Lampiran Hasil Nilai Topic Coherence.....	48



## INTISARI

Pariwisata Yogyakarta merupakan destinasi yang kaya akan budaya, sejarah, dan keindahan alam. Dengan pertumbuhan pesat penggunaan media sosial, data dari *Platform* seperti *Twitter* menjadi sumber informasi yang berharga untuk memahami opini dan tren terkini. Pariwisata Yogyakarta, sebagai destinasi yang kaya akan budaya dan sejarah, sering menjadi fokus diskusi di media sosial. Tujuan dari Penelitian ini adalah menelaah topik pariwisata Yogyakarta dari *Twitter*. Dataset yang diperoleh dalam penelitian ini dari *crawling* data menggunakan *API key Twitter*. Penelitian ini ingin melihat topik apa saja yang sering masyarakat bicarakan mengenai pariwisata Yogyakarta pada masing-masing kelas di media sosial *Twitter* dalam periode September 2021 – Januari 2024. Penelitian ini menggunakan tahapan dari pengumpulan data, text preprocessing, dan menerapkan metode *Topic Modelling*, khususnya *Latent Dirichlet Allocation (LDA)*, pada dataset *Twitter* untuk mengungkap topik-topik utama terkait pariwisata Yogyakarta. Hasil penelitian ini pengujian kinerja pemodelan topik dengan metode *LDA* dapat dilihat dari nilai *coherence score*, semakin tinggi nilai *coherence* suatu topik, semakin mudah diinterpretasikan oleh manusia dan *Perplexity* merupakan salah satu standar pengukuran yang dapat digunakan untuk menilai kinerja model yang baik dari model tersebut ditunjukkan dengan nilai *perplexity* yang lebih rendah. Nilai *coherence score* yang ditunjukkan pada num topic ke-1 sebesar 0.331047, untuk nilai *perplexity* ditunjukkan dengan nilai yang tinggi terletak pada num topic ke-3 sebesar -8.830172565520245. diharapkan dapat memberikan wawasan mendalam tentang topik-topik yang sering dibahas dan berkonsentrasi pada penerapan sistem pemodelan topik untuk membangun sistem keputusan topik berita yang menggunakan metode *Latent Dirichlet Allocation (LDA)*. Pada Penelitian ini efektif dalam menggunakan metode *LDA* untuk menentukan topik berita yang mencakup tiga kategori topik yang sering dibicarakan pada masing-masing kelas.

**Kata kunci:** Yogyakarta, media social, *Platform*, *Twitter*, *Topic Modelling*, *Latent Dirichlet Allocation (LDA)*, *crawling*, *coherence score*, *perplexity*

## ***ABSTRACT***

Yogyakarta Tourism is a destination rich in culture, history, and natural beauty. With the rapid growth of social media usage, data from platforms such as Twitter is a valuable source of information to understand current opinions and trends. Yogyakarta tourism, as a destination rich in culture and history, is often the focus of discussion on social media. The purpose of this research is to examine the topic of Yogyakarta tourism from Twitter. The dataset obtained in this research is from crawling data using Twitter API keys. This research wants to see what topics people often talk about Yogyakarta tourism in each class on Twitter social media in the period September 2021 - January 2024. This research uses the stages of data collection, text preprocessing, and applying the Topic Modeling method, specifically Latent Dirichlet Allocation (LDA), to the Twitter dataset to reveal the main topics related to Yogyakarta tourism. The results of this study test the performance of topic modeling with the LDA method can be seen from the coherence score value, the higher the coherence value of a topic, the easier it is to be interpreted by humans and Perplexity is one of the measurement standards that can be used to assess the performance of a good model of the model indicated by a lower perplexity value. The coherence score value shown in the 1st num topic is 0.331047, for the perplexity value shown with a high value located in the 3rd num topic of -8.830172565520245. is expected to provide deep insight into topics that are often discussed and concentrate on the application of topic modeling systems to build news topic decision systems that use the Latent Dirichlet Allocation (LDA) method. This research is effective in using the LDA method to determine news topics that include three categories of frequently discussed topics in each class.

***Keyword:*** Yogyakarta, social media, Platform, Twitter, Topic Modeling, Latent Dirichlet Allocation (LDA), *crawling, coherence score, perplexity.*