

**ANALISIS TOPIC MODELLING PARIWISATA BANTUL MENGGUNAKAN
LATENT DIRICHLET ALLOCATION (LDA)**

SKRIPSI

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



disusun oleh
HENDRA WASI SANJAYA
22.21.1558

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2024

**ANALISIS TOPIC MODELLING PARIWISATA BANTUL MENGGUNAKAN
LATENT DIRICHLET ALLOCATION (LDA)**

SKRIPSI

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



disusun oleh

HENDRA WASI SANJAYA

22.21.1558

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS TOPIC MODELLING PARIWISATA BANTUL MENGGUNAKAN
LATENT DIRICHLET ALLOCATION (LDA)**

yang disusun dan diajukan oleh

Hendra Wasi Sanjaya
22.21.1558

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 1 April 2024

Dosen Pembimbing,


Theopilus Bayu Sasongko, S.Kom, M.Eng
NIK. 190302375

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS TOPIC MODELLING PARIWISATA BANTUL MENGGUNAKAN
LATENT DIRICHLET ALLOCATION (LDA)**

yang disusun dan diajukan oleh

Hendra Wasi Sanjaya

22.21.1558

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Juli 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dr. Ferry Wahyu Wibowo, S.Si., M. Cs.
NIK. 190302235

Arifiyanto Hadinegoro, S. Kom, MT
NIK. 190302289

Theopilus Bayu Sasongko, S. Kom, M. Eng
NIK. 190302375



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Juli 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 : **Hendra Wasi Sanjaya**
NIM : **22.21.1558**

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS TOPIC MODELLING PARIWISATA BANTUL MENGGUNAKAN LATENT DIRICHLET ALLOCATION (LDA)

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom, M.Eng

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, 23 Juli 2024

Yang Menyatakan,



Hendra Wasi Sanjaya

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 tercurahkan 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. Diri saya sendiri, karena telah membuktikan bahwa bisa sejauh ini
2. Ayah dan Ibu, yang sampai saat ini masih terus memberikan doa, dukungan, serta motivasi
3. Seluruh keluarga saya yang selalu memberikan segala dukungan, doa, serta semangat.
4. Kepada bapak Theopilus Bayu Sasongko, S.Kom, M.Eng yang telah memberikan bimbingan dalam skripsi ini serta memberikan ilmu baru dalam pengerjaan skripsi.
5. Kepada teman-teman transfer Angkatan 22, terima kasih banyak sudah menemani serta membantu saya mulai dari awal perkuliahan hingga Skripsi ini selesai disusun.

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 S1 Informatika.
3. Pak Theopilus Bayu Sasongko, S.Kom, M.Eng sebagai dosen pembimbing yang baik, Ikhlas, ramah dan humble serta memberikan solusi sehingga permasalahan ataskesulitan dalam penulisan skripsi ini dapat terselesaikan.
4. Kedua orang tua saya serta semua keluarga yang senantiasa turut mendoakan, memberikan dukungan dan semangat dalam menjalani kuliah dan menyelesaikan skripsi ini.
5. Seluruh teman teman 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

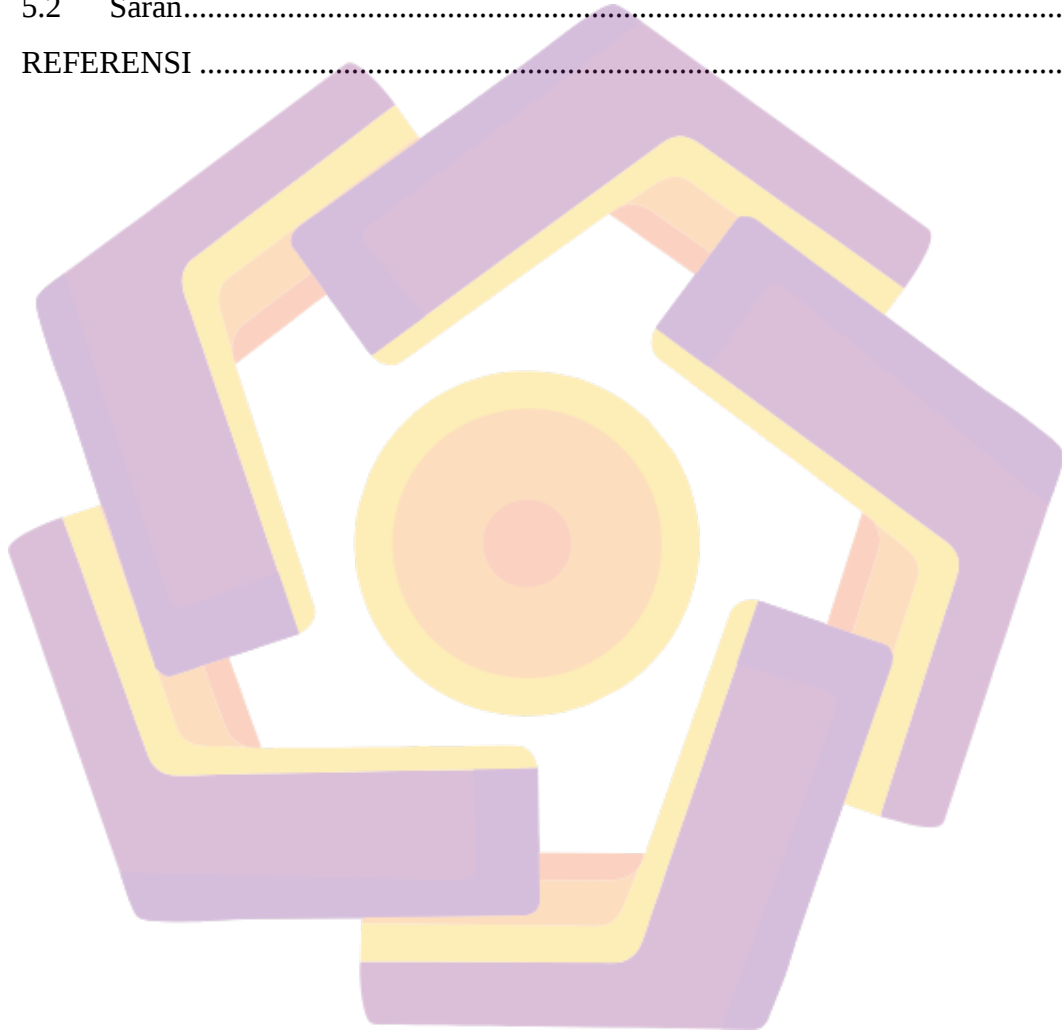
Hendra Wasi Sanjaya

DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
INTISARI.....	xii
<i>ABSTRACT</i>	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	17
2.2.1 Analisis Identifikasi Masalah.....	17
2.2.2 Mempelajari Literatur	17
2.2.3 Pengumpulan data.....	17
2.2.4 Twitter.....	17
2.2.5 Text mining.....	18
2.2.6 Preprocessing	18

2.2.7	Topic Modeling.....	21
2.2.8	LDA	21
2.2.9	Topik coherence.....	23
2.2.10	Perplexity	24
BAB III METODE PENELITIAN.....		25
3.1	Alur Penelitian	25
3.2	Studi Pustaka.....	25
3.3	Analisis Masalah.....	26
3.4	Implementasi Hasil.....	26
3.4.1	Pengumpulan Data.....	27
3.4.2	Text Preprocessing.....	27
3.4.3	Penentuan Cluster	28
3.4.4	Pembuatan Model Latent Dirichlet Allocation.....	30
3.5	Analisa Hasil	31
3.6	Dokumentasi	31
3.7	Alat dan Bahan.....	31
3.7.1	Alat Penelitian.....	32
3.7.2	Bahan Penelitian	33
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Pengumpulan Data	34
4.2	Text Preprocessing.....	35
4.2.1	Casefolding	36
4.2.2	Remove Punctuation.....	37
4.2.3	Tokenizing	38
4.2.4	Normalization	38
4.2.5	Stemming.....	39
4.2.6	Stopword Removing	40
4.3	Penentuan Cluster.....	41

4.4	Pembuatan Model LDA	45
4.5	Visualisasi Data Teks	49
BAB V PENUTUP.....		50
5.1	Kesimpulan	50
5.2	Saran.....	50
REFERENSI		52



DAFTAR GAMBAR

Gambar 2. 1 Proses Teks Mining.....	18
Gambar 2. 2 Alur Kerja Preprocessing	19
Gambar 2. 3 Plate Notation Metode LDA	22
Gambar 3. 1 Alur Penelitian.....	25
Gambar 3. 2 Implementasi Hasil.....	27
Gambar 3. 3 Text Preprocessing	27
Gambar 3. 4 Grafik Model LDA.....	30
Gambar 4. 1 Sample Data PariwisataBantul.csv.....	34
Gambar 4. 2 Topic Coherence	43
Gambar 4. 3 Grafik Perplexity	44
Gambar 4. 4 Wordcloud Topik 1	47
Gambar 4. 5 Wordcloud Topik 2	48
Gambar 4. 6 Visualisasi Wordcloud	49

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	9
Tabel 3. 1 Rentang Waktu Kegiatan Penelitian	31
Tabel 4. 1 Sample Full Text.....	35
Tabel 4. 2 Sample Casefolding	36
Tabel 4. 3 Sample Remove Punctuation.....	37
Tabel 4. 4 Sample Tokenizing	38
Tabel 4. 5 Sample Normalization.....	39
Tabel 4. 6 Sample Stemming	40
Tabel 4. 7 Sample Stopword Removal.....	41
Tabel 4. 8 Sampel Hasil Preprocessing.....	42
Tabel 4. 9 Nilai Koherensi Topik.....	43
Tabel 4. 10 Nilai Perplexity	45
Tabel 4. 11 Model LDA Topik 1	46
Tabel 4. 12 Model LDA Topik 2	47

INTISARI

Pariwisata merupakan salah satu sektor penting dalam perekonomian Indonesia. Kabupaten Bantul, yang terletak di Yogyakarta, terkenal dengan kekayaan wisatanya, mulai dari wisata alam, budaya, hingga kuliner.

Di era serba digital ini, media sosial menjadi platform yang banyak digunakan wisatawan untuk mencari informasi dan berbagi pengalaman tentang wisata. Twitter atau yang sekarang berganti nama menjadi X, sebagai salah satu platform media sosial populer, memiliki banyak pengguna yang aktif berbagi informasi salah satunya tentang wisata Bantul.

Penelitian ini bertujuan untuk menganalisis topik wisata Bantul yang populer di Twitter dengan menggunakan metode Latent Dirichlet Allocation (LDA). Dataset menggunakan hashtag #BantulTourism, #VisitBantul, #WisataBantul, dll, yang terkait pariwisata Bantul untuk diidentifikasi topik-topik utamanya.

Penelitian ini diharapkan dapat memberikan manfaat bagi dua pihak: wisatawan dan pemerintah Kabupaten Bantul. Bagi wisatawan, hasil penelitian ini dapat membantu mereka dalam memilih tempat wisata yang sesuai dengan minat dan preferensi mereka. Bagi pemerintah Kabupaten Bantul, hasil penelitian ini dapat menjadi masukan berharga dalam mengembangkan strategi dan kebijakan untuk meningkatkan sektor pariwisata di wilayahnya.

Kata Kunci: Pariwisata, Bantul, Topic Modelling, Latent Dirichlet Allocation (LDA), Twitter, Analisis Data, Minat Wisatawan, Pengembangan Pariwisata, Strategi Pariwisata, Kebijakan Pariwisata

ABSTRACT

Tourism is one of the crucial sectors in the Indonesian economy. Bantul Regency, located in Yogyakarta, is well-known for its diverse tourism attractions, ranging from natural, cultural, to culinary experiences.

In this digital era, social media has become a widely used platform for tourists to seek information and share their travel experiences. Twitter, now rebranded as X, is one of the popular social media platforms with many active users sharing information, including about tourism in Bantul.

This research aims to analyze popular tourism topics in Bantul on Twitter using the Latent Dirichlet Allocation (LDA) method. The dataset utilizes hashtags related to Bantul tourism, such as #BantulTourism, #VisitBantul, #WisataBantul, etc., to identify the main topics.

This research is expected to benefit two parties: tourists and the Bantul Regency government.

For tourists, the research results can assist them in selecting tourist destinations that match their interests and preferences. For the Bantul Regency government, the research outcomes can serve as valuable input for developing strategies and policies to improve the tourism sector in its region.

Keyword: *Tourism, Bantul Regency, Topic Modelling, Latent Dirichlet Allocation (LDA), Twitter, Data Analysis, Tourist Preferences, Tourism Development, Policy Maki*