

## Pemodelan Topik pada Ulasan Google Maps Candi Borobudur Menggunakan Latent Dirichlet Allocation

Reonardh<sup>1\*</sup>, Sunneng Sandino Berutu<sup>2</sup>, Kristian Juri Damai Lase<sup>3</sup>, Jatmika<sup>4</sup>

<sup>1,2,3,4</sup> Universitas Kristen Immanuel Yogyakarta, Indonesia

Jl. Solo Km. 11,1 Yogyakarta, Indonesia

\*Corresponding author e-mail : reonardh2042@student.ukrimuniversity.ac.id

### ABSTRAK

Penelitian ini bertujuan untuk menggali wawasan yang mendalam dari ulasan *Google Maps* wisatawan tentang Candi Borobudur menggunakan metode Pemodelan Topik *Latent Dirichlet Allocation* (LDA). Candi Borobudur, salah satu situs warisan dunia yang terkenal di Indonesia, menerima banyak ulasan dari wisatawan yang berbeda-beda. Namun, untuk mengidentifikasi tren, preferensi, dan aspek yang paling penting bagi pengunjung, diperlukan analisis yang lebih dalam. Dengan menggunakan pendekatan LDA, mengelompokkan ulasan-ulasan ini kedalam topik-topik latensial yang memungkinkan untuk mengeksplorasi pola-pola umum dan perbedaan-perbedaan dalam persepsi dan pengalaman pengunjung terhadap Candi borobudur. Proses pemodelan topik ini mencakup beberapa langkah, dimulai dari persiapan data, pra-pemrosesan data, wordcloud, topic coherence serta pemodelan topik. Adapun nilai koherensi paling baik diperoleh dalam model *Latent dirichlet allocation* (LDA) adalah topik terbaik dengan jumlah 4 topik. Sedangkan dari hasil analisis pemodelan topik pada metode *Latent dirichlet allocation* (LDA) pada Kata "candi" (0.050883695), "borobudur" (0.02406518), "tiket" (0.018967053), dan "masuk" (0.018507887), menyoroti fokus pada aspek kuil, aksesibilitas, dan biayamasuk. Kata "jalan" (0.012353696) dan "parkir" (0.011362941) menunjukkan perhatian terhadap infrastruktur, sementara "area" (0.01039175) mengacu pada lingkungan sekitar candi. Istilah "rb" (0.009597249) dan "harga" (0.0092093265) merujuk pada biaya dalam rupiah, dan "beli" (0.008586656) terkait dengan pembelian tiket. Hasil ini memberikan wawasan tentang fokus utama dalam pembahasan Borobudur, mencakup kuil, aksesibilitas, infrastruktur, dan biaya.

**Kata kunci :** Google Maps, Pemodelan Topik, *Latent Dirichlet Allocation* (LDA), Nilai Koherensi

### ABSTRACT

*This research aims to extract deep insights from tourists' Google Maps reviews of Borobudur Temple using the Latent Dirichlet Allocation (LDA) Topic Modeling method. Borobudur Temple, one of the famous world heritage sites in Indonesia, receives many reviews from different tourists. However, to identify trends, preferences, and aspects that are most important to visitors, a deeper analysis is required. Using the LDA approach, categorizing these reviews into latent topics makes it possible to explore common patterns and differences in visitors' perceptions and experiences of Borobudur Temple. This topic modeling process includes several steps, starting from data preparation, data pre-processing, wordcloud, topic coherence and topic modeling. The best coherence value obtained in the Latent dirichlet allocation (LDA) model is the best topic with a total of 4 topics. While based on the results of the analysis of topic modeling with the Latent dirichlet allocation (LDA) method on Borobudur reviews get 4 topic models The best coherence value obtained in the Latent dirichlet allocation (LDA) model is the best topic with a total of 4 topics. While from the results of the topic modeling analysis on the Latent dirichlet allocation (LDA) method on the words "temple" (0.050883695), "borobudur" (0.02406518), "ticket" (0.018967053), and "entrance" (0.018507887), highlighting the focus on aspects of the temple, accessibility, and entrance fees. The words "road" (0.012353696) and "parking" (0.011362941) indicate attention to infrastructure, while "area" (0.01039175) refers to the temple's surroundings. The terms "rb" (0.009597249) and "harga" (0.0092093265) refer to costs in rupiah, and "beli" (0.008586656) is related to ticket purchases. These results provide insight into the main focus of the Borobudur discussion, covering temples, accessibility, infrastructure, and costs.*

**Keywords:** Google maps, topic modelling, *Latent Dirichlet Allocation* (LDA), Coherence Score.

## I. PENDAHULUAN

Borobudur adalah candi Buddha yang megah dan merupakan salah satu situs bersejarah terpenting di Indonesia, yang terletak di Magelang, Jawa Tengah. Candi yang dibangun sekitar abad ke-9 Masehi ini merupakan salah satu keajaiban arsitektur dan kepercayaan Buddha di Asia Tenggara[1]. Taman Wisata Candi Borobudur masih menjadi destinasi dengan tingkat kunjungan terbanyak di Jawa Tengah dengan jumlah wisatawan sebanyak

63.427 orang. [2]. Dengan popularitasnya yang terus meningkat, jumlah ulasan wisatawan tentang Candi Borobudur di *platform daring* seperti situs *web* perjalanan dan media sosial juga semakin bertambah. Ulasan-ulasan ini tidak hanya mencerminkan pengalaman individu wisatawan, tetapi juga mencakup beragam perspektif, preferensi, dan harapan [3].

Namun, dalam melihat kumpulan ulasan yang besar ini, mengidentifikasi pola umum, tren, dan aspek yang paling penting bagi pengunjung dapat menjadi tugas yang menantang. Latent Dirichlet Allocation (LDA) Topic Modeling menawarkan pendekatan yang ampuh untuk mengekstraksi struktur topik dari kumpulan data teks.[4].

Dengan mengaplikasikan metode ini pada ulasan wisatawan tentang Candi Borobudur, kita dapat mengidentifikasi topik-topik utama yang secara alami muncul dalam percakapan online tentang situs tersebut. Metode ini sangat populer digunakan karena dapat menganalisis dokumen berukuran besar dan membantu mengidentifikasi pola tertentu pada sebuah dokumen dengan cara membangkitkan beberapa jenis topik yang berbeda[5].

Riset ini bertujuan untuk menemukan pola tertentu pada ulasan wisatawan tentang Candi Borobudur. Dengan menggunakan metode LDA, kita dapat meningkatkan pemahaman mengenai apa yang membuat kunjungan ke Candi Borobudur sangat berharga bagi para pengunjung. Melalui pemahaman mengenai topik-topik yang paling sering muncul dalam ulasan, serta hubungan antara topik-topik tersebut, kita dapat meningkatkan pengalaman wisatawan yang berkunjung ke Candi Borobudur [6].

Dengan demikian, hasil analisis ini diharapkan dapat memberikan masukan berharga bagi pihak-pihak yang terlibat dalam pengelolaan dan pemasaran destinasi pariwisata, serta meningkatkan pengalaman wisatawan yang berkunjung ke Candi Borobudur. Perbaikan di atas akan membuat teks Anda lebih ringkas dan lebih mudah dipahami oleh pembaca non-teknis[7].

## II. METODE

Metode penelitian yang disusun dalam penelitian ini mencakup berbagai tahapan, mulai dari Pengumpulan data, pra-pemrosesan data, wordcloud, topic coherence, Pemodelan topik.

### Pengumpulan Data

Fase ini diperlukan untuk memiliki korpus teks yang dapat digunakan untuk menjalankan analisis. Tujuannya adalah untuk mengumpulkan laporan perawatan dan mengubahnya dalam bentuk yang dapat dianalisis. Proses pengambilan data dari *Google review* menggunakan *Web Scraper*[7].

### Pra-pemrosesan data

Fase pra-pemrosesan data memerlukan pelaksanaan aktivitas berikut:

- Cleaning*. proses mengidentifikasi dan menangani masalah pada data seperti nilai yang hilang, duplikat, dan *outlier* untuk memastikan dataset bersih dan siap untuk dianalisis. Langkah-langkah tersebut meliputi pemeriksaan data awal, mengisi atau menghapus nilai yang hilang, menghapus baris duplikat, menangani pencilan, dan normalisasi data[8].
- Casa folding*. Pelipatan huruf adalah bentuk prapemrosesan teks Pelipatan huruf besar-kecil adalah cara yang paling mudah dan efektif, walaupun seringkali terlewatkan. Tujuan pelipatan huruf adalah mengubah seluruh huruf pada dokumen ke dalam huruf kecil. Jenis huruf kecil yang disetujui hanya huruf 'a' hingga 'z'. Huruf selain huruf akan dihapus dan dianggap sebagai pembatas[9].
- Normalisasi. Teknik yang dipakai dalam menciptakan kumpulan entity sesuai karakteristik atribut dari entity tersebut agar sebuah entity yang sederhana dapat terbentuk, tidak redundan, mudah diadaptasi, dan fleksibel, agar dapat dijamin bahwa database tersebut memiliki kualitas yang bagus. Normalisasi seringkali dilakukan pada database untuk meminimalisir redundansi data, memastikan integritas data[10].
- Tokenizing. Tokenizing atau disebut juga dengan Tahap Analisis Leksikal adalah tahap pemotongan teks kedalam bentuk yang lebih kecil, yaitu token. Tahap ini sekaligus menghilangkan angka, karakter tanda baca, dan karakter lain yang dianggap kurang berpengaruh dalam pengolahan data[11].
- Stopword. Stopword salah satu tahap prapemrosesan teks yang berfungsi untuk menghapus kata-kata yang tak relevan dalam kalimat berdasarkan daftar stopwords. Kata-kata

yang akan dihilangkan terdiri dari kata-kata yang paling umum ditemukan dalam bahasa korpus. Kata-kata tambahan dapat didefinisikan dan ditambahkan ke daftar jika perlu[12].

- f. *Stemming data*. Tahap stemming adalah tahap juga dibutuhkan untuk meminimalisir banyaknya index yang berlainan dari sebuah data agar sebuah data yang memiliki akhiran atau awalan akan segera kembali ke wujud dasarnya. Di samping itu, juga untuk mengelompokkan kata-kata lain yang memiliki kata dasar dan makna yang sama tetapi mempunyai bentuk berbeda sebab mendapatkan afiks yang lain[13].

### Word Cloud

*word cloud* sebuah visualisasi yang menampilkan kata-kata dari teks tertentu, dengan jenis huruf yang lebih besar untuk kata-kata yang lebih sering muncul di dalam teks. Awan kata ini dipakai untuk menampilkan kata-kata yang paling penting atau sering muncul dalam teks yang sedang dianalisis, dan dapat dipakai untuk mengevaluasi konten teks, menemukan topik, atau mencari ide baru[14].

### Wordlink

*WordLink* jaringan atau graf yang menghubungkan kata-kata berdasarkan frekuensi kemunculannya bersama dalam sebuah teks. Kata-kata ini diwakili sebagai node dalam graf, dan hubungan antar kata (*co-occurrence*) diwakili oleh edges atau garis penghubung antar node. Ketebalan dan panjang edges dapat menunjukkan kekuatan hubungan atau frekuensi kemunculan bersama kata-kata tersebut[15].

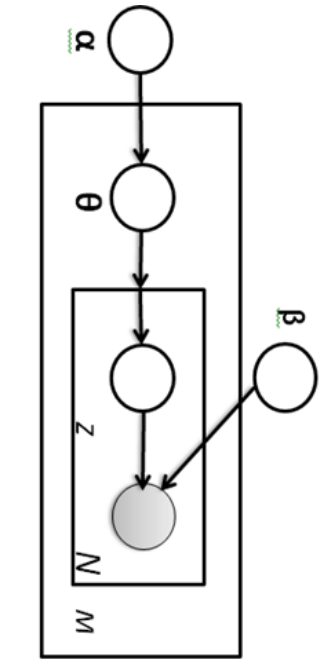
### Topic Coherence

koherensi topik adalah langkah menentukan jumlah model topik dalam pemodelan topik. Makin tinggi nilai *topic coherence*, maka model yang dihasilkan akan semakin baik. Lebih lanjut, Jumlah topik yang memiliki nilai koherensi topik tertinggi digunakan sebagai pemodelan topik. berikut ini adalah hasil grafik pada koherensi topik.

### Pemodelan Topik

Pada tahap ini, algoritma Latent Dirichlet Allocation (LDA) digunakan untuk memodelkan teks. LDA memandang dokumen sebagai campuran dari berbagai topik laten yang *mendasarinya*. Tiap dokumen dalam korpus dianalisis sebagai kombinasi acak dari topik-topik ini, dengan masing-masing

topik memiliki representasi probabilistik.



Gambar 1. Representasi Graphical Model LDA

Tujuan dari pemodelan ini adalah untuk mengidentifikasi jumlah topik yang paling optimal, meminimalkan kebingungan, dan memaksimalkan koherensi antar topik. Pengaturan parameter dan hiperparameter dalam algoritma LDA akan membantu mengoptimalkan model.

### Kelemahan Metode

Meskipun LDA adalah metode yang kuat, ada beberapa keterbatasan yang perlu diperhatikan:

- Keterbatasan Representasi Data:** Pra-pemrosesan seperti stemming dan penghapusan stopwords dapat menghilangkan makna kontekstual penting, sehingga hasil analisis tidak selalu mencerminkan keseluruhan kompleksitas teks asli.
- Hilangnya Makna Kontekstual:** Langkah-langkah pra-pemrosesan, meski bermanfaat untuk menyederhanakan data, berisiko menghilangkan konteks yang kaya dalam teks, terutama yang terkait dengan struktur sintaksis dan hubungan antar kata.
- Ketergantungan pada Visualisasi Terbatas:** Visualisasi seperti word cloud dan WordLink memberikan wawasan dasar tentang frekuensi kata, tetapi tidak mampu menangkap hubungan yang lebih dalam antara topik-topik atau sentimen dalam teks.

Selain itu, meskipun LDA dapat mengidentifikasi topik-topik dalam dokumen,

algoritma ini memiliki keterbatasan dalam menangani kompleksitas hubungan antar topik dan memerlukan penentuan jumlah topik yang optimal secara cermat.

### III. HASIL DAN PEMBAHASAN

Tahap pertama Sebelum mulai melakukan proses analisis adalah melakukan *pre-processing* data. Tahapan ini bertujuan untuk menghapus atau memperbaiki teks mengenai data yang tidak perlu diperlukan. Preprocessing data pada umumnya bersifat dinamis dan berbeda-beda pada setiap data. Dalam penelitian ini, tahap preprocessing data dibagi menjadi 6 tahap, yaitu *cleaning*, *case folding* menghilangkan tanda baca, normalisasi, *tokenize*, *stopword removal* dan *steming data*.

Tabel 1. Pra-pemrosesan data

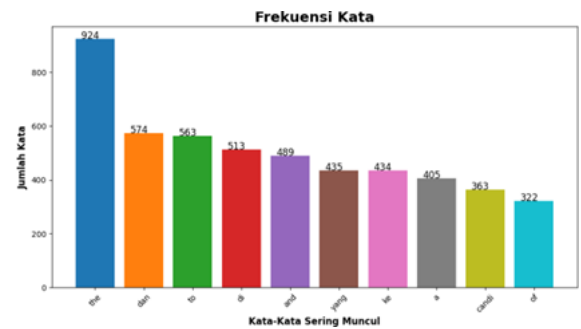
No	Sebelum	Sesudah
<i>Cleaning</i>	tidak diragukan lagi sebagai situs pariwisata kelas internasional, sayangnya fasilitas dan penataannya belum maksimal.	tidak diragukan lagi sebagai situs pariwisata kelas internasional sayangnya fasilitas dan penataannya belum maksimal
<i>Case folding</i>	tidak diragukan lagi sebagai situs pariwisata kelas internasional sayangnya fasilitas dan penataannya belum maksimal	tidak diragukan lagi sebagai situs pariwisata kelas internasional sayangnya fasilitas dan penataannya belum maksimal
<i>Normalisasi</i>	tidak diragukan lagi sebagai situs pariwisata kelas internasional sayangnya fasilitas dan penataannya belum maksimal	tidak diragukan lagi sebagai situs pariwisata kelas internasional sayangnya fasilitas dan penataannya belum maksimal
<i>Tokenization</i>	tidak diragukan lagi sebagai situs pariwisata kelas internasional sayangnya fasilitas dan penataannya belum maksimal	['tidak', 'diragukan', 'lagi', 'sebagai', 'situs', 'pariwisata', 'kelas', 'internasional', 'sayangnya', 'fasilitas', 'dan', 'penataan', 'ya', 'belum', 'maksimal']
<i>Stopword removal</i>	['tidak', 'diragukan', 'lagi', 'sebagai', 'situs', 'pariwisata', 'kelas', 'internasional', 'sayangnya', 'fasilitas', 'dan', 'penataan', 'ya', 'belum', 'maksimal']	'diragukan', 'situs', 'pariwisata', 'kelas', 'internasional', 'sayangnya', 'fasilitas', 'penataan', 'ya', 'maksimal', ragu situs pariwisata kelas internasional sayang fasilitas tata ya maksimal
<i>Stemming data</i>	'diragukan', 'situs', 'pariwisata', 'kelas', 'internasional', 'sayangnya', 'fasilitas', 'penataan', 'ya', 'maksimal',	

### WordCloud

Dari hasil pre-processing dapat di visualisasikan dengan word cloud , kata yang lebih besar merupakan kata yang sering muncul.



Gambar 2. Word cloud

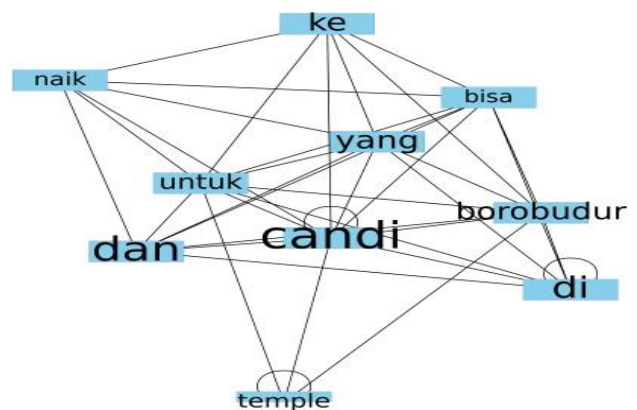


Gambar 3. Plot bar Frekuensi kata

Hasil dari Gambar 2 Terlihat bahwa beberapa kata yang mempunyai tingkat frekuensi tinggi diberi tanda dengan besaran kata yang lebih besar dari kata lain dalam teks, Temple , dan, yang, untuk, candi.

### Wordlink

Hasil dari preprocessing dapat divisualisasikan dalam bentuk *wordlink*. Menunjukkan jaringan kata (*wordlink*) dalam kata terkait candi borobudur.



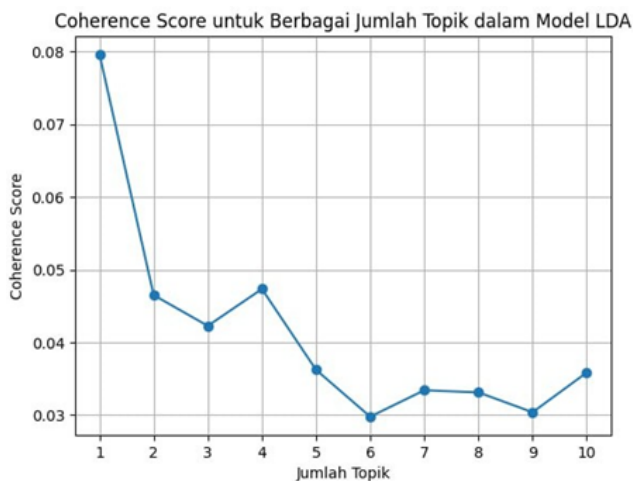
Gambar 4. Wordlink

Pada Gambar 4 menunjukkan jaringan kata kunci (*wordlink*) dalam kata terkait Candi Borobudur, di mana setiap node mewakili kata dan setiap edge menunjukkan frekuensi kemunculan bersama antar kata. Kata "candi" adalah yang paling menonjol, menunjukkan bahwa ini adalah topik utama, dengan kata-kata seperti "Borobudur", "yang", "dan", "di",

"untuk", "ke", "bisa", "naik", dan "temple" terhubung erat, menunjukkan konteks dan hubungan yang berkaitan dengan Candi Borobudur dalam teks. Jaringan ini membantu memahami keterkaitan antar kata dan topik yang dibahas.

### Topic Coherence

*Topic coherence* untuk menentukan jumlah topik terbaik yang akan jadi acuan dalam pemodelan topik.



Gambar 5. Topic coherence

Gambar 5 memperlihatkan puncak koherensi grafik topik berada pada jumlah topik ke-4 qatau dalam kata lain, jumlah topik terbaik yang dibentuk yaitu sebanyak 4 topik.

### Pemodelan Topik

Hasil topic coherence menjadi acuan dalam pemodelan topik menggunakan pendekatan algoritma LDA, yang bertujuan untuk mengidentifikasi dan memahami topic utama , pada topik 1

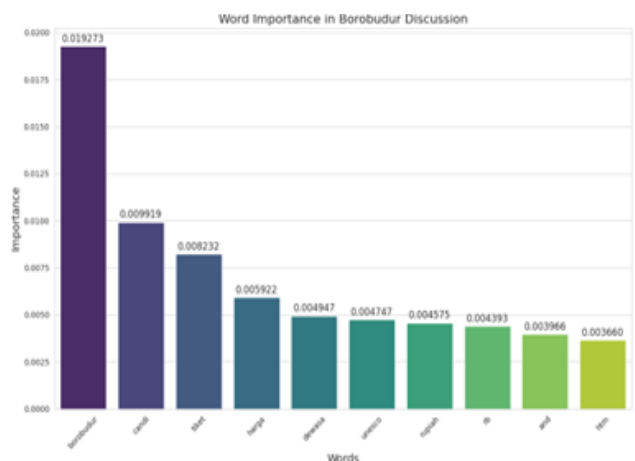
Tabel 2. Model LDA Topik 1

Word	Importance
borobudur	0.019273106
candi	0.009919346
tiket	0.008232462
harga	0.0059215017
dewasa	0.0049471413
unesco	0.004746627
rupiah	0.0045747594
rb	0.0043930835
and	0.003965942
_htm	0.0036600754

Dari tabel diatas menampilkan Kata "borobudur" memiliki nilai tertinggi (0.019273106), menegaskan pusat diskusi. "Candi" (0.009919346) menunjukkan fokus pada aspek kuil. "Tiket" (0.008232462) dan "harga" (0.0059215017) mencerminkan perhatian pada akses dan biaya masuk. Kata "dewasa" dengan nilai 0.0049471413

menunjukkan bahwa terdapat pembahasan khusus mengenai tarif untuk pengunjung dewasa. Kata "unesco" dengan nilai 0.004746627 menyoroti status Borobudur sebagai situs warisan dunia yang dilindungi oleh UNESCO. Satuan mata uang Indonesia, "rupiah", dan singkatannya "rb", juga tercantum dengan nilai masing-masing 0.0045747594 dan 0.0043930835, menekankan pembicaraan tentang biaya dalam konteks mata uang lokal.

Maka model juga dapat ditampilkan dalam bentuk plot bar seperti berikut ini:



Gambar 6. Plot bar topik 1

Dari beberapa data yang didapatkan yaitu pada topik pertama pemodelan LDA ulasan yang sering muncul adalah Candi , borobudurdan ticket. Sehingga dapat disimpulkan bahwa ulasan topik 1 cenderung membahas candi borobudur masuk sejarah keajaiban dunia

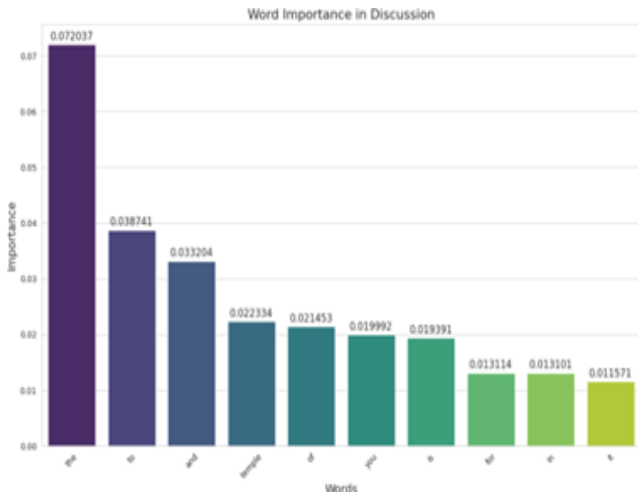
Tabel 3. Model LDA Topik 2

Word	Importance
The	0.07203673
to	0.038741235
and	0.033203796
temple	0.02233378
of	0.021452738
you	0.019991862
is	0.01939113
for	0.013114031
in	0.013100657
it	0.011570966

Dari tabel diatas menampilkan Kata "the" memiliki nilai tertinggi (0.07203673), menunjukkan dominasi dalam penggunaan bahasa Inggris. Kata-kata penghubung seperti "to" (0.038741235), "and" (0.033203796), dan "of" (0.021452738) juga menonjol, mencerminkan struktur kalimat umum. Kata "temple" (0.02233378) menunjukkan fokus pada aspek kuil. Kata "you" (0.019991862) dan "is" (0.01939113) sering digunakan dalam kalimat penjelasan atau deskriptif. Kata "for" (0.013114031)

dan "in" (0.013100657) digunakan dalam konteks penjelasan lokasi atau tujuan. Kata "it" (0.011570966) sering muncul dalam konteks deskriptif.

Maka model juga dapat ditampilkan dalam bentuk plot bar seperti berikut ini:



Gambar 7. Plot bar topik 2

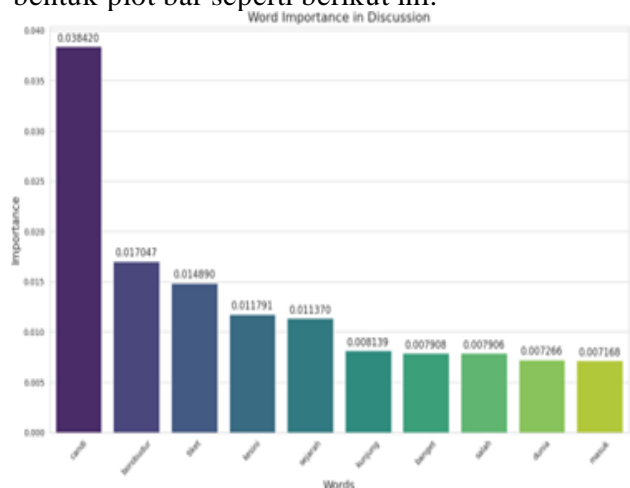
Dari beberapa data yang didapatkan yaitu pada topik kedua pemodelan LDA ulasan yang sering muncul adalah the , to dan and.

#### Model LDA Topic 3

Word	Importance
candi	0.038419504
borobudur	0.017046614
tiket	0.014889521
kesini	0.011791167
sejarah	0.011369621
kunjung	0.008138716
banget	0.007908338
salah	0.0079055615
salah	0.0072655287
dunia masuk	0.007167975

Dari tabel diatas menampilkan Kata "candi" memiliki bobot tertinggi (0.038419504), menekankan fokus utama pada aspek kuil Borobudur. Kata "borobudur" sendiri memiliki bobot 0.017046614, mempertegas pusat diskusi pada situs ini. Kata "tiket" (0.014889521) dan "masuk" (0.007167975) mencerminkan perhatian terhadap aksesibilitas dan biaya masuk. Kata "kesini" (0.011791167) dan "kunjung" (0.008138716) menunjukkan minat dalam aktivitas kunjungan ke Borobudur. Kata "sejarah" (0.011369621) menyoroti pentingnya aspek historis situs ini. Kata "banget" (0.007908338) dan "salah" (0.0079055615) mungkin mencerminkan opini atau reaksi pengunjung terhadap pengalaman mereka di Borobudur. Kata "dunia" (0.0072655287) menegaskan status Borobudur sebagai situs yang diakui secara internasional.

Maka model juga dapat ditampilkan dalam bentuk plot bar seperti berikut ini:



Gambar 8. Plot bar topik 3

Dari beberapa data yang didapatkan yaitu pada topik ketiga pemodelan LDA ulasan yang sering muncul adalah candi , tiket dan borobudur. Sehingga dapat disimpulkan bahwa ulasan topik 3 cenderung membahas anda masuk kedalam candinya.

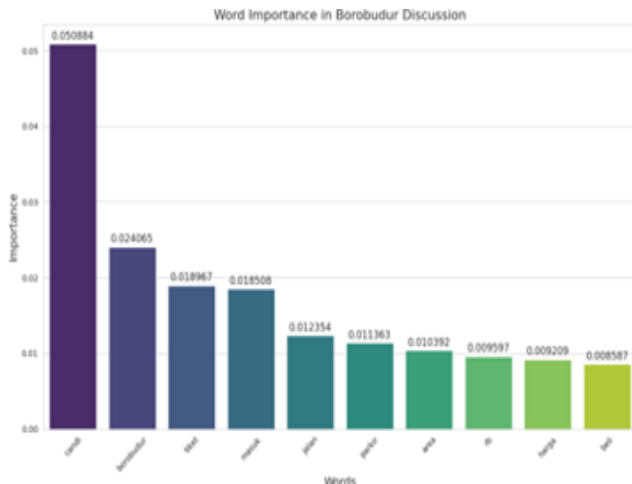
#### Model LDA Topic 4

Word	Importance
candi	0.050883695
borobudur	0.02406518
tiket	0.018967053
masuk	0.018507887
jalan	0.012353696
parkir	0.011362941
area	0.01039175
rb	0.009597249
harga	0.0092093265
beli	0.008586656

Dari tabel diatas menampilkan Kata Kata "candi" memiliki bobot tertinggi (0.050883695), menekankan fokus utama pada aspek kuil Borobudur. Kata "borobudur" sendiri memiliki bobot 0.02406518, mempertegas pusat diskusi pada situs ini. Kata "tiket" (0.018967053) dan "masuk" (0.018507887) mencerminkan perhatian terhadap aksesibilitas dan biaya masuk. Kata "jalan" (0.012353696) dan "parkir" (0.011362941) menunjukkan perhatian terhadap infrastruktur dan fasilitas pendukung di sekitar Borobudur. Kata "area" (0.01039175) mempertegas fokus pada lingkungan dan wilayah sekitar candi. Kata "rb" (0.009597249) dan "harga" (0.0092093265) mengacu pada informasi tentang biaya dalam satuan mata uang lokal, rupiah. Kata "beli" (0.008586656) mencerminkan aktivitas pembelian tiket atau barang terkait kunjungan ke Borobudur.

Maka model juga dapat ditampilkan dalam

bentuk plot bar seperti berikut ini:



Gambar 8. Plot bar topik 4

Dari beberapa data yang didapatkan yaitu pada topik keempat pemodelan LDA ulasan yang sering muncul adalah candi, borobudur dan tiket. Sehingga dapat disimpulkan bahwa ulasan topik 4 cenderung membahas Tiket masuk area candi borobudur

#### IV. KESIMPULAN

Adapun nilai koherensi paling baik yaitu topik terbaik dengan jumlah 4 topik. Sedangkan dari hasil analisis pemodelan topik pada metode Latent dirichlet allocation (LDA) pada Kata "candi" dan "borobudur" memiliki bobot tertinggi, menandakan bahwa pembahasan utama berkisar pada aspek kuil dan situs Borobudur itu sendiri. Kata "tiket" dan "masuk" mengindikasikan perhatian signifikan terhadap aksesibilitas dan biaya masuk ke situs. Ini menunjukkan bahwa informasi mengenai cara masuk dan harga tiket adalah topik yang sering dibahas. Kata "jalan", "parkir", dan "area" menunjukkan bahwa infrastruktur dan fasilitas pendukung di sekitar Borobudur juga menjadi perhatian penting bagi pengunjung. Ini mencakup kondisi jalan menuju lokasi, ketersediaan tempat parkir, dan area sekitar candi. Kata "rb" dan "harga" memperlihatkan bahwa banyak diskusi berkaitan dengan biaya dalam mata uang lokal, rupiah. Ini menegaskan perlunya informasi yang jelas tentang harga-harga yang relevan bagi pengunjung lokal. Kata "beli" mengindikasikan bahwa aktivitas pembelian, baik tiket masuk maupun barang-barang terkait, merupakan bagian penting dari pengalaman pengunjung di Borobudur.

#### V. SARAN

Untuk penelitian selanjutnya, disarankan menggunakan Pldavis untuk visualisasi interaktif model LDA. Pldavis membantu mengevaluasi koherensi topik, mengidentifikasi tumpang tindih, dan menyesuaikan hyperparameter. Gunakan alat ini untuk analisis multibahasa, multisumber, dan

temporal guna memahami perubahan topik seiring waktu. Integrasikan juga analisis sentimen untuk menambah wawasan tentang persepsi topik. Visualisasi hasil dengan Pldavis mempermudah interpretasi dan presentasi temuan, memberikan gambaran yang lebih jelas dan menarik.

#### DAFTAR PUSTAKA

- [1] Y. Afrillia, L. Rosnita, and D. Siska, "Analisis Sentimen Ciutan Twitter Terkait Penerapan Permendikbudristek Nomor 30 Tahun 2021 Menggunakan TextBlob dan Support Vector Machine," *G-Tech: Jurnal Teknologi Terapan*, vol. 6, no. 2, pp. 387–394, Oct. 2022, doi: 10.33379/gtech.v6i2.1778.
- [2] Y. Matira and I. Setiawan, 'Pemodelan Topik pada Judul Berita Online Detikcom Menggunakan Latent Dirichlet Allocation', *Estimasi: Journal of Statistics and Its Application*, vol. 4, no. 1, pp. 2721–379, 2023, doi: 10.20956/ejsa.vi.24843.
- [3] R. Abdillah, E. Haerani, and R. M. Candra, "Analisis Sentimen Ulasan Aplikasi Wetv Untuk Peningkatan Layanan Menggunakan Metode Support Vector Machine," *Journal of Information System Research (JOSH)*, vol. 4, no. 3, pp. 865–873, Apr. 2023, doi: 10.47065/josh.v4i3.3353.
- [4] W. A. Nawang Sari and H. Dwi Purnomo, 'Topic Modeling Using the Latent Dirichlet Allocation Method on Wikipedia Pandemic Covid-19 Data in Indonesia', *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 5, pp. 1223–1230, 2022, doi: 10.20884/1.jutif.2022.3.5.321.
- [5] Sari and Wilujeng Ayu Nawang, "Pemodelan Topik Menggunakan Metode Latent Dirichlet Allocation Pada Data Wikipedia Pandemic Covid-19 di Indonesia", <https://repository.uksw.edu/handle/123456789/26150>.
- [6] M. Taufiq Anwar, D. Riandhita Arief Permana, "Analisis Sentimen Masyarakat Indonesia Terhadap Produk Kendaraan Listrik Menggunakan VADER," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, pp. 783–792, Mar. 2023,
- [7] D. Aryanti, "Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naïve Bayes Classifier," *Journal of Information System Research (JOSH)*, vol. 3, no. 4, pp. 524–531, Jul. 2022, doi: 10.47065/josh.v3i4.1944.
- [8] L. O. Sihombing, H. Hannie, and B. A. Dermawan, "Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier," *Edumatic: Jurnal Pendidikan Informatika*, vol. 5, no. 2, pp. 233–242, Dec.

- 2021, doi: 10.29408/edumatic.v5i2.4089.
- [9] Rianita Giovanni Katryn, 'Text Preprocessing Tahap Awal dalam Natural Language Processing (NLP)', <https://medium.com/mandiri-engineering/text-preprocessing-tahap-awal-dalam-natural-language-processing-nlp-bc5fbb6606a>.
- [10] R. Rinandyaswara, Y. A. Sari, and M. T. Furqon, 'PEMBENTUKAN DAFTAR STOPWORD MENGGUNAKAN TERM BASED RANDOM SAMPLING PADA ANALISIS SENTIMEN DENGAN METODE NAÏVE BAYES (STUDI KASUS: KULIAH DARING DI MASA PANDEMI)', vol. 9, no. 4, 2022, doi: 10.25126/jtiik.202294707.
- [11] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," Jurnal Pengembangan IT (JPIT), vol. 6, no. 3, Jul. 2021.
- [12] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," Jurnal Sains Komputer & Informatika (J-SAKTI), vol. 5, no. 2, pp. 697–711, Sep. 2021.
- [13] Nurdin, K. F., Sutanto, T. E., & Santoso, A. (2023). Analisa Pemodelan Topik Berita Daring Menggunakan Semi-Supervised dan Fully Unsupervised Latent Dirichlet Allocation. Jurnal Pendidikan Tambusai, 7(2), 8043–8053. <https://www.palupos.com>
- [14] W. Sejati, A. Singh Bist, and A. Tambunan, "Pengembangan Analisis Sentimen dalam Rekayasa Software Engineering menggunakan tinjauan literatur sistematis," Jurnal Manajemen Pendidikan dan Teknologi Informasi, vol. 2, no. 1, pp. 95–103, Sep. 2023
- [15] A. Dwiki, A. Putra, and S. Juanita, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN," Jurnal Teknik Informatika dan Sistem Informasi, vol. 8, no. 2, pp. 636–646, Jun. 2021.