



Analisis Topik pada Forum Pariwisata di Indonesia Menggunakan Latent Dirichlet Allocation

Valentinus Roby Hananto^{1*}

¹Program studi Sistem Informasi, Universitas Dinamika,
Jl. Raya Kedung Baruk 98, Surabaya 60298, Telp. (031) 8721731

*Email Penulis Korespondensi: valentinus@dinamika.ac.id

Abstrak

Di era digital ini, data-data pariwisata di internet berkembang secara pesat. Jumlah data yang sangat besar ini dapat dimanfaatkan untuk meningkatkan proposisi nilai dalam pengembangan smart tourism. Indonesia sebagai salah satu tujuan utama pariwisata dunia memiliki banyak potensi pariwisata yang menarik turis dari berbagai demografis. Turis-turis ini sudah banyak menggunakan media internet seperti forum untuk membahas atau mengomentari pariwisata, khususnya di Indonesia. Penelitian ini bertujuan untuk menelaah topik pariwisata dari media forum di Indonesia. Data yang digunakan dalam penelitian ini diperoleh dari sebuah forum yaitu Kaskus dengan total 6.940 post dalam periode tahun 2008 hingga 2021. Pemodelan topik dilakukan menggunakan metode Latent Dirichlet Allocation (LDA). Topik dominan yang dihasilkan menggambarkan topik pariwisata seperti apa yang paling banyak dibahas di Indonesia. Topik populer yang ditemukan meliputi wisata alam, paket tur, hotel, kebudayaan, transportasi, sejarah, dan wisata laut. Klusterisasi topik dari penelitian ini memberikan wawasan tentang ekstraksi informasi dari review pariwisata di Indonesia.

Kata kunci—*topic modeling, Latent Dirichlet Allocation, pariwisata Indonesia, forum internet*

Abstract

In this digital era, tourism data on the internet is growing rapidly. This huge amount of data can be utilized to increase the value proposition in the development of smart tourism. Indonesia as one of the world's main tourism destinations has a lot of tourism potential that attracts tourists from various demographics. These tourists have used a lot of internet media such as forums to discuss or comment on tourism, especially in Indonesia. This study aims to examine tourism topics from media forums in Indonesia. The data used in this study were obtained from a popular web forum in Indonesia, namely Kaskus, with a total of 6,940 posts in the period 2008 to 2021. Topic modeling was carried out using the Latent Dirichlet Allocation (LDA) method. The resulting dominant topics can describe what tourism topics are most discussed in Indonesia. Popular topics found include nature tourism, tour packages, hotels, cultural heritage, transportation, history, and marine tourism. The topic clustering in this study can provide insights about information extraction from tourism reviews in Indonesia.

Keywords—*topic modeling, Latent Dirichlet Allocation, Indonesia tourism, internet forum*

1. PENDAHULUAN

Pertumbuhan aktivitas media sosial dan konten yang dihasilkan pengguna di internet merupakan peluang bagi industri pariwisata. Dengan dukungan teknologi informasi, *smart tourism* dapat diimplementasikan dengan mengubah sejumlah besar data menjadi proposisi nilai [1]. Elemen *smart tourism* menekankan pada peningkatan pariwisata yang dimediasi teknologi melalui personalisasi, pemahaman konteks, dan pemantauan secara *real-time* [2]. Implementasi *smart tourism* bergantung pada ketersediaan data dan juga adanya sistem yang mampu mengubah data menjadi proposisi nilai. Brandt dkk. mengidentifikasi tiga proposisi nilai dari ekosistem *smart tourism*: kehadiran, pendekatan lingkungan, dan pendekatan topical [3]. Sebagai contoh, proses untuk mencari tempat wisata, atraksi, dan acara populer dapat dilakukan dengan bantuan analisis media sosial.

Indonesia adalah salah satu tujuan wisata paling populer di dunia. Berdasarkan *The Travel & Tourism Competitiveness Report 2021* [4], Indonesia menduduki peringkat ke-32 dari 117 negara secara keseluruhan. Peringkat Indonesia ini naik 12 tingkat dari periode sebelumnya, sehingga Indonesia sekarang menjadi negara tujuan wisata nomor 6 di Asia setelah Jepang, Singapura, China, Korea Selatan, dan Hong Kong. Sejalan dengan popularitas wisata ini, ada banyak forum internet dan juga media sosial di Indonesia yang juga menjadi media diskusi pariwisata.

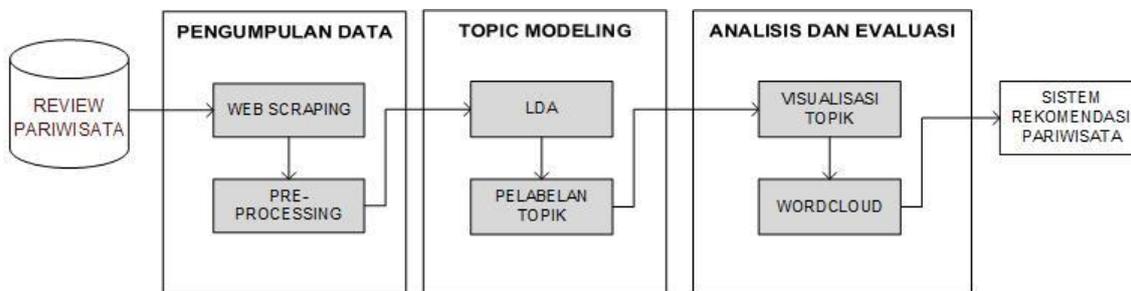
Kegunaan analitik media sosial untuk mendukung *smart tourism* telah diusulkan dalam beberapa penelitian. Figueredo dkk. menganalisis foto-foto dari media sosial yang relevan dengan preferensi tujuan pariwisata [5]. Temuan mereka berguna untuk mengembangkan sistem rekomendasi yang memudahkan turis mencari ide pariwisata. Sebuah studi tentang ekosistem pariwisata cerdas menggunakan pesan media sosial telah dilakukan di San Francisco [3]. Menggunakan 600.000 data Twitter yang diberi *geotag* di San Francisco, analisis spasial dan semantik dilakukan untuk menyajikan informasi bagi pemangku kepentingan di industri pariwisata. Pemanfaatan data Twitter untuk menganalisis pariwisata juga sudah dilakukan peneliti di Jepang, di mana data yang terkumpul diolah untuk menganalisis tren pariwisata yang sering diulas [6].

Pendekatan yang relatif baru untuk menganalisis data *review* dari internet adalah menggunakan *topic modeling*. *Topic modeling* adalah metode statistik yang diterapkan untuk menemukan tema dan tren laten dalam kumpulan dokumen teks. Data dari web forum internet, juga dapat dicirikan sebagai dokumen teks yang mengandung tema laten. Yuyan dkk. membangun *topic model* dari 24,222 *online reviews* di Jiuzhaigo, China. Model ini mewakili jenis topik apa yang paling banyak diposting oleh turis dan digunakan stakeholder terkait untuk pengembangan pariwisata [7]. Studi pemodelan topik juga dilakukan dengan menggunakan data *review* hotel di Korea untuk mengklasifikasikan pengelompokan topik yang menarik menurut pengunjung [8]. Klasifikasi ditemukan untuk membantu manajer dalam meningkatkan kepuasan pelanggan dan bagaimana peneliti dapat lebih tepat mengukur kepuasan pelanggan di industri perhotelan. Meski banyak penelitian yang mengeksplorasi konten media sosial, termasuk di Indonesia, namun belum banyak penelitian yang mengeksplorasi topik pariwisata dari forum internet. Menemukan topik-topik ini dapat memberikan wawasan berharga bagi orang-orang untuk mengilustrasikan tren dan citra pariwisata. Dalam penelitian ini, metode pemodelan topik diusulkan untuk mendapatkan topik yang paling penting dan berharga terkait pariwisata di Indonesia. Data dari forum internet diambil untuk membangun model topik, kemudian topik yang dihasilkan dapat digunakan untuk evaluasi.

Bahasan dalam artikel ini disusun ke dalam struktur berikut. Bagian 2 menjelaskan metode penelitian untuk analitik media sosial dan pemodelan topik. Hasil eksperimen dari penelitian ini dijelaskan pada Bagian 3. Bagian 4 menyajikan kesimpulan dari penelitian ini. Saran untuk pengembangan penelitian berikutnya terdapat pada Bagian 5.

2. METODE PENELITIAN

Sistem umum yang diusulkan untuk menemukan dan menganalisis topik pariwisata dari review pada forum internet dapat dilihat pada Gambar 1.



Gambar 1. Gambaran umum sistem

2.1 Tahapan Pengumpulan Data

Pada tahap awal, dataset akan dikumpulkan dari berbagai platform online melalui proses *crawling* data. Salah satu metode yang umum untuk digunakan adalah *web scraping*, yaitu proses pengumpulan data dari suatu *website*. *Web scraping* biasanya mengacu pada proses yang melibatkan otomatisasi. Beberapa situs web tidak membolehkan untuk *web scraping*, sementara yang lain tidak keberatan. Jika *web scraping* digunakan untuk tujuan pendidikan atau penelitian, kemungkinan besar tidak akan ada masalah. Tetap saja, ada baiknya dilakukan riset pendahuluan dan memastikan bahwa tidak ada ketentuan layanan apa pun yang dilanggar sebelum memulai proyek skala besar.

Web scraping berfungsi layaknya sebuah bot yang menjelajahi halaman situs web yang berbeda dan menyalin semua konten. Saat menjalankan kode, bot akan mengirimkan permintaan ke server dan datanya terkandung dalam respons yang dikirim. Proses berikutnya adalah mengurai data respons dan mengekstrak bagian yang diinginkan. Secara umum, proses *web scraping* terdiri dari tahapan berikut:

1. Memeriksa HTML situs web yang ingin dilakukan *crawling*
2. Mengakses URL situs web menggunakan kode dan mengunduh semua konten HTML di halaman
3. Format konten yang diunduh diubah menjadi format yang dapat dibaca
4. Mengekstrak informasi yang berguna dan menyimpan ke dalam format terstruktur
5. Untuk informasi yang ditampilkan di beberapa halaman situs web, langkah 2–4 dapat diulangi untuk mendapatkan informasi yang lengkap.

Sebelum pemodelan topik diterapkan untuk mengelompokkan dokumen, data yang sudah dikumpulkan perlu diproses terlebih dahulu (*pre-processing*) untuk menghapus hal-hal yang tidak perlu (misalnya tanda baca, angka, simbol, URL, dan kata hubung). *Review* yang memiliki panjang kurang dari 20 kata dihapus, karena biasanya belum memiliki makna yang kuat. *Review* yang memiliki duplikat atau konten yang sama persis juga dihapus. Kemudian, setiap *review* diiris menjadi kata-kata terpisah (*tokenization*). Setiap kata yang menjadi token kemudian diubah menjadi bentuk dasarnya (*lemmatization*). Setelah *pre-processing* teks selesai, kamus kata digunakan untuk menghasilkan matriks *document-term*, yang merupakan pemetaan setiap kata dengan frekuensinya. Matriks *document-term* ini digunakan sebagai objek input untuk membangun *topic modeling*.

2.2 Tahapan Topic Modeling

Metode *topic modeling* yang digunakan dalam penelitian ini adalah *Latent Dirichlet Allocation* (LDA). LDA adalah model probabilistik generatif untuk mengelompokkan data berupa dokumen teks [9]. LDA adalah *machine learning* yang bersifat *unsupervised*. Metode ini bertujuan untuk memodelkan topik yang berasal dari beberapa dokumen, di mana topik

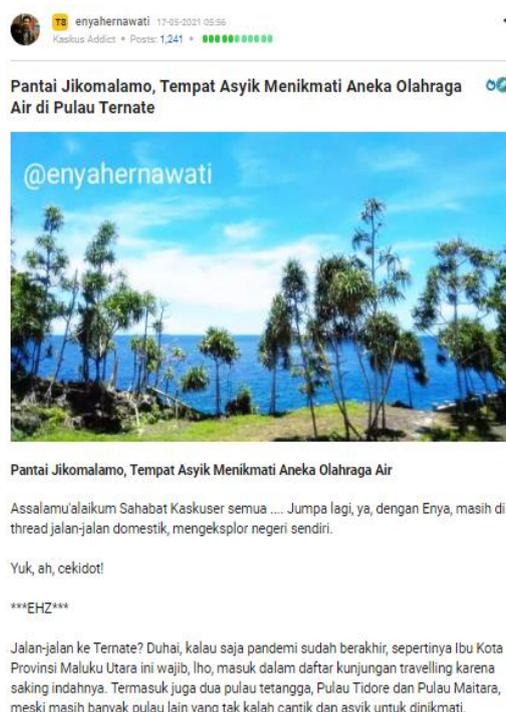
didefinisikan sebagai distribusi atas istilah atau kosa kata yang muncul dalam dokumen. Terdapat tiga langkah utama dalam prosedur generatif yang digunakan LDA. Pertama, topik dipilih secara acak berdasarkan jumlah topik yang ditentukan untuk setiap dokumen. Kedua, sebuah kata dipilih dari distribusi kata-kata yang terkait dengan topik yang dipilih. Kemudian, proses akan diulang untuk semua kata dalam dokumen. Menerapkan LDA berarti membalikkan proses generatif untuk mengekstrak topik laten dari dokumen.

2.3 Tahapan Analisis dan Evaluasi

Pada tahapan ini, hasil dari *topic modeling* dapat digunakan untuk analisis lebih lanjut. Pertama-tama dapat dilakukan visualisasi kluster dari LDA, untuk melihat persebaran kluster yang sudah terbentuk. Berikutnya, kata-kata kunci dari tiap kluster dapat dipelajari menggunakan bantuan *wordcloud*. Dengan menginterpretasikan kata-kata kunci ini, tiap topik dapat diberikan nama/label yang sesuai. Nama topik ini juga dapat dikonfirmasi dengan melihat isi dari *review* yang ada dalam tiap kluster.

3. HASIL DAN PEMBAHASAN

Data *review* untuk penelitian ini didapatkan dari proses *crawling* dari suatu forum yang berisi ulasan destinasi wisata di Indonesia. Untuk eksperimen ini, *crawling* data dilakukan menggunakan *tool* Scrapy di Python. Scrapy digunakan untuk mengekstrak data dari Kaskus (<https://www.kaskus.co.id/>) yang merupakan salah satu forum internet terbesar di Indonesia. Contoh tampilan ulasan wisata dari *website* tersebut dapat dilihat pada Gambar 2. Dari hasil *crawling* data, didapatkan dokumen sejumlah 6.940 (lihat struktur data pada Tabel 1).



Gambar 2. Contoh ulasan wisata di forum Kaskus

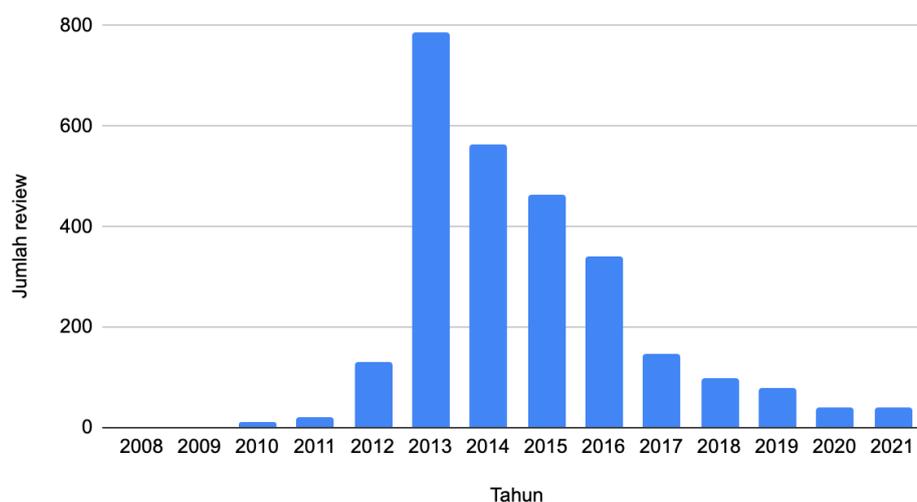
Tabel 1. Struktur data

Sumber data	www.kaskus.co.id
Rentang waktu	Tahun 2008-2020

Total dokumen	6.940
Jumlah kata	100.103
Rata-rata jumlah kalimat per review	16
Rata-rata jumlah kata per review	355

Frekuensi posting dari waktu ke waktu yang dikumpulkan selama periode ini juga dapat dianalisis. Gambar 3 menunjukkan grafik frekuensi *review* dari tahun 2008-2022. Ini menunjukkan peningkatan jumlah *review* dari tahun 2008 hingga 2013. Namun, angka tersebut menunjukkan penurunan sejak tahun 2014 hingga 2021, yang juga mengindikasikan turunnya pengguna dari forum tersebut.

Jumlah *review* per tahun

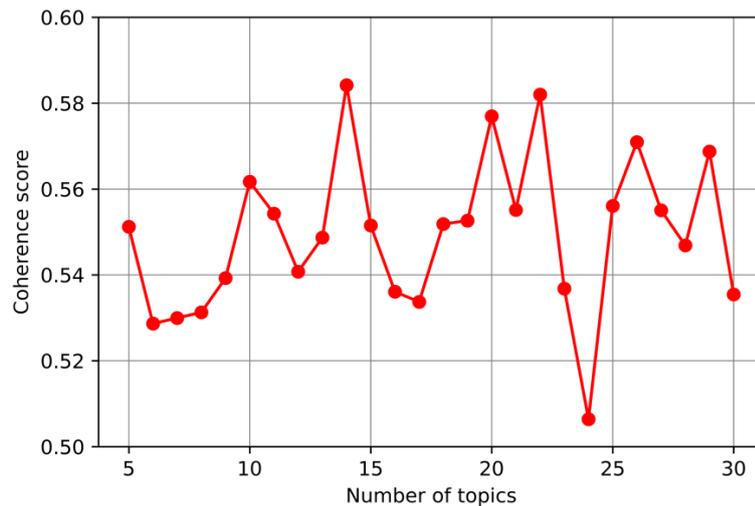


Gambar 3. Frekuensi jumlah *review* pariwisata pada forum Kaskus

Setelah melalui proses *pre-processing* teks, beberapa kata yang sering muncul seperti “ane”, “spoiler”, “gan”, “agan”, “orang”, “Indonesia”, dapat diidentifikasi dengan melihat peringkat frekuensi kata. Kata-kata umum ini cenderung dominan dan akan berdampak negatif pada hasil *topic modeling*, karena akan sering muncul di setiap kluster. Oleh karena itu, kata-kata yang sering muncul dalam 5% dokumen dan tidak memiliki arti khusus untuk topik tertentu dikeluarkan dari daftar kata. Setiap istilah unik dalam kamus diberi frekuensinya membentuk matriks *document-term*, yang selanjutnya akan digunakan sebagai objek input untuk melatih model LDA.

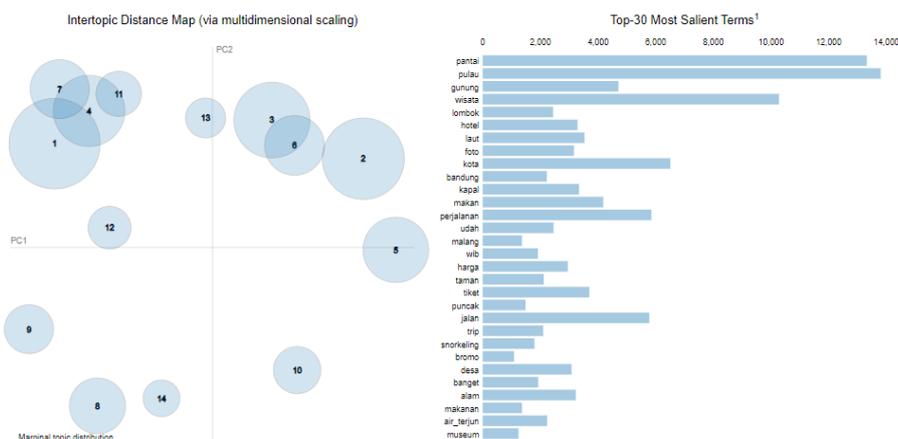
Setelah dilakukan *pre-processing*, *Bag of Words* (BoW) model dapat dihasilkan yang akan digunakan untuk masukan tahapan berikutnya, yaitu membangun *topic modeling* menggunakan metode LDA. Pemodelan LDA diterapkan menggunakan paket Tomotopy dengan Python. Dalam proses pengembangan LDA, beberapa iterasi percobaan telah dilakukan untuk menemukan jumlah kluster topik yang optimal. Jumlah topik (k) dipilih mulai dari 5 kemudian ditambah 1, misalnya $k = [5, 6, 7, \dots, 30]$, topik. Hasilnya, ada 26 *topic model* yang dibangun yang akan dikomparasi dengan menghitung skor koherensi. Ukuran koherensi C_v diimplementasikan dalam penelitian ini, didasarkan pada jendela geser yang menggunakan informasi *mutual pointwise* yang dinormalisasi (NPMI) dan kesamaan kosinus sebagai ukuran [10]. Skor koherensi C_v yang lebih tinggi menunjukkan korelasi yang lebih kuat antara semua kata dalam setiap topik, yang menunjukkan model yang lebih baik. Gambar 4 menunjukkan perubahan skor koherensi untuk data yang diberikan. Dengan membuat grafik yang membandingkan C_v dengan jumlah topik yang digunakan, maka k yang optimal akan dipilih ketika C_v mencapai puncak dan tidak ada lagi peningkatan yang signifikan. Seperti terlihat pada

gambar 3, maka nilai C_v optimal tercapai saat jumlah topik (k) adalah 14. Jumlah kluster 14 ini digunakan dalam pembuatan topic modeling menggunakan LDA. Tidak selalu mungkin untuk berhasil menentukan jumlah topik yang optimal hanya berdasarkan skor koherensi; dengan demikian, inspeksi manusia terhadap hasil pemodelan masih dibutuhkan untuk menginterpretasikan keluaran pemodelan. Setelah menentukan jumlah topik menjadi 14, evaluasi manual menemukan bahwa topiknya sudah cukup beragam dan tidak terlalu luas. 14 topik dipilih selanjutnya digunakan untuk analisis lebih lanjut. Tema yang diekstraksi digunakan untuk membuat kelas untuk model pengetahuan.



Gambar 4. Grafik nilai koherensi (C_v) berdasarkan jumlah topik

Visualisasi LDA dibuat menggunakan bagan interaktif LDAvis dengan Python (Gambar 5). Sistem ini dikembangkan oleh Sievart dan Shirley [11] untuk lebih memahami model LDA. Di sisi kiri, topik diplot sebagai gelembung dalam bidang dua dimensi yang pusatnya ditentukan dengan menghitung jarak antar topik. Semakin besar gelembungnya, semakin umum topik tersebut. Model topik yang baik akan memiliki gelembung yang tidak tumpang tindih yang tersebar di seluruh bagan. Memilih gelembung di panel kiri secara interaktif mengungkapkan istilah yang paling relevan untuk kelompok topik tertentu di panel sebelah kanan.



Gambar 5. Visualisasi LDA

Berdasarkan *output* dari LDA model yang didapat dari eksperimen, terdapat 14 kluster topik yang dihasilkan. Setiap dokumen yang sudah dikumpulkan kemudian dikelompokkan ke

dalam kluster berdasarkan bobot keanggotaan terhadap masing-masing kluster, dan kemudian dipilih bobot topik yang tertinggi. Untuk membantu pemahaman terhadap hasil LDA model, dilakukan analisis terhadap kata-kata dengan bobot tertinggi dari masing-masing kluster topik. Hasil pembobotan kata ini dapat divisualisasikan ke dalam bentuk *wordcloud*, seperti pada gambar 6.

Word Cloud per Topic



Gambar 6. Wordcloud per topik

Berdasarkan jumlah dokumen dari tiap-tiap kluster topik, topik 2 adalah yang paling dominan dengan jumlah 27% dari total dokumen. Namun topik ini dapat diabaikan, karena banyak berisi kata-kata yang terlalu umum dan tidak spesifik untuk makna tertentu seperti: sudah, banget, jalan, langsung, spoiler. Topik dominan berikutnya (topik 4) memiliki kata kunci dalam urutan bobot terkait: wisata, air, alam, kawasan, air_terjun. Dengan menemukan kata kunci topik, analisis dokumen yang sebagian besar disumbangkan oleh topik tertentu akan membantu menafsirkan topik tersebut. Dari sini dapat diinterpretasikan topik terkait wisata alam yang banyak membahas pariwisata alam di Indonesia. Dengan membaca dokumen dalam suatu

topik, topik tersebut akan dikonfirmasi. Contoh kata kunci dan *review* dari 5 topik tertinggi dapat dilihat pada Tabel 2.

Tabel 2. Kata kunci dan contoh review pada 5 topik terpopuler

No.	Topik	10 kata kunci	Persentase <i>review</i>	Contoh <i>review</i>
1	Topik 4 (wisata alam)	Wisata, air, alam, Kawasan, air_terjun, menikmati, keindahan, memiliki, kota, salah	14,4%	“Minahasa Tenggara sudah dikenal dan populer di antara orang-orang Eropa karena hasil bumi dan keindahan alamnya. Kabupaten yang masuk di Sulawesi Utara ini sering kali disebut sebagai salah satu tempat pariwisata bahari di Indonesia. karena memiliki alam yang indah dan dikenal juga sebagai salah satu Spot untuk menyelam di Sulawesi Utara...”
2	Topik 5 (paket tur)	Makan, trip, perjalanan, malam, biaya, tiket, peserta, tanggal, mei, sewa	11,9%	“Halo Jalan2ers, Masih belum tau Liburan Natal mau kemana? gabung yuk sama kita. Saya dan beberapa orang teman rencana berangkat tanggal 25 - 28 Des'14. Kita masih mencari 2-3 teman jalan lagi untuk sharing cost kebeberapa tempat berikut ini : Kawah Ijen (Banyuwangi)...”
3	Topik 1 (hotel/ akomodasi)	Hotel, harga, tiket, wisata, paket, murah, liburan, fasilitas, kamar, tour	10%	“Perkiraan baget Menginap di penginapan gili trawangan di mulai dari harga 150.000 sampai dengan 6.500.000 ada di pulau ini. apabila anda pergi berlibur ke gili trawangan jangan lupa persiapkan baget sesuai dengan saku anda. untuk dekat pantai harga rata-rata dari 650.000 sampai dengan 6.5 jt...”
4	Topik 7 (budaya)	Masyarakat, budaya, memiliki, dunia, daerah, desa, festival, salah, acara, negara	6,4%	“Sebagai daerah wisata yang baru muncul di Kabupaten Pangandaran, menawarkan wisata alam dan budaya berupa rafting dan berwisata budaya dengan rutin di gelarnya budaya lokal yang telah menjadi salah satu ikon kebudayaan Kabupaten Pangandaran yaitu kesenian Badud. Cara yang di tempuh Margacinta cukup unik untuk mengenalkan potensi wisata yang ada, dengan mengenalkan kesenian yang di gelar rutin...”

5	Topik 0 (transportasi)	Perjalanan, jalan, masuk, jalur, arah, terminal, bus, malam, langsung, kereta	5,7%	“Jakarta kian berbenah dalam berbagai hal, termasuk dalam meningkatkan minat wisatawan berkunjung ke ibukota negara Indonesia tersebut. Direncanakan, dalam waktu dekat para wisatawan, baik domestik dan mancanegara dapat menikmati wisata Jakarta menggunakan Bus Wisata dan akan menjadi ikon baru wisata Kota Jakarta...”
---	---------------------------	---	------	--

4. KESIMPULAN

Dalam penelitian ini, pemodelan topik menggunakan LDA dilakukan pada *review* pariwisata yang diposting di forum internet di Indonesia. Dari segi frekuensi *review* dalam periode dari tahun 2014 hingga tahun 2021 mengalami penurunan, sehingga dapat menjadi evaluasi pengelola forum untuk pengembangan sistem. Dari segi analisis kontekstual, model topik yang dibangun menghasilkan total 14 kluster topik, dengan topik dominan tentang wisata alam, paket tur, hotel/akomodasi, budaya, dan transportasi. Dengan memanfaatkan klusterisasi topik ini, dapat dilakukan analisis dari *review* yang ada serta rekomendasi yang diperlukan untuk pengembangan pariwisata di Indonesia.

5. SARAN

Dengan menghitung skor koherensi, jumlah topik yang dihasilkan dalam penelitian ini dipilih menjadi 14. Namun, menentukan jumlah topik yang paling sesuai adalah bagian yang paling menantang dalam pemodelan topik. Jika nomor topik diatur lebih rendah, maka beberapa topik terlalu luas dan perlu dipisahkan. Di sisi lain, jika nomor topik diatur lebih tinggi, maka ditemukan beberapa topik serupa yang dapat digabungkan. Metrik seperti skor koherensi dapat digunakan untuk membantu menentukan jumlah topik, tetapi tidak ada jumlah topik yang sempurna yang dapat menghilangkannya sama sekali. Oleh karena itu, mengeksplorasi metode yang lebih baru untuk memilih sejumlah topik yang sesuai atau mengeksplorasi pemodelan topik yang bersifat *semi-supervised* dapat dipertimbangkan untuk penelitian di masa mendatang.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tim Redaksi Jurnal Teknik Politeknik Negeri Sriwijaya yang telah memberi kesempatan, sehingga artikel ilmiah ini dapat diterbitkan.

DAFTAR PUSTAKA

- [1] Gretzel, U., Sigala, M., Xiang, Z., and Koo, C. (2015). Smart tourism: foundations and developments. *Electronic Markets*, 25(3), 179–188.
- [2] Buhalis, D., and Amaranggana, A. (2015). Smart Tourism Destinations Enhancing Tourism Experience Through Personalisation of Services, In: *Information and Communication Technologies in Tourism 2015*, Tussyadiah, I., and Inversini, A. (Eds.), 377-389. Springer: Cham.

- [3] Brandt, T., Bendler, J., and Neumann, D. (2017). Social media analytics and value creation in urban smart tourism ecosystems. *Information and Management*, 54(6), 703–713.
- [4] World Economic Forum. (2021). *Travel & Tourism Development Index 2021: Rebuilding for a Sustainable and Resilient Future*. <https://www.weforum.org/reports/travel-and-tourism-development-index-2021/> (diakses pada 25 Juni, 2023)
- [5] M. Figueredo, N. Cacho, A. Thome, A. Cacho, F. Lopes and M. Araujo, "Using social media photos to identify tourism preferences in smart tourism destination," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 2017, pp. 4068-4073, doi: 10.1109/BigData.2017.8258423.
- [6] Hananto, V. R., Serdült, U., Kryssanov, V. V. (2020, March). Discovering tourism topics from social media: A case study of Japan. In *CEUR Workshop Proceedings* (pp. 83-89). CEUR-WS.
- [7] Luo, Y., Tong, T., Zhang, X., Yang, Z. and Li, L. (2023), "Exploring destination image through online reviews: an augmented mining model using latent Dirichlet allocation combined with probabilistic hesitant fuzzy algorithm", *Kybernetes*, Vol. 52 No. 3, pp. 874-897. <https://doi.org/10.1108/K-07-2021-0584>.
- [8] Sutherland I, Sim Y, Lee SK, Byun J, Kiatkawsin K. Topic Modeling of Online Accommodation Reviews via Latent Dirichlet Allocation. *Sustainability*. 2020; 12(5):1821. <https://doi.org/10.3390/su12051821>
- [9] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- [10] Röder, M., Both, A., and Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures, In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining - WSDM '15*, 399–408. ACM: Shanghai.
- [11] Sievert, C., and Shirley, K. E. (2014). LDAvis: A method for visualizing and interpreting topics, In: *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70. Association for Computational Linguistics: Baltimore.