

ANALISIS DATA REVIEW HOTEL DI GOOGLE MAPS MELALUI TEXT MINING (STUDI KASUS: KABUPATEN BANDUNG)

Jorgha Akram Aryandi¹, Yahya Aliman Aidil Basith², Munawir³, dan Devi Aprianti Rimadhani Agustini⁴
^{1,2,3,4}Program Studi Teknik Komputer, Kampus Daerah Cibiru, Universitas Pendidikan Indonesia
Email: jorghaaryandi@upi.edu¹, yahyaaliman@upi.edu², munawir@upi.edu³, deviaprianti@upi.edu⁴

Abstrak

Penelitian ini bertujuan untuk menganalisis data review hotel di Google Maps melalui text mining di Kabupaten Bandung. Metode text mining digunakan untuk menggali informasi yang terkandung dalam teks ulasan pengguna hotel. Data review hotel diambil dari platform Google Maps dan diolah menggunakan metode Latent Dirichlet Allocation (LDA). Hasil analisis menunjukkan adanya bobot setiap kata dalam ulasan pengguna terkait hotel di Kabupaten Bandung. Temuan ini memberikan wawasan tentang kebersihan, fasilitas, pelayanan, dan pengalaman pengguna yang dapat digunakan sebagai bahan evaluasi dan pengembangan hotel di daerah tersebut. Hasil penelitian didapatkan faktor daya tarik yang paling dominan ialah Kualitas dan Pelayanan (37.60%), kemudian Fasilitas dan Lingkungan (34.54%), dilanjut dengan Aktivitas dan Hiburan (27.85%). Setelah analisis LDA, data tersebut diperkuat dengan Analysis of Variance (ANOVA) One Way. Hasil yang diperoleh dari ANOVA adalah p-value bernilai 0.48578981747778355. Dalam kasus ini, p-value (0.48578981747778355) lebih besar dari tingkat signifikansi yang umumnya digunakan (0.05), yang menunjukkan bahwa tidak ada perbedaan signifikan antara kelompok fasilitas dan lingkungan, aktivitas dan hiburan, serta kualitas dan pelayanan pada hotel-hotel yang dianalisis.

Kata Kunci: Google Maps, Hotel, Kabupaten Bandung, LDA, Review

Abstract

This study aims to analyze hotel review data on Google Maps through text mining in Bandung Regency. The text mining method is used to dig up the information contained in the text of hotel user reviews. Hotel review data is taken from the Google Maps platform and processed using the Latent Dirichlet Allocation (LDA) method. The results of the analysis show that there is weight for each word in user reviews regarding hotels in Bandung Regency. These findings provide insights about cleanliness, facilities, service, and user experience that can be used as material for evaluating and developing hotels in the area. The results showed that the most dominant attractiveness factors were Quality and Service (37.60%), then Facilities and Environment (34.54%), followed by Activities and Entertainment (27.85%). After LDA analysis, the data is strengthened by One Way Analysis of Variance (ANOVA). The results obtained from ANOVA are the p-value of 0.48578981747778355. In this case, the p-value (0.48578981747778355) is greater than the commonly used significance level (0.05), indicating that there are no significant differences between the groups of facilities and environment, activities and entertainment, and quality and service at the hotels analyzed.

KeyWords: Google Maps, Hotel, Kabupaten Bandung, LDA, Review

I. PENDAHULUAN

Pentingnya ulasan *online* sudah akhir-akhir ini. Dengan keberadaannya di mana-mana dan keluasannya, sekarang dapat melihat tingkat kepuasan pelanggan di satu sisi tanpa banyak kesulitan, dan bahkan terus menerapkan umpan balik untuk meningkatkan pengalaman pelanggan secara keseluruhan. Karena pengaruh ulasan terus berkembang, pelanggan sangat dipengaruhi oleh ulasan dalam hal pengambilan keputusan [1]. Misalnya, ulasan *online* adalah faktor penentu utama bagi pelanggan saat memutuskan apakah akan membeli produk tertentu atau mengunjungi hotel tertentu atau tidak. Kami menemukan ribuan ulasan yang berisi berbagai faktor penilaian yang memengaruhi preferensi konsumen saat ini. Namun, informasi tersebut umumnya diberikan oleh komentar sederhana atau peringkat bintang belaka. Berdasarkan ulasan sederhana, tidak mudah untuk mendapatkan gambaran menyeluruh tentang apa yang diinginkan konsumen secara keseluruhan atau faktor apa yang dirujuk konsumen saat membuat keputusan kunjungan kembali.

Semakin berkembangnya bisnis hotel membuat persaingan pada industri hotel semakin ketat, setiap hotel berusaha untuk meningkatkan tingkat hunian kamarnya [2]. Dalam hal ini, para pengusaha harus memiliki keunggulan kompetitif agar mampu bersaing dan bertahan dengan para pesaing hotel lainnya. Para pengusaha hotel dapat melakukan kegiatan promosi untuk dapat menarik pelanggan baru, seperti mengiklankan jasa hotel melalui *website* ataupun *e-commerce* [2]. Salah satu faktor yang dapat mempengaruhi pelanggan dalam menentukan hotel ialah *review* atau ulasan hotel [3].

Ulasan *online* wisatawan telah menjadi sumber informasi penting untuk sektor perhotelan. Popularitas yang lebih luas menghasilkan data ulasan yang di *posting* di *Google Maps* ini menciptakan peluang bagi sektor perhotelan untuk menganalisis data untuk menentukan faktor-faktor yang memengaruhi pengalaman pelanggan dan pengambilan keputusan [3]. Berbagai platform dijadikan sebagai alat promosi hotel untuk menarik pengunjung. Selain memberikan informasi terkait hotel, platform ini bertindak sebagai sumber penting data publik dalam bentuk ulasan *online* pengunjung, yang menggambarkan pengalaman berpariwisata wisatawan [4].

Industri perhotelan semakin kompetitif karena layanan serupa disediakan oleh semua pelaku bisnis perhotelan, sehingga pelaku bisnis perhotelan perlu fokus untuk memuaskan pelanggan mereka dengan memberikan pelayanan yang lebih baik daripada pesaingnya [5]. Selain itu, kepuasan pelanggan adalah bagian dari proses mereka mengevaluasi layanan yang diberikan. Pada saat yang sama, memuaskan pengunjung untuk mendapatkan loyalitas. Oleh karena itu, penting dalam industri perhotelan untuk memberikan layanan berkualitas tinggi sehingga kepuasan pelanggan tercapai [5].

Data ulasan *online* khususnya telah dianggap sebagai faktor penting untuk mempengaruhi keputusan konsumen dan dinilai sebagai aset berdasarkan informasi yang berharga [6]. Dalam penelitian ini, kami menggunakan metode *review online* dari *google maps* untuk mengumpulkan semua informasi dari *reviewer* tentang hotel dan menganalisis daya tarik hotel di Kabupaten Bandung. Pendekatan yang kami lakukan adalah mengevaluasi persepsi pengunjung dengan mengidentifikasi kalimat dari opini atau ulasan mengenai kepuasan pelanggan berdasarkan fasilitas, aktivitas pengunjung, kualitas, maupun pelayanan hotel.

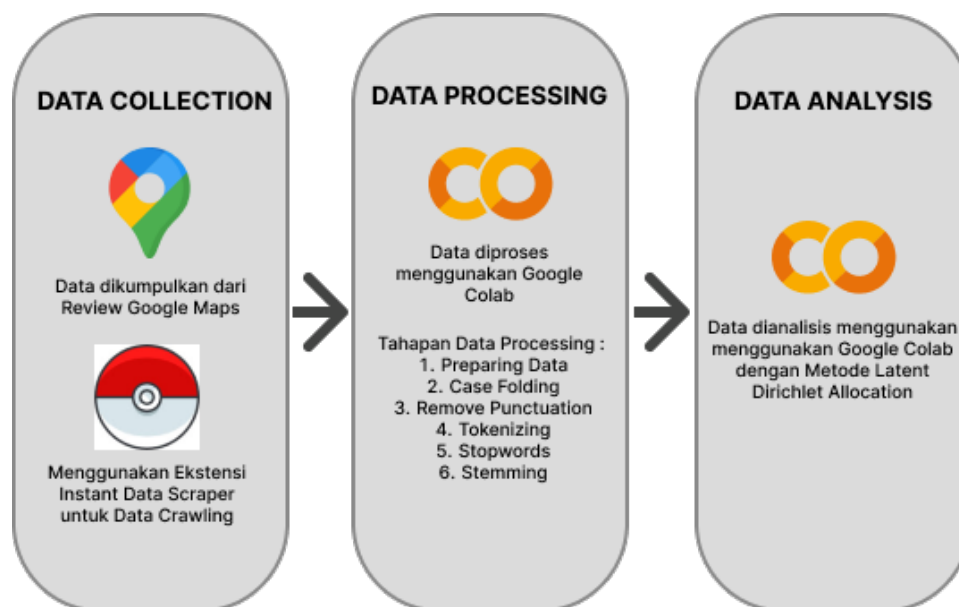
Adapun penelitian terdahulu berjudul Deteksi Topik Tentang Tokoh Publik Politik Menggunakan *Latent Dirichlet Allocation (LDA)*. Pada penelitian tersebut menggunakan metode TF-IDF untuk mengetahui nilai rata-rata setiap *tweet* dan mengetahui *tweet* mana saja yang mengandung kata-kata terbanyak dalam *output* LDA. Dengan menggunakan metode TF-IDF, user akan lebih mudah mengetahui makna dari *output* yang dihasilkan pada tahapan LDA [7]. Penelitian selanjutnya dilakukan oleh Rizki Ramadandi, dkk. berjudul Pemodelan Topik Menggunakan Metode *Latent Dirichlet Allocation* dan *Gibbs Sampling* kemudian dilakukan dengan menggunakan metode koherensi topik pengukuran skor UCI dengan hasil penelitian menyebutkan diperoleh lima buah topik optimal pada kedua konfigurasi pengujian [8]. Penelitian lainnya berjudul Analisis Pemodelan Topik Ulasan Aplikasi BNI, BCA dan BRI menggunakan *Latent Dirichlet Allocation* menghasilkan adanya perbedaan dan kesamaan dalam kualitas layanan antara ketiga aplikasi tersebut [9].

Berdasarkan fungsi-fungsi tersebut, penelitian ini akan mengevaluasi faktor dominan daya tarik hotel berdasarkan *review online* pengunjung hotel. Kategori evaluasi diambil berdasarkan dari referensi [6]. Lebih jelasnya dapat dilihat pada Tabel I.

Tabel I: Kategori topik dan variabel evaluasi daya tarik hotel

No	Kategori	Variabel
1	Fasilitas dan Lingkungan	Fasilitas dan Lingkungan Hotel
2	Aktivitas dan Hiburan	Aktivitas dan Hiburan Hotel
3	Kualitas dan Pelayanan	Kualitas dan Pelayanan Hotel

II. METODE PENELITIAN



Gambar 1: Tahapan Penelitian

- 1) *Data Collection:* Pada tahap ini, kami melakukan pengumpulan data dari *Review Google Maps* dan di *crawling* menggunakan ekstensi *Instant Data Scraper*.
- 2) *Data Processing:* Pada tahap ini, dilakukan pemrosesan data menggunakan *Google Colab* dengan bahasa *Python* yang terdiri dari beberapa tahapan seperti:
 - *Preparing Data:* Merujuk pada langkah-langkah yang dilakukan untuk mempersiapkan data mentah sebelum diproses lebih lanjut.

- *Case Folding*: Merujuk pada proses mengubah semua huruf dalam teks menjadi huruf kecil atau huruf besar. Hal ini dilakukan untuk menghindari perbedaan huruf besar dan huruf kecil yang mempengaruhi pemrosesan dan analisis teks.
 - *Remove Punctuation*: Merupakan langkah dalam pemrosesan teks yang melibatkan penghapusan semua tanda baca dari teks. Tanda baca seperti koma, titik, tanda tanya, tanda seru, dan lain-lain., dihilangkan untuk membersihkan teks sebelum langkah-langkah selanjutnya.
 - *Tokenizing*: Merupakan proses memecah teks menjadi unit-unit yang lebih kecil yang disebut "token". Token bisa berupa kata-kata, frasa, atau simbol tertentu yang memiliki makna dalam konteks tertentu. Tokenisasi umumnya melibatkan pemisahan kata-kata dari kalimat dengan menggunakan spasi sebagai pemisah utama atau menggunakan tanda baca sebagai pemisah.
 - *Stopwords*: Merujuk pada kata-kata yang umumnya dianggap tidak memiliki makna penting dalam pemrosesan teks karena kemunculannya yang sangat sering. Contoh *stopwords* dalam bahasa Indonesia adalah 'yg', 'bgt', 'wah', dan lain-lain. Biasanya, *stopwords* dihapus dari teks untuk mengurangi dimensi dan meningkatkan efisiensi analisis teks.
 - *Stemming*: Merupakan proses pengubahan kata-kata menjadi bentuk dasarnya atau "akar kata" dengan menghilangkan awalan, akhiran, atau infleksi lainnya. Tujuan dari *stemming* adalah untuk mengurangi kata-kata yang memiliki akar kata yang sama agar dapat dihitung sebagai satu entitas dalam analisis teks. Misalnya, kata "berjalan", "berjalanlah", dan "berjalanlah" akan *distem* menjadi "jalan" agar dapat dianggap sebagai satu kata yang sama.
- 3) *Data Analysis*: Pada tahap ini, kami melakukan analisis data menggunakan metode *Latent Dirichlet Allocation (LDA)* dan kemudian dianalisis dengan *ANOVA One Way*.

A. Studi Kasus Hotel

Hotel yang dipilih untuk dilakukan analisis adalah beberapa hotel terbaik di Kabupaten Bandung terkhusus di wilayah Soreang dan Ciwidey. Berikut Tabel II disajikan detail hotel.

Tabel II: Hotel sebagai Studi Kasus

No	Nama Hotel	Lokasi
1	Grand Sunshine Resort & Convention Soreang Bandung	Jl. Raya Soreang No.06, Pamekaran, Kec. Soreang, Kabupaten Bandung, Jawa Barat 40912
2	Forest Hills Hotel Ciwidey	Jl. Propinsi, Sukajadi, Kec. Soreang, Kabupaten Bandung, Jawa Barat 40911
3	Sutan Raja Hotel and Convention Centre	Jalan Raya Soreang KM.17 No.10 Cingcin, Pamekaran, Kec. Soreang, Kabupaten Bandung,
4	Hotel Soreang	Jl. Terusan Al-Fathu, A Kp. Sangkan Betah No.44, RT.003/RW.014, Parungserab, Kec. Soreang, Kabupaten Bandung, Jawa Barat 40319

B. Data Collection

Data yang di ekstraksi dari *Google Maps* digunakan untuk menemukan lokasi hotel. Kami mengumpulkan data dari ulasan *online Google Maps* berdasarkan lokasi hotel. Kami menggunakan *web crawler* untuk mengekstrak semua data dari ulasan Pengguna *Google Maps*. *Web crawler* adalah sebuah program yang dapat mencari informasi yang sesuai dengan kata kunci dalam sebuah *website* [10]. Semua data yang diekstraksi disimpan dalam *file excel* dengan data pengguna, evaluasi dalam bentuk skor bintang dan teks opini dari *reviewer*. Tabel III menunjukkan data yang dikumpulkan dari ulasan *online*.

Tabel III: Distribusi Total Review Online (Per 21 Juni 2023)

No	Nama Hotel	Total Review	Review yang dianalisis
1	Grand Sunshine Resort & Convention Soreang Bandung	5985	880
2	Forest Hills Hotel Ciwidey	163	163
3	Sutan Raja Hotel and Convention Centre	5322	730
4	Hotel Soreang	503	502

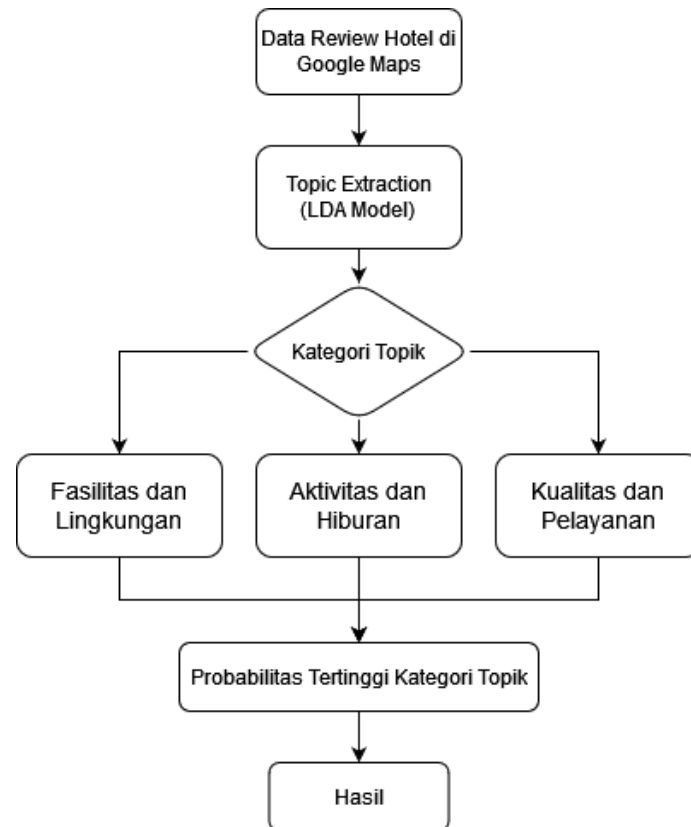
C. Topic Extraction

Kami mengumpulkan 100 ulasan dari situs *web google* untuk 4 hotel terbaik di Kabupaten Bandung. Kemudian, dilakukan ekstraksi data melalui *website google maps* ke lokasi yang ditentukan untuk masing-masing hotel. Tinjauan metodologi yang dilakukan dan alur sistem yang diusulkan ditunjukkan pada Gambar 2. Studi ini menggunakan perintah operasi otomatis untuk mengumpulkan data dari situs *web Google Maps*.

Model *Latent Dirichlet Allocation (LDA)* diperkenalkan untuk mengurangi beban kerja dan menangani data yang tidak seimbang [11] untuk menyelidiki topik laten dari korpus [12]. Secara umum, model LDA berasumsi bahwa dokumen terdiri dari distribusi topik dan topik terdiri dari distribusi kata yang koheren secara semantik. Ini adalah algoritma tanpa pengawasan yang memodelkan setiap dokumen sebagai campuran topik untuk menghasilkan ringkasan topik secara otomatis dalam hal distribusi probabilitas diskrit atas kata-kata untuk setiap topik, dan selanjutnya menyimpulkan distribusi diskrit per-dokumen

atas topik [12]. LDA banyak digunakan untuk mengekstrak informasi topik laten dari sejumlah besar dokumen. Dengan demikian, penelitian ini bermaksud untuk memanfaatkan model LDA untuk menganalisis topik pada *review* hotel di Kabupaten Bandung.

Model LDA menentukan alokasi kata untuk topik yang berbeda, kemudian dilakukan identifikasi probabilitas kata tertinggi di setiap topik. Kami menggunakan jumlah kata sebanyak 10 kata teratas setiap topik dan menyajikan bobot yang sesuai dari setiap kata. Bobot setiap kata menunjukkan *relative importance* di dalam topik.



Gambar 2: Prosedur Identifikasi Daya Tarik Hotel

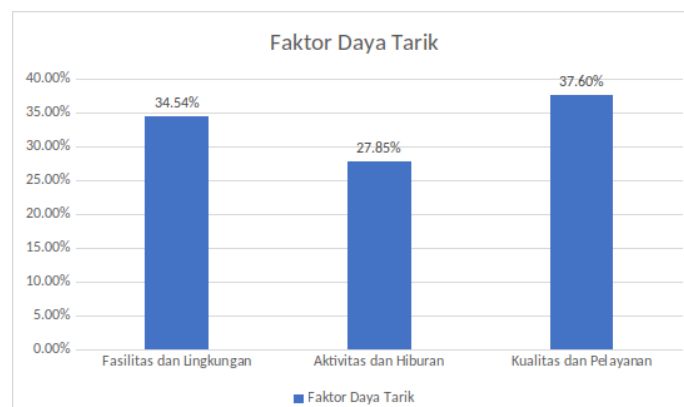
III. HASIL DAN PEMBAHASAN

Dari hasil penelitian ini terdapat 3 kategori topik mengenai prioritas kata kunci untuk mengidentifikasi daya tarik pengunjung hotel berdasarkan *review online*. Kemudian kami mengklasifikasikan kategori topik menjadi tiga kategori yaitu: Fasilitas dan Lingkungan, Aktivitas dan Hiburan, Kualitas dan Pelayanan. Kami menamai tiga topik yang di-ekstraksi. Istilah kata kunci teks ulasan 4 hotel akan dibahas pada bagian berikut ini. Adapun topik yang sesuai dengan untuk *review* Grand Sunshine Hotel, kita dapat membedakan setiap topik dengan meringkas 10 kata teratas yang sesuai. Bobot total dari setiap topik mewakili tingkat kepentingan relatif dari setiap topik.

Tabel IV: Kategori Topik Grand Sunshine Hotel

Topik 1. Fasilitas dan Lingkungan 0.1820			
kamar	0.0232	bersih	0.0172
kolam	0.0229	nyaman	0.0169
anak	0.0212	fasilitas	0.0162
ramah	0.0173	luas	0.0152
renang	0.0172	enak	0.0149
Topik 2. Aktivitas dan Hiburan 0.0738			
anak	0.014	zoo	0.0059
makan	0.0103	fasilitas	0.0059
bagus	0.0077	kesini	0.0053
acara	0.0076	nyaman	0.0053
kamar	0.0065	cocok	0.0053
Topik 3. Kualitas dan Pelayanan 0.0671			
kamar	0.0154	ruang	0.0048
anak	0.0121	bintang	0.0044
mewah	0.0073	makan	0.0043
layan	0.0055	handuk	0.004
bagus	0.0054	bayar	0.0039

Bobot (nilai probabilitas) dari setiap kata yang diperluas menunjukkan proporsi frekuensi kata yang diperluas dalam setiap topik dari jumlah total kata. Bobot relatif setiap topik dihitung sebagai jumlah bobot relatif dari 10 kata.

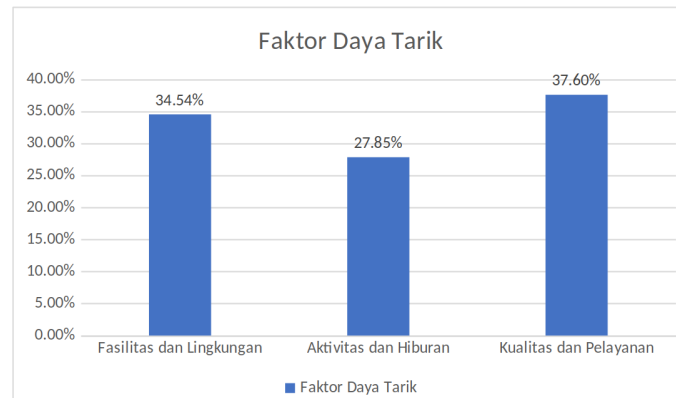


Gambar 3: Bobot persentase *Review Online* Grand Sunshine Resort

Hasil penelitian menunjukkan bahwa “Fasilitas dan Lingkungan” merupakan faktor terpenting daya tarik hotel di Grand Sunshine Resort & Convention Soreang Bandung. Banyak pengunjung berbagi pengalaman mereka tentang fasilitas dan pelayanan di hotel. Meringkas *review online* faktor daya tarik pengunjung hotel, karena topik telah dikategorikan ke dalam variabel daya tarik, proses penentuan prioritas topik difokuskan untuk mengevaluasi kategori topik dari semua statistik probabilitas semua taman tematik. Untuk membandingkan setiap hotel, ringkasan prioritas topik diilustrasikan sebagai berikut Tabel V.

Tabel V: Ringkasan dari Kategori Topik *Review* Hotel

Nama Hotel	Kategori Topik		
	Fasilitas danLingkungan	Aktivitas danHiburan	Kualitas dan Pelayanan
Grand Sunshine Resort & Convention Soreang Bandung	0.182	0.0738	0.0671
Forest Hills Hotel Ciwidey	0.1152	0.1097	0.1523
Sutan Raja Hotel and Convention Centre	0.1338	0.154	0.2154
Hotel Soreang	0.1644	0.1427	0.2134
Rata-Rata	0.1489	0.12	0.162
Persentase	34.54%	27.85%	37.60%



Gambar 4: Distribusi Faktor Daya Tarik Hotel di Kabupaten Bandung

Gambar 4 menunjukkan persentase ringkasan kategori topik hotel dari teks ulasan pengunjung dan rata-rata yang dijelaskan pada Tabel V sebelumnya. Faktor daya tarik hotel yang dominan adalah faktor Kualitas dan Pelayanan yang rata-rata persentasenya adalah 37.60%. Dilanjut dengan faktor Fasilitas dan Lingkungan sebesar 34.54% dan faktor Aktivitas dan Hiburan dengan persentase 27.85%. Untuk membandingkan hipotesis kami bahwa daya tarik hotel adalah tiga kategori, "Fasilitas dan Lingkungan", "Aktivitas dan Hiburan", dan "Kualitas dan Pelayanan" dengan ekstraksi topik dari ulasan *online*, kami menggunakan *Analysis of Variance (ANOVA)*. *ANOVA One Way* digunakan untuk membandingkan rata-rata atau mean dari satu variabel dependen di antara beberapa kelompok yang berbeda secara statistik. Tujuan dari *ANOVA One Way* adalah untuk menentukan apakah terdapat perbedaan yang signifikan antara kelompok-kelompok tersebut. Analisis menggunakan ANOVA dilakukan di *Google Colab* dengan beberapa *syntax* yang diperintahkan. Hasil yang diperoleh dari analisis ialah seperti berikut:

```

Hasil Analisis ANOVA:
Statistic: 0.7831220699675596
p-value: 0.48578981747778355
    
```

Berdasarkan hasil analisis ANOVA yang diperoleh, statistik uji adalah 0.7831220699675596, dan nilai p-value adalah 0.48578981747778355. Pada analisis ANOVA, nilai p-value adalah ukuran signifikansi statistik. Jika p-value lebih kecil dari tingkat signifikansi yang ditentukan (biasanya 0.05), maka kita dapat menolak hipotesis nol dan menyimpulkan bahwa terdapat perbedaan yang signifikan antara kelompok-kelompok yang dibandingkan. Namun, jika p-value lebih besar dari tingkat signifikansi, kita gagal menolak hipotesis nol dan menyimpulkan bahwa tidak ada perbedaan signifikan antara kelompok-kelompok tersebut.

Dalam kasus ini, p-value (0.48578981747778355) lebih besar dari tingkat signifikansi yang umumnya digunakan (0.05), yang menunjukkan bahwa tidak ada perbedaan signifikan antara kelompok fasilitas dan lingkungan, aktivitas dan hiburan, serta kualitas dan pelayanan pada hotel-hotel yang dianalisis.

Namun, penting untuk dicatat bahwa hasil ini hanya berdasarkan data yang diberikan dan mungkin terdapat faktor-faktor lain yang tidak diperhitungkan dalam analisis ini. Selain itu, interpretasi yang lebih rinci dari hasil ini juga dapat dilakukan dengan mempertimbangkan konteks dan tujuan penelitian yang lebih spesifik.

IV. SIMPULAN

Penelitian ini menginformasikan bahwa ulasan *online* dapat digunakan untuk memahami daya tarik hotel di Kabupaten Bandung. Ulasan *online* memberikan informasi yang berharga dalam mengevaluasi keindahan karena data yang melimpah tersedia dan mencerminkan opini publik melalui teks. Berdasarkan ekstraksi topik, penelitian ini bertujuan untuk mengembangkan atau meningkatkan industri perhotelan. Analisis varians digunakan untuk mengidentifikasi faktor-faktor yang dianggap mempengaruhi daya tarik hotel, yaitu "Fasilitas dan Lingkungan", "Aktivitas dan Hiburan", dan "Kualitas dan Pelayanan". Hasil analisis dengan Model LDA dan ANOVA menunjukkan bahwa hipotesis nol diterima, mengindikasikan bahwa ketiga faktor tidak ada perbedaan yang signifikan. Dengan demikian, dapat disimpulkan bahwa daya tarik hotel terdiri dari tiga faktor, yaitu "Fasilitas dan Lingkungan", "Aktivitas dan Hiburan", dan "Kualitas dan Pelayanan".

Dalam 4 hotel terbaik di Kabupaten Bandung yang diteliti, faktor fasilitas dominan ditemukan di Grand Sunshine Resort & Convention Soreang Bandung (30.56%), Faktor Aktivitas dan Hiburan ditemukan di Sutan Raja Hotel and Convention Centre (32.06%), dan Faktor Kualitas dan Pelayanan ditemukan di Sutan Raja Hotel and Convention Centre (33.22%). Namun, persentase bobot untuk setiap faktor tidak memiliki perbedaan yang signifikan di antara ke-4 hotel tersebut.

Hasil ekstraksi topik menunjukkan bahwa daya tarik hotel dapat dikategorikan menjadi tiga faktor, yaitu Fasilitas dan Lingkungan, Aktivitas dan Hiburan, dan Kualitas maupun Pelayanan. Faktor Kualitas dan Pelayanan adalah faktor dominan dalam daya tarik hotel. Secara khusus, analisis ini menunjukkan bahwa variasi aktivitas yang ditawarkan di hotel dipengaruhi

oleh ketersediaan fasilitas. Fasilitas dan Lingkungan juga menjadi faktor yang meningkatkan daya tarik masyarakat untuk menginap atau sekedar mengunjungi hotel. Hasil analisis ini diharapkan dapat memberikan informasi yang berharga dalam pengembangan hotel.

PUSTAKA

- [1] E. Kim, M. Q. Ding, X. Wang, dan S. Lu, "Does Topic Consistency Matter? A Study of Critic and User Reviews in the Movie Industry," *J Mark*, no. September 2023, doi: 10.1177/00222429221127927.
- [2] K. A. Fachrudin, D. L. Tarigan, dan M. F. Iman, "Analisis Rating dan Harga Kamar Hotel Bintang Lima di Indonesia," *Jurnal Akuntansi, Keuangan, dan Manajemen*, vol. 3, no. 3, hlm. 237–252, 2022, doi: 10.35912/jakman.v3i3.1107.
- [3] G. Roy, "Travelers' online review on hotel performance – Analyzing facts with the Theory of Lodging and sentiment analysis," *Int J Hosp Manag*, vol. 111, hlm. 103459, Mei 2023, doi: 10.1016/J.IJHM.2023.103459.
- [4] Y. J. Kim dan H. S. Kim, "The Impact of Hotel Customer Experience on Customer Satisfaction through Online Reviews," *Sustainability (Switzerland)*, vol. 14, no. 2, hlm. 1–13, 2022, doi: 10.3390/su14020848.
- [5] D. Gelen, J. Cardenas, D. M. Gurung, D. Han, H. Ban, dan H. Kim, "The Text Mining from Online Customer Reviews: Implications for Luxury Hotel in Busan," *Culinary Science & Hospitality Research*, vol. 28, no. 2, hlm. 67–80, 2022, doi: 10.20878/cshr.2022.28.2.008.
- [6] Munawir, L. N. Fujianti, M. D. Koerniawan, dan B. Dewancker, "Text Mining by Social Network Data towards Developing Attractiveness of Urban Park. Case Study: Thematic Parks in Bandung City, Indonesia," *IOP Conf Ser Earth Environ Sci*, vol. 738, no. 1, 2021, doi: 10.1088/1755-1315/738/1/012057.
- [7] F. N. Hikmah, S. Basuki, dan Y. Azhar, "Deteksi Topik Tentang Tokoh Publik Politik Menggunakan Latent Dirichlet Allocation (LDA)," *Jurnal Repositor*, vol. 2, no. 4, hlm. 415–426, 2020, doi: 10.22219/repositor.v2i4.52.
- [8] R. Ramadandi, N. Yusliani, O. Arsalan, R. Kurniati, dan R. F. Isnanto, "Pemodelan Topik Menggunakan Metode Latent Dirichlet Allocation dan GIBBS Sampling," hlm. 74–79, 2021.
- [9] B. A. Tondang, M. R. Fadhil, M. N. Perdana, A. Fauzi, dan U. S. Janitra, "Analisis pemodelan topik ulasan aplikasi BNI , BCA , dan BRI menggunakan latent dirichlet allocation Modeling analysis of BNI , BCA , and BRI application review topics using latent dirichlet allocation," vol. 4, hlm. 114–127, 2023, doi: 10.37373/infotech.v4i1.601.
- [10] V. Shrivastava, "A Methodological study of web crawler," *Journal of Engineering Research and Application*, ISSN- 2248-9622, vol. 8, no. 11, 2018.
- [11] S. Tahvili, L. Hatvani, E. Ramentol, R. Pimentel, W. Afzal, dan F. Herrera, "A novel methodology to classify test cases using natural language processing and imbalanced learning," *Eng Appl Artif Intell*, vol. 95, hlm. 103878, Okt 2020, doi: 10.1016/J.ENGAPPAI.2020.103878.
- [12] S. Zhou, P. Kan, Q. Huang, dan J. Silbernagel, "A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura," *J Inf Sci*, no. August 2020, hlm. 1–15, 2021, doi: 10.1177/01655515211007724.