



PERBANDINGAN EFEKTIVITAS MODEL *NON-NEGATIVE MATRIX FACTORIZATION* (NMF) DAN TOP2VEC UNTUK PEMODELAN TOPIK (STUDI KASUS: *TRAVELOKA REVIEWS OF HOTEL KEMPINSKI JAKARTA*)

Dzakki Damar Rabbani¹, Fathir Habibie²

^{1,2}Departemen Statistika Bisnis, Fakultas Vokasi, Institut Teknologi Sepuluh Nopember (ITS)

¹dzakkirabbani99@gmail.com

Abstrak

Ulasan pelanggan di platform *Online Travel Agent* (OTA) seperti Traveloka telah menjadi aset digital krusial yang secara langsung memengaruhi reputasi dan keputusan pemesanan dalam industri perhotelan. Namun, volume ulasan dalam bentuk teks tidak terstruktur yang masif menyulitkan manajemen untuk mengekstrak wawasan strategis secara efisien. Informasi berharga mengenai sentimen terhadap fasilitas, kualitas layanan, dan aspek pengalaman lainnya berisiko terabaikan, sehingga menciptakan kebutuhan mendesak akan metode analisis otomatis. Penelitian ini bertujuan untuk membandingkan efektivitas dua pendekatan pemodelan topik *Non-Negative Matrix Factorization* (NMF) yang berbasis ko-okurensi statistik, dan Top2Vec yang berbasis kedekatan semantik, pada data ulasan Hotel Kempinski Jakarta. Setelah melalui tahap pra-pemrosesan teks, data yang didominasi ulasan positif ini dianalisis menggunakan kedua model. Hasilnya, model NMF berhasil mengekstrak 10 topik yang lebih spesifik dan granular seperti 'kebersihan kamar' dan 'sarapan pagi', dengan mencapai skor koherensi (C_v) sebesar 0.4813. Sebaliknya, model Top2Vec secara otomatis mengidentifikasi 3 topik konseptual yang lebih luas seperti 'pengalaman menginap keluarga', dengan skor koherensi 0.3140. Berdasarkan keunggulan skor koherensi yang signifikan, model NMF terbukti lebih baik dan dinilai paling efektif untuk mengekstrak topik yang koheren serta mudah diinterpretasikan. Temuan ini mengindikasikan bahwa untuk analisis teks ulasan dengan tujuan identifikasi aspek layanan yang spesifik, pendekatan NMF memberikan hasil yang lebih optimal.

Kata Kunci: Hotel Kempinski Jakarta, Natural Language Processing, NMF, Pemodelan Topik, Top2Vec

Article History:

Received: July 2025

Reviewed: July 2025

Published: July 2025

Plagiarism Checker No 234

Prefix DOI :

10.8734/Koehesi.v1i2.365

Copyright : Author

Publish by : Koehesi



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

PENDAHULUAN

Era digital saat ini, industri perhotelan mengalami transformasi signifikan yang didorong oleh kemajuan teknologi informasi. *Platform Online Travel Agent* (OTA) seperti Traveloka telah menjadi kanal utama bagi konsumen untuk mencari, memesan, dan memberikan ulasan mengenai pengalaman menginap mereka. Ulasan pelanggan, atau *user-generated content* (UGC), telah menjadi salah satu faktor paling berpengaruh dalam proses pengambilan keputusan konsumen, bahkan melebihi informasi pemasaran yang disediakan oleh pihak hotel itu sendiri [1]. Ketersediaan ulasan dalam volume besar menjadi sumber data yang sangat berharga untuk memahami persepsi dan kepuasan pelanggan secara mendalam.

Hotel Kempinski Jakarta, sebagai salah satu hotel mewah terkemuka di Indonesia, secara konsisten menerima ratusan hingga ribuan ulasan pelanggan di *platform* Traveloka. Ulasan-ulasan ini mengandung wawasan krusial mengenai berbagai aspek layanan, mulai dari kualitas kamar, kebersihan, fasilitas, hingga keramahan staf. Namun, volume dan sifat data yang tidak



terstruktur (berupa teks naratif) menjadi tantangan besar. Menganalisis data ini secara manual tidak hanya memakan waktu dan tenaga, tetapi juga rentan terhadap bias subjektif dari analisis [2]. Akibatnya, banyak wawasan penting yang dapat digunakan untuk peningkatan kualitas layanan dan strategi bisnis berisiko terlewatkan.

Dalam menjawab tantangan tersebut, diperlukan pendekatan komputasional yang mampu mengekstraksi tema-tema atau topik-topik utama dari kumpulan teks secara otomatis. *Topic modeling* adalah salah satu teknik dalam bidang *Natural Language Processing* (NLP) yang dirancang khusus untuk tujuan ini [3]. Metode ini memungkinkan analisis ribuan ulasan secara efisien untuk mengidentifikasi pola-pola tersembunyi, seperti keluhan yang sering muncul atau pujian terhadap aspek layanan tertentu, yang dapat menjadi dasar pengambilan keputusan berbasis data.

Dalam *domain topic modeling*, terdapat berbagai algoritma dengan pendekatan yang berbeda. Dua model yang menonjol adalah *Non-Negative Matrix Factorization* (NMF) dan *Top2Vec*. NMF merupakan model faktorisasi matriks klasik yang efektif dalam mengurai korpus teks menjadi matriks topik dan matriks bobot kata [4]. Di sisi lain, *Top2Vec* adalah model yang lebih modern yang memanfaatkan *word embeddings* untuk menemukan topik berdasarkan kedekatan semantik antar kata dan dokumen. Keunggulan *Top2Vec* terletak pada kemampuannya menemukan topik yang lebih koheren secara kontekstual dan secara otomatis menentukan jumlah topik yang optimal [5]. Perbedaan fundamental antara pendekatan matematis (NMF) dan semantik (*Top2Vec*) ini memunculkan pertanyaan mengenai model mana yang lebih efektif untuk studi kasus ulasan perhotelan.

Oleh karena itu, penelitian ini bertujuan untuk menerapkan dan membandingkan efektivitas model NMF dan *Top2Vec* dalam mengidentifikasi topik-topik utama dari ulasan pelanggan Hotel Kempinski Jakarta yang dipublikasikan di *platform* Traveloka. Hasil dari penelitian ini diharapkan dapat memberikan dua kontribusi utama. Pertama, memberikan wawasan yang dapat ditindaklanjuti bagi manajemen Hotel Kempinski Jakarta untuk meningkatkan kualitas layanan berdasarkan umpan balik otentik dari pelanggan. Kedua, menyajikan perbandingan metodologis yang dapat menjadi referensi bagi para peneliti atau praktisi di industri perhotelan dalam memilih model pemodelan topik yang paling sesuai untuk menganalisis data ulasan pelanggan.

URAIAN PENELITIAN

Hotel Kempinski Jakarta

Hotel Kempinski Jakarta adalah salah satu hotel bintang lima yang berlokasi strategis di pusat kota Jakarta, Indonesia. Sebagai bagian dari jaringan perhotelan mewah Kempinski yang bertaraf internasional, hotel ini dikenal dengan standar layanannya yang tinggi, fasilitas premium, dan posisinya sebagai ikon kemewahan di ibu kota [6]. Dalam konteks penelitian ini, Hotel Kempinski Jakarta berfungsi sebagai objek studi kasus, di mana ulasan-ulasan pelanggan yang ditinggalkan di *platform* digital seperti Traveloka menjadi sumber data primer untuk dianalisis. Ulasan tersebut merefleksikan persepsi dan pengalaman nyata pelanggan terhadap berbagai aspek layanan hotel.

Ulasan Pelanggan (*Customer Review*)

Ulasan pelanggan, yang juga dikenal sebagai *User-Generated Content* (UGC), adalah bentuk umpan balik yang diberikan oleh konsumen mengenai produk, layanan, atau pengalaman yang mereka terima. Ulasan ini biasanya dipublikasikan di platform digital seperti media sosial, situs *e-commerce*, atau *Online Travel Agents* (OTA) [7]. Dalam konteks industri perhotelan, ulasan pelanggan menjadi sumber informasi krusial yang tidak hanya memengaruhi keputusan pemesanan calon tamu lain, tetapi juga berfungsi sebagai alat evaluasi kinerja bagi manajemen hotel untuk mengidentifikasi keunggulan dan area yang memerlukan perbaikan [8].



Natural Language Processing (NLP)

Natural Language Processing (NLP) atau Pemrosesan Bahasa Alami adalah cabang dari ilmu komputer dan kecerdasan buatan (AI) yang berfokus pada interaksi antara komputer dengan bahasa manusia. Tujuan utama NLP adalah untuk memungkinkan mesin dalam membaca, memahami, menginterpretasikan, dan menghasilkan bahasa manusia dengan cara yang bernilai. Teknik-teknik dalam NLP mencakup berbagai tugas, seperti analisis sentimen, terjemahan mesin, peringkasan teks otomatis, dan pemodelan topik, yang memungkinkan ekstraksi wawasan dari data teks dalam skala besar [9].

Pemodelan Topik (Topic Modeling)

Pemodelan topik adalah sebuah metode statistik dan *unsupervised machine learning* yang digunakan untuk menemukan struktur tematik atau "topik" abstrak yang tersembunyi dalam suatu kumpulan dokumen (korpus). Algoritma pemodelan topik tidak memerlukan label data terlebih dahulu; sebaliknya, ia menganalisis pola kemunculan bersama kata-kata dalam dokumen untuk mengelompokkannya ke dalam topik-topik tertentu. Hasilnya adalah representasi setiap topik sebagai distribusi kata dan setiap dokumen sebagai distribusi topik, sehingga memungkinkan analisis tematik konten secara kuantitatif dan efisien [10].

Non-Negative Matrix Factorization (NMF)

Non-Negative Matrix Factorization (NMF) adalah salah satu algoritma yang dapat digunakan untuk pemodelan topik. NMF bekerja dengan cara menguraikan (memfaktorkan) matriks data input, seperti matriks term-dokumen, menjadi dua matriks berdimensi lebih rendah. Salah satu matriks merepresentasikan topik (sebagai kumpulan kata), dan matriks lainnya merepresentasikan bobot setiap topik dalam setiap dokumen. Keunikan NMF terletak pada kendala non-negatifnya, yang memastikan bahwa semua elemen dalam matriks hasil bersifat positif, sehingga hasilnya lebih mudah diinterpretasikan sebagai bagian-bagian yang saling menambah [11].

Word Embeddings

Word Embeddings adalah representasi kata dalam bentuk vektor numerik berdimensi rendah. Dalam pendekatan ini, kata-kata dari suatu korpus dipetakan ke dalam ruang vektor di mana kata-kata yang memiliki makna atau konteks serupa akan memiliki posisi yang berdekatan. Representasi vektor ini memungkinkan algoritma *machine learning* untuk memahami hubungan semantik antar kata, yang menjadi dasar bagi model-model NLP yang lebih canggih [12].

Top2Vec

Top2Vec adalah algoritma pemodelan topik modern yang memanfaatkan *word embeddings* untuk mengidentifikasi topik. Berbeda dengan model tradisional yang mengandalkan frekuensi kata, *Top2Vec* secara bersamaan membuat *embedding* untuk dokumen dan kata, lalu mengidentifikasi klaster-klaster dokumen yang padat di dalam ruang vektor tersebut. Pusat (*centroid*) dari setiap klaster ini dianggap sebagai vektor topik. Keunggulan utama *Top2Vec* adalah kemampuannya untuk secara otomatis menentukan jumlah topik yang relevan dan menghasilkan topik yang lebih koheren secara semantik karena didasarkan pada kedekatan kontekstual, bukan sekadar ko-okurensi kata [13].

METODE PENELITIAN

Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh melalui proses *web scraping* dari laman aplikasi layanan pemesanan hotel, yaitu Traveloka. Data yang dikumpulkan berupa ulasan pengguna (*user reviews*) terhadap Hotel Kempinski Jakarta, mencakup *rating* (penilaian bintang) dan isi ulasan teks (*review content*) yang dipublikasikan oleh pengguna aplikasi. Pengambilan data dilakukan untuk periode 1 Januari hingga 31 Desember 2024, sehingga mencerminkan opini dan persepsi pelanggan sepanjang tahun tersebut.



Variabel Penelitian

Adapun variabel-variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan	Skala	Satuan
X ₁	Rating	Rasio	Skala 1-10
X ₂	Review	Data Teks	-

Struktur Data

Struktur data yang digunakan dalam penelitian ini dapat dituliskan dalam Tabel 3.2 sebagai berikut.

Tabel 3.2 Struktur Data

<i>n</i>	X ₁ (Rating)	X ₂ (Review)
1	10,0	“puas bgt. next kesini lg bersama keluarga. overall is good”
2	10,0	“Kamar besar view bundaran HI juga dapet. Ke Grand Indonesia ada akses khususnya. Staffnya friendly friendly. Makanannya enak enak. Recommended staycay disini kalo lagi punya extra rejeki.”
...
...
1427	10,0	Sempurna! Hotel ini merupakan hotel terbaik, ketika saya pemesanan untuk grand room deluxe, saya mendapatkan salon kamar suite! Dan saya menyukainya!

Langkah Analisis

Berikut merupakan langkah analisis yang digunakan pada penelitian ini.

1. Pengumpulan & Pemuatan Data: Data ulasan diekstraksi dan dimuat ke dalam lingkungan analisis Python menggunakan library Pandas.
2. Pra-pemrosesan Teks: Tahap ini krusial untuk membersihkan dan menstandarisasi data teks. Prosesnya meliputi:
 - a. *Case Folding*: Mengubah semua teks menjadi huruf kecil.
 - b. *Punctuation Removal*: Menghapus tanda baca.
 - c. *Tokenization*: Memecah kalimat menjadi kata-kata (token).
 - d. *Stopword Removal*: Menghapus kata-kata umum dalam Bahasa Indonesia yang tidak membawa makna signifikan (misalnya, "dan", "di", "yang").
 - e. *Stemming*: Mengubah kata-kata ke bentuk dasarnya menggunakan library Sastrawi.
3. Analisis Karakteristik Data: Menganalisis data yang telah bersih untuk memahami karakteristiknya, seperti distribusi sentimen dan frekuensi kata kunci utama. Langkah ini dilakukan melalui visualisasi data (diagram batang dan *word cloud*) untuk mendapatkan wawasan awal mengenai *dataset*.
4. Vektorisasi Data (untuk NMF): Teks yang telah bersih diubah menjadi representasi numerik menggunakan TF-IDF *Vectorizer*, yang mengukur pentingnya sebuah kata dalam sebuah dokumen relatif terhadap keseluruhan korpus.



5. Implementasi Model NMF: Model NMF diterapkan pada matriks TF-IDF untuk mengekstraksi sejumlah topik yang telah ditentukan sebelumnya.
6. Implementasi Model Top2Vec: Model Top2Vec diterapkan langsung pada data teks yang telah dibersihkan (namun umumnya tanpa *stemming* dan *stopword removal* yang agresif, karena model ini memanfaatkan konteks kalimat).
7. Evaluasi Model: Kualitas topik dari kedua model diukur menggunakan skor koherensi topik (C_v) dengan bantuan library Gensim.
8. Analisis Perbandingan: Hasil dari kedua model, baik secara kualitatif (topik yang dihasilkan) maupun kuantitatif (skor koherensi), dibandingkan untuk menentukan model yang lebih efektif.
9. Menginterpretasikan hasil analisis.
10. Menarik Kesimpulan dan saran.

HASIL DAN PEMBAHASAN

Hasil Pra-pemrosesan Teks

Tahap pra-pemrosesan berhasil mengubah ulasan mentah yang beragam menjadi format teks yang terstruktur, yang siap untuk dianalisis oleh mesin. Proses ini secara efektif menghilangkan informasi yang tidak relevan, seperti (tanda baca, URL, emoji) menyeragamkan kosakata (melalui *case folding* dan *stemming*). Tabel 4.1 di bawah ini menunjukkan contoh hasil dari sebuah ulasan pada tahapan pra-pemrosesan teks.

Tabel 4.1 Proses Pra-pemrosesan Teks

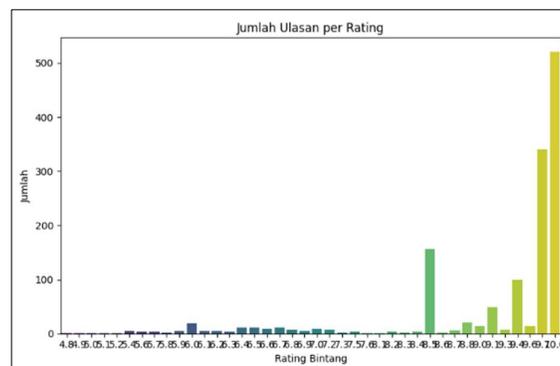
Tahap	Contoh Kalimat
Teks Asli	"Hotel yang sangat bagus dari segi pelayanan, fasilitas dan view dari dalam kamar 🍷 🍷"
Hasil Akhir (ulasan bersih)	"hotel bagus segi layan fasilitas view kamar"

Karakteristik Data Ulasan

Setelah teks dibersihkan, berikut merupakan karakteristik dari data ulasan yang akan dijelaskan sebagai berikut.

1) Sebaran Data Rating

Berikut merupakan sebaran data *rating* yang ditampilkan pada Gambar 4.1.



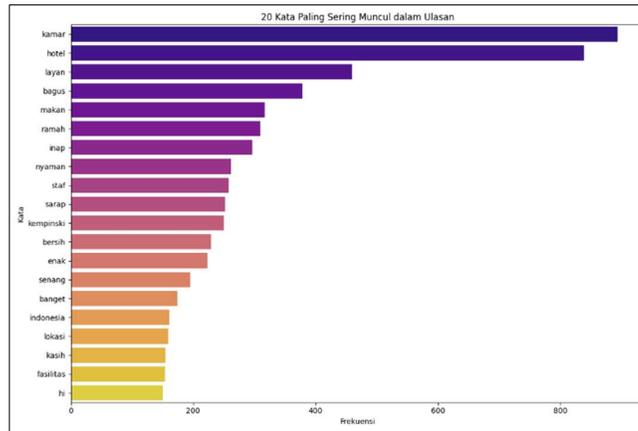
Gambar 4.1 Sebaran Data Rating

Berdasarkan gambar 4.1 sebaran data rating menunjukkan dominasi ulasan dengan nilai tinggi, terutama pada rating 10, yang dapat mencerminkan kualitas produk/jasa yang sangat memuaskan atau adanya bias dalam perilaku pemberian ulasan oleh pengguna.



2) Kata Kunci Utama Ulasan

Berikut merupakan kata kunci utama ulasan yang ditampilkan pada Gambar 4.2.



Gambar 4.2 Kata Kunci Utama Ulasan

Berdasarkan gambar 4.2 bahwa topik utama yang dibahas oleh pengguna dalam ulasan sangat berkaitan dengan pengalaman menginap dan kualitas pelayanan. Kata “kamar” dan “hotel” muncul dengan frekuensi tertinggi, menandakan bahwa aspek fisik dan kenyamanan akomodasi menjadi fokus utama perhatian tamu. Kata “layan”, “ramah”, dan “staf” juga muncul cukup sering, mencerminkan pentingnya kualitas pelayanan dan interaksi dengan petugas hotel dalam membentuk kepuasan pelanggan.

3) Wordcloud Semua Ulasan

Berikut merupakan wordcloud dari semua ulasan yang ditampilkan pada Gambar 4.3.



Gambar 4.3 Kata Kunci Utama Ulasan

Berdasarkan Gambar 4.3 bahwa *wordcloud* dari semua ulasan memberikan gambaran visual yang memperkuat hasil analisis frekuensi kata. Kata-kata yang tampil dengan ukuran besar seperti “kamar”, “hotel”, “layan”, “makan”, “bagus”, dan “ramah” menandakan bahwa topik-topik tersebut paling sering disebut oleh para pengguna. Ini menunjukkan bahwa aspek fisik kamar, kualitas layanan, dan pengalaman kuliner merupakan elemen yang paling berkesan bagi tamu.

B. Implementasi Model Pemodelan Topik

Setelah memahami karakteristik data, kedua model pemodelan topik diimplementasikan untuk mengekstrak struktur tematik yang lebih dalam untuk mengukur tingkat Efektivitas Perbandingan dari Model *Non-Negative Matrix Factorization* (NMF) dan *Top2Vec*.



1) *Topik dari Model Non-Negative Matrix Factorization (NMF)*

Berikut merupakan hasil topik dari Model *Non-negative Matrix Factorization (NMF)* yang mampu mengelompokkan kata-kata berdasarkan tingkat ko-okurensi atau kemunculan bersamaan dalam satu dokumen. Hasil pemodelan disajikan pada Tabel 4.2, yang menampilkan 10 topik utama.

Tabel 4.2 Topik yang Dihasilkan oleh Model NMF

No	Contoh Kalimat
1	bagus, tinggal, hotel, bersih, sarap, staycation, habis, fasilitas, enak, pandang
2	hotel, jakarta, indonesia, bintang, kota, kempinski, suka, inap, salah, grand
3	kamar, luas, mandi, sarap, bersih, hi, pandang, bundar, jam, deluxe
4	inap, fasilitas, perfect, service, checkin, cepat, indonesia, harga, mewah
5	layan, hebat, sempurna, prima, ramah, tingkat, kamar, hotel, suasana
6	staf, ramah, bantu, sarap, alam, ruang, benarbenar, inap, anak-anak, bayi
7	nyaman, bersih, tinggal, tenang, senang, rekomendasi, keluarga, kamar, mall, amp
8	lokasi, strategis, pusat, fasilitas, kota, belanja, mal, makan, suasana, mall
9	senang, terima, kasih, kempinski, inap, alam, tinggal, milik, traveloka, kali
10	makan, enak, banget, nya, oke, staff, breakfast, variasi, lengkap, ramahramah

2) *Topik dari Model Top2Vec*

Berikut merupakan hasil topik dari Model Top2Vec yang bekerja berdasarkan kedekatan semantik, secara otomatis mengidentifikasi bahwa terdapat 3 topik yang optimal dalam dataset ini. Topik yang dihasilkan cenderung lebih konseptual, seperti yang ditunjukkan pada Tabel 4.3.

Tabel 4.3 Topik yang Dihasilkan oleh Model Top2Vec

No	Contoh Kalimat
1	hotel, jakarta, breakfastnya, staffnya, restaurant, booking, traveloka, barang, keluarga, belanja
2	senang, keluarga, satusatunya, barang, kerja, siang, kesan, sejarah, sayang, kasih
3	jakarta, strategis, indonesia, senang, siang, keluarga, sejarah, booking, staffnya, kerja

C. *Analisis Perbandingan Efektivitas Kinerja Model*

Dalam menentukan model mana yang lebih efektif, dilakukan perbandingan kuantitatif menggunakan metrik Skor Koherensi (C_v). Hasil evaluasi, yang diekstraksi dari output notebook, disajikan pada Tabel 4.4.

Tabel 4.4 Analisis Perbandingan Efektivitas Kinerja Model

Metode	Jumlah Topik	Skor Koherensi (C_v)
1	10	0.4813
2	3	0.3140



Berdasarkan Tabel 4.4 terkait analisis perbandingan efektivitas kinerja model, bahwa model *Non-Negative Matrix Factorization* (NMF) terbukti secara kuantitatif lebih unggul daripada model Top2Vec. Hal ini didasarkan pada perolehan Skor Koherensi (C_v) yang lebih tinggi secara signifikan, di mana model NMF mencapai nilai 0.4813, sementara model Top2Vec memperoleh nilai 0.3140. Nilai yang lebih tinggi pada NMF ini menunjukkan bahwa topik-topik yang dihasilkannya memiliki koherensi semantik yang lebih baik dan lebih mudah diinterpretasikan, sehingga menjadikan NMF sebagai model yang lebih efektif dan optimal untuk pemodelan topik dalam penelitian ini.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, dapat disimpulkan sebagai berikut.

1. Proses pra-pemrosesan teks berhasil mengubah data ulasan mentah yang beragam menjadi format teks terstruktur yang bersih. Tahapan ini secara efektif menghilangkan informasi tidak relevan seperti tanda baca dan emoji serta menyeragamkan kosakata melalui *case folding* dan *stemming*, sehingga data siap untuk dianalisis oleh model.
2. Karakteristik data ulasan didominasi oleh rating bernilai tinggi, terutama rating 10, yang mengindikasikan tingkat kepuasan pelanggan yang tinggi. Analisis frekuensi dan *wordcloud* menunjukkan bahwa kata kunci utama yang paling sering muncul adalah "kamar", "hotel", "layan", "makan", dan "ramah", yang menegaskan bahwa aspek akomodasi fisik, kualitas pelayanan, dan pengalaman kuliner menjadi elemen paling berkesan bagi pengguna.
3. Implementasi model pemodelan topik berhasil mengekstrak struktur tematik dari data ulasan. Model *Non-Negative Matrix Factorization* (NMF) menghasilkan 10 topik yang didasarkan pada ko-okurensi kata, seperti topik terkait kebersihan, fasilitas, dan staf. Sementara itu, model Top2Vec secara otomatis mengidentifikasi 3 topik optimal yang lebih konseptual berdasarkan kedekatan semantik, seperti topik seputar pengalaman menginap bersama keluarga dan proses pemesanan.
4. Analisis perbandingan efektivitas kinerja model secara kuantitatif menunjukkan bahwa model *Non-Negative Matrix Factorization* (NMF) terbukti lebih unggul dibandingkan model Top2Vec. Hal ini didasarkan pada perolehan Skor Koherensi (C_v) yang lebih tinggi secara signifikan, di mana NMF mencapai nilai 0.4813 sementara Top2Vec memperoleh nilai 0.3140, menjadikan NMF sebagai model yang lebih efektif dan optimal untuk pemodelan topik dalam penelitian ini.

Saran

Berdasarkan hasil penelitian yang telah dipaparkan, saran praktis yang dapat dilakukan untuk para pemangku kepentingan yaitu, bagi manajemen Hotel Kempinski Jakarta, disarankan untuk memanfaatkan hasil pemodelan topik dari model NMF sebagai dasbor pemantauan kepuasan pelanggan, di mana sepuluh topik spesifik yang ditemukan dapat menjadi metrik kunci untuk evaluasi kinerja operasional. Tim marketing juga dapat mempertajam strategi komunikasi dengan menonjolkan aspek yang paling berkesan seperti "kamar", "layan", dan "makan" dalam materi promosi untuk memperkuat citra positif. Lebih luas lagi, bagi industri perhotelan, studi ini menunjukkan bahwa pendekatan analisis otomatis pada ulasan pelanggan merupakan metode efektif untuk mendapatkan wawasan kompetitif.

Bagi pihak *platform Online Travel Agent* (OTA) seperti Traveloka disarankan untuk mempertimbangkan integrasi fitur peringkasan topik otomatis pada halaman ulasan guna meningkatkan nilai tambah bagi pengguna dan mitra hotel.

Untuk penelitian selanjutnya, terdapat beberapa arah pengembangan yang potensial. Disarankan untuk memperluas studi dengan melakukan perbandingan model NMF dengan algoritma yang lebih baru seperti BERTopic, serta mengintegrasikan analisis topik dengan analisis sentimen untuk mendapatkan wawasan yang lebih dalam. Selain itu, perluasan ruang lingkup data juga krusial, baik dengan menggunakan data dari periode waktu yang lebih panjang



untuk mengidentifikasi tren, maupun dengan melakukan studi komparatif terhadap hotel kompetitor untuk memperoleh gambaran pasar yang lebih komprehensif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, bimbingan, dan bantuan dalam penyelesaian penelitian ini. Terima kasih pertama disampaikan kepada Institut Teknologi Sepuluh Nopember (ITS) atas fasilitas dan kesempatan yang diberikan kepada penulis untuk melaksanakan penelitian ini sebagai bagian dari program studi D4 Statistika Bisnis.

Penulis juga menyampaikan rasa hormat dan terima kasih yang mendalam kepada Ibu Noviyanti Santoso, Ph.D., dosen Mata Kuliah Metode Data Tidak Terstruktur kelas 6B, yang telah memberikan arahan, saran, serta dukungan moral selama proses penelitian dan penyusunan paper ini.

Ucapan terima kasih yang tulus juga disampaikan kepada keluarga dan teman-teman yang selalu memberikan semangat, dukungan, dan pengertian yang sangat berarti bagi penulis dalam menyelesaikan penelitian ini.

Akhir kata, penulis berterima kasih kepada semua pihak yang telah berkontribusi, baik secara langsung maupun tidak langsung, dalam menyelesaikan penelitian ini. Semoga hasil penelitian ini dapat memberikan manfaat bagi pembaca dan berkontribusi positif dalam pengembangan ilmu pengetahuan di bidang analisis teks dan pemodelan topik.

DAFTAR PUSTAKA

- [1] B. A. & P. H. E. M. M. S. Sparks, "Hotel booking intentions as a function of the number of pictures and consumer-generated reviews," *Journal of Hospitality & Tourism Research*, vol. 37, no. 2, pp. 154-173, 2013.
- [2] B. Liu., "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, p. 1-167, 2012.
- [3] D. M. Blei., "Probabilistic Topic Models," *Communications of the ACM*, vol. 55, no. 4, pp. 77-84, 2012.
- [4] D. & W. H. L. Kuang, "Non-negative matrix factorization for interactive topic modeling and document clustering," in *Proceedings of the 24th International Conference on World Wide Web*, Florence, 2015.
- [5] D. Angelov, "Top2Vec: Distributed Representations of Topics," arXiv, 2020.
- [6] H. K. Jakarta, "Company Profile & Service Directory Kempinski Group Indonesia," Kempinski Group Indonesia, Jakarta, 2024.
- [7] B. Liu, "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1-167, 2012.
- [8] A. S. & S. F. Cantalops, "New consumer behavior: A review of research on eWOM and hotels," *International Journal of Hospitality Management*, vol. 36, pp. 41-51, 2014.
- [9] D. & M. J. H. Jurafsky, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (3rd ed.)*, Prentice Hall: Pearson International Edition, 2019.
- [10] D. M. Blei, "Probabilistic Topic Models," *Communications of the ACM*, vol. 55, no. 4, pp. 77-84, 2012.
- [11] D. & W. H. Kuang, "Non-negative matrix factorization for interactive topic modeling and document clustering," in *Proceedings of the 24th International Conference on World Wide Web*, Florence, Italy, 2015.



- [12] T. C. K. C. G. & D. J. Mikolov, "Efficient Estimation of Word Representations in Vector Space," arXiv preprint arXiv:1301.3781, 2013.
- [13] D. Angelov, "Top2Vec: Distributed Representations of Topics," arXiv, 2020.