

Analisis Sentimen Berbasis Aspek pada Layanan Hotel di Wilayah Kabupaten Banyumas dengan Word2Vec dan Random Forest

Sena Wijayanto^{1*}, Dedy Agung Prabowo², Daniel Yeri Kristiyanto³, M Yoka Fathoni⁴

^{1,3,4}Prodi Sistem Informasi, Fakultas Informatika, Institut Teknologi Telkom Purwokerto, Purwokerto

²Prodi Teknik Informatika, Fakultas Informatika, Institut Teknologi Telkom Purwokerto, Purwokerto

^{1,2,3,4}Jln. DI Panjaitan No.128, Kota Purwokerto, 53147, Indonesia

email: ¹sena@ittelkom-pwt.ac.id, ²dedy@ittelkom-pwt.ac.id, ³daniel@ittelkom-pwt.ac.id, ⁴myokafathoni@ittelkom-pwt.ac.id

Abstract – In the tourism industry, hotels have an important role to help tourists because they provide lodging, especially for tourists from out of town. The quality of hotel services can be seen from the opinions given by visitors who have stayed at the hotel. This study aims to conduct a sentiment analysis of the reviews given by hotel visitors. The review data was taken from Traveloka using web scrapping. The method used for feature extraction is word2vec. For sentiment classification, the method used is random forest. The best experimental results were obtained from the experimental results using the number of trees 100, 200, and 300 with an accuracy of 82%-83%.

Abstrak – Dalam industri pariwisata, hotel memiliki peran penting untuk membantu wisatawan karena menyediakan penginapan terutama bagi wisatawan dari luar kota. Kualitas layanan hotel dapat dilihat dari opini-opini yang diberikan oleh pengunjung yang telah menginap di hotel tersebut. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan yang diberikan oleh pengunjung hotel. Data ulasan tersebut diambil dari Traveloka menggunakan web scrapping. Metode yang digunakan untuk ekstraksi fitur adalah word2vec. Untuk klasifikasi sentimen, metode yang digunakan adalah random forest. Hasil percobaan terbaik didapatkan dari hasil percobaan dengan menggunakan jumlah tree 100, 200, dan 300 dengan hasil akurasi sebesar 82%-83%.

Kata Kunci – analisis sentimen; ulasan hotel; word2vec; random forest

I. PENDAHULUAN

Berwisata merupakan salah satu kebutuhan yang perlu dipenuhi oleh semua orang. Hal tersebut dikarenakan wisata menjadi sarana dalam menghilangkan penat. Kaum milenial melakukan wisata paling tidak minimal setahun sekali [1]. Dalam industri pariwisata, hotel memiliki peran yang penting dalam membantu wisatawan yang berkunjung dengan menyediakan penginapan terutama bagi wisatawan dari luar kota. Di wilayah Kabupaten Banyumas terdapat 179 hotel baik hotel berbintang maupun hotel non berbintang [2]. Hotel-hotel tersebut akan menyediakan berbagai pelayanan untuk para wisatawan yang menginap.

Kualitas layanan hotel dapat dilihat dari opini-opini yang diberikan oleh pengunjung yang menginap di hotel [3]. Para wisatawan akan menuliskan pengalaman tentang layanan yang dialaminya pada platform online seperti pada media sosial, Traveloka, TripAdvisor, google maps, dsb. Pengalaman yang dituliskan tersebut dapat berupa ulasan positif, negatif, maupun netral atau biasa saja [4]. Ulasan-ulasan dari pengunjung hotel akan mempengaruhi keputusan dari calon pengunjung baru hotel tersebut [5]. Ulasan tersebut juga dapat digunakan oleh pihak hotel untuk memahami pendapat dari pengunjung tentang layanan yang diterima sehingga layanan-layanan yang ada dapat diperbaiki [6]. Namun jumlah ulasan yang banyak dan beragam akan menyebabkan pihak hotel kesulitan memahami ulasan-ulasan tersebut [7]. Maka perlu membuat terobosan untuk mengatasi permasalahan tersebut sehingga pihak hotel akan memahami ulasan-ulasan yang diterima secara efektif dan efisien.

Analisis sentimen merupakan salah satu teknik yang dapat digunakan untuk menganalisis ulasan dari pengunjung hotel. Analisis sentimen dapat digunakan untuk melihat suatu opini yang ditujukan untuk hotel termasuk positif maupun negatif [8]. Terdapat permasalahan yang dihadapi ketika melakukan analisis sentimen yaitu data tidak terstruktur dan kurangnya akurasi hasil klasifikasi [9]. Banyak penelitian yang melakukan analisis sentimen pada suatu ulasan [3][4][10]. Namun analisis sentimen pada penelitian tersebut hanya mengelompokkan sentimen positif dan negatif saja. Berdasarkan permasalahan tersebut maka penelitian ini akan melakukan analisis sentimen hingga pada aspek-aspek layanan yang ada di hotel terutama hotel di wilayah Kabupaten Banyumas menggunakan word2vec dan random forest. Luaran dari penelitian ini diharapkan mengetahui sentimen positif atau negatif pada tingkat aspek terkait layanan hotel di wilayah Kabupaten Banyumas.

II. PENELITIAN YANG TERKAIT

Penelitian-penelitian sebelumnya yang membahas tentang analisis sentimen diantaranya adalah penelitian yang dilakukan oleh Fauzi pada tahun 2018 menganalisis sentimen pada bahasa Indonesia menggunakan *Random Forest*. Pada penelitian tersebut menghasilkan akurasi sebesar 82%. Ekstraksi fitur yang digunakan adalah *bag of word* dengan berbagai macam pembobotan yaitu *Binary TF*, *Raw TF*, *Logarithmic TF* dan *TF.IDF* namun tidak menghasilkan perbedaan hasil yang signifikan [11].

*) penulis korespondensi: Sena Wijayanto

Email: sena@ittelkom-pwt.ac.id

Pada penelitian yang dilakukan oleh Setiowati dan Helen pada tahun 2018 melakukan analisis sentimen pada hotel di Yogyakarta. Pada penelitian ini opini masyarakat yang didapatkan dari *website* pemesanan hotel dikelompokkan menjadi positif, negatif dan netral. Hasil yang didapatkan ketika menggunakan algoritma *Random Forest* adalah ketika menggunakan ekstraksi fitur TF dan tanpa *stopword*[3].

Penelitian lainnya yaitu dilakukan oleh Baskoro, dkk pada tahun 2021 melakukan analisis sentimen menggunakan *Random Forest* dan *TF.IDF* pada ulasan hotel di Purwokerto. Ulasan hotel yang diperoleh dikelompokkan menjadi ulasan positif dan ulasan negatif. Hasil akurasi yang dihasilkan adalah 87,23 %. Proses stemming yang dilakukan tidak berkontribusi dalam peningkatan akurasi [7].

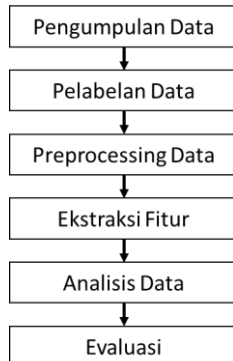
Selanjutnya penelitian yang dilakukan oleh khomsah pada tahun 2021 untuk menganalisis sentimen pada komentar di *Youtube*. Metode yang dipakai adalah *Random Forest* dan *Word2Vec* untuk mengekstraksi fitur. Pada percobaan yang dilakukan didapatkan ukuran *window* yang menghasilkan akurasi terbesar adalah 10. Hasil akurasi yang dihasilkan yaitu 88,95% [12].

Penelitian yang dilakukan Amien, dkk pada tahun 2021 menganalisis kategori aspek pada ulasan berbahasa Indonesia. Pada penelitian ini membandingkan kinerja *Naïve Bayes (NB)*, *Support Vector Machine (SVM)*, dan *Random Forest (RF)*. Hasil yang didapatkan adalah dengan menggunakan metode *Random Forest* akan menghasilkan kinerja paling bagus[13].

III. METODE PENELITIAN

A. Alur Penelitian

Gambar 1 menunjukkan tahapan-tahap yang dilaksanakan pada penelitian ini.



Gbr. 1 Alur penelitian.

B. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data dengan cara menggunakan *web scrapping*. Data diperoleh dari *website* Traveloka dengan mengambil ulasan-ulasan pada hotel di Kabupaten Banyumas. Data ulasan yang diambil yaitu user, ulasan, tanggal, dan nilai *rating*. Contoh hasil pengumpulan data ulasan hotel yang telah didapatkan dapat dilihat di tabel I.

C. Pelabelan Data

Tahap selanjutnya yaitu menambahkan label. Label tersebut yaitu label sentimen berupa positif atau negatif dan juga label aspek yaitu aspek yang diulas pada ulasan hotel. Aspek-aspek tersebut antara lain *service*, *hotel*, *room*, *location*, *price*, *restaurant* [14].

D. Preprocessing

Setelah data berupa ulasan-ulasan hotel diberikan label, langkah selanjutnya adalah *preprocessing*. Di tahap ini data ulasan yang telah didapatkan akan dibersihkan agar data memiliki kualitas yang baik. Selain itu ulasan yang sama persis atau duplikat akan dihapus. Langkah-langkah *preprocessing* yang dilakukan yaitu *tokenization*, menghilangkan simbol/angka, *case folding*, dan *stopword removal*. *Tokenization* berfungsi untuk memecah kalimat menjadi kumpulan kata. *Case folding* digunakan untuk mengubah huruf besar pada ulasan menjadi huruf kecil. *Stopword removal* bertujuan menghilangkan kata-kata tidak bermakna atau tidak mewakili dokumen.

TABEL I
CONTOH HASIL PENGUMPULAN DATA ULASAN HOTEL

User	Date	Ratin g	Ulasan
Heri S.	9/18/2022	8.5	Bagus, lumayan nyaman untuk keluarga.
Nicky g.	5/23/2022	5.4	Sprei dan sarung bantal bau tidak sedap, resepsionis kurang menyenangkan.

E. Ekstraksi Fitur

Setelah data dilakukan *preprocessing* selanjutnya dilakukan ekstraksi fitur. Sebelum melakukan ekstraksi fitur, data ulasan akan dibagi 2 yaitu data training sebanyak 70% dan data testing sebanyak 30%. Ekstraksi fitur pada penelitian ini menggunakan *Word2Vec*. *Word2Vec* merupakan teknik atau metode untuk merepresentasikan atau mengubah kata menjadi vektor pada suatu N dimensi. Parameter *window* yang digunakan pada *Word2Vec* yaitu 10.

F. Analisis Data

Langkah Selanjutnya merupakan analisis sentimen menggunakan *Random Forest*. Pada penelitian ini akan dilakukan perbandingan percobaan banyak *tree* yaitu 100, 200, dan 300.

G. Evaluasi

Evaluasi yang digunakan pada penelitian ini yaitu akurasi. Akurasi dapat menghitung rasio dari hasil prediksi terhadap data sebenarnya. Perhitungan akurasi menggunakan rumus berikut :

$$Akurasi = \frac{(TP+TN)}{(TP+FP+FN+TN)} p(x, y) ; (0 \leq x \leq M - 1, 0 \leq y \leq N - 1) \quad (1)$$

Keterangan :

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

IV. HASIL DAN PEMBAHASAN

Hasil dari *web scrapping* pada *website* Traveloka dilakukan *preprocessing* sehingga didapatkan data akhir sebanyak 1407

ulasan. Setelah itu dilakukan pelabelan data. Untuk pelabelan sentimen, dari hasil analisis ulasan dengan ratingnya didapatkan hasil ketika rating kurang dari 8 maka akan memiliki sentimen negatif, sedangkan untuk ulasan yang memiliki rating 8 atau lebih memiliki sentimen positif. Dan untuk pelabelan aspek dilakukan secara manual sesuai dengan isi dari ulasan. Contoh hasil pelabelan dapat dilihat pada tabel II.

TABEL II
CONTOH HASIL PELABELAN DATA

Ratin g	Ulasan	Class	Aspect
8.5	Bagus, lumayan nyaman untuk keluarga.	Positif	Room
5.4	Sprei dan sarung bantal bau tidak sedap, resepsionis kurang menyenangkan.	Negatif	Service

Untuk ekstraksi fitur, metode yang digunakan adalah Word2Vec. Metode ini akan merepresentasikan kata menjadi vektor. Contoh hasil dari Word2Vec bisa dilihat di tabel III.

TABEL III
CONTOH HASIL WORD2VEC

Kata	Vektor
lumayan	0.06705109775066376
baru	0.11563082784414291
dingin	0.0821416974067688
bagus	0.11563082784414291

Untuk klasifikasi sentimen menggunakan metode *Random Forest*. Dilakukan 3 kali percobaan untuk mencari parameter terbaik pada *Random Forest* yaitu menggunakan banyak *tree* sebesar 100, 200, dan 300. Hasil klasifikasi dapat dilihat di tabel IV.

TABEL IV
HASIL KLASIFIKASI SENTIMEN

Banyak tree	Akurasi
100	82%
200	83%
300	83%

Dari hasil tersebut bisa dilihat terdapat perubahan akurasi namun tidak terlalu signifikan. Akurasi terbesar yang didapatkan yaitu sebesar 83%.

V. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen pada ulasan hotel di Kabupaten Banyumas. Analisis menggunakan metode *Random Forest*. Hasil percobaan yang didapatkan dengan percobaan menggunakan banyak *tree* sebesar 100, 200, dan 300 memiliki hasil akurasi yang tidak jauh berbeda, yaitu sebesar 82-83%.

UCAPAN TERIMA KASIH

Ucapan terima kasih penulis kepada pihak Kemdikbudristek dan Institut Teknologi Telkom Purwokerto yang telah membantu dan memberikan dukungan terkait dengan penelitian yang dilakukan seperti bantuan fasilitas penelitian, dana hibah, dan lainnya.

4 DAFTAR PUSTAKA

- [1] Ramayani, Yusuf, and M. Veranita, "Minat Berwisata Kaum Milenial di Era New," *J. Kepariwisata Indonesia*, vol. 15, no. 2, pp. 158–167, 2021.
- [2] BPS Kabupaten Banyumas, "Jumlah Hotel dan Jasa Akomodasi di Kabupaten Banyumas, 2015-2018." <https://banyumaskab.bps.go.id/statictable/2020/03/31/245/jumlah-hotel-dan-jasa-akomodasi-menurut-klasifikasi-di-kabupaten-banyumas-2015---2018.html> (accessed Feb. 17, 2022).
- [3] Y. Setiowati and A. Helen, "Klasifikasi Analisis Sentimen Mengenai Hotel di Yogyakarta," *SCAN-Jurnal Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 1–10, 2018.
- [4] O. Somantri, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 5, no. 2, pp. 191–196, 2019, [Online]. Available: www.google.com/maps.
- [5] S. A. Amira, S. Utama, and H. Fahmi, "Penerapan Metode Support Vector Machine untuk Analisis Sentimen pada Review Pelanggan Hotel," *Edu Komputika J.*, vol. 7, no. 2, pp. 40–48, 2020, doi: 10.15294/edukomputika.v7i2.42608.
- [6] R. Cahyani, Indriati, and P. Pandu Adikara, "Analisis Sentimen terhadap Ulasan Hotel menggunakan Boosting Weighted Extreme Learning Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 8, pp. 7767–7773, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [7] B. B. Baskoro, I. Susanto, and S. Khomsah, "Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)," *J. INISTA*, vol. 3, no. 2, pp. 21–29, 2021, doi: 10.20895/INISTA.V3.
- [8] R. A. Setiawan, A. K. A. Esthetika, E. M. O. Nurhayanto, Y. Asmara, and A. Wahyudi, "Analisis Sentimen Hotel di Nusa Tenggara Barat Menggunakan Algoritma SVM Hotel Sentiment Analysis in West Nusa Tenggara Using the SVM Algorithm," in *Seminar Multimedia dan Artificial Intelligence : Optimalisasi Artificial Intelligence di Era Revolusi Industri 4.0 dan Society 5.0*, 2021, vol. 4, pp. 149–155.
- [9] S. Khomsah, A. F. Hidayatullah, and A. S. Aribowo, "Comparison of the Effects of Feature Selection and Tree-Based Ensemble Machine Learning for Sentiment Analysis on Indonesian YouTube Comments," *Lect. Notes Electr. Eng.*, vol. 746 LNEE, pp. 161–172, 2021, doi: 10.1007/978-981-33-6926-9_15/COVER.
- [10] W. Widayat, "Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 3, p. 1018, 2021, doi: 10.30865/mib.v5i3.3111.
- [11] M. A. Fauzi, "Random forest approach for sentiment analysis in Indonesian language," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 1, pp. 46–50, 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
- [12] S. Khomsah, "Sentiment Analysis On YouTube Comments Using Word2Vec and Random Forest," *Telematika*, vol. 18, no. 1, p. 61, 2021, doi: 10.31315/telematika.v18i1.4493.
- [13] S. Amien, P. Perdana, T. Bharata Aji, and R. Ferdiana, "Aspect Category Classification dengan Pendekatan Machine Learning Menggunakan Dataset Bahasa Indonesia (Aspect Category Classification with Machine Learning Approach Using Indonesian Language Dataset)," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 3, pp. 229–235, 2021.
- [14] S. Cahyaningtyas, D. Hatta Fudholi, and A. Fathan Hidayatullah, "Deep Learning for Aspect-Based Sentiment Analysis on Indonesian Hotels Reviews," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, 2021, doi: 10.22219/kinetik.v6i3.1300.