

Analisis Sentimen Aspek pada Opini Wisatawan menggunakan Metode Naïve Bayes (Studi Kasus: Perum Perhutani)

Fasya Ghassani Hadiyan¹, Fitra Abdurrachman Bachtiar²

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹fasya1996@gmail.com, ²fitra.bachtiar@ub.ac.id

Abstrak

Perum Perhutani merupakan BUMN (Badan Usaha Milik Negara) yang salah satu portofolio bisnisnya adalah mengelola dan mengembangkan 768 ekowisata di Jawa dan Madura. Untuk mengembangkan objek wisata yang dikelolanya Perhutani belum dapat melakukannya dengan optimal dikarenakan kurangnya data dan informasi mengenai preferensi wisatawan mengenai objek-objek wisata di samping upaya yang dilakukannya. Analisis sentimen menggunakan *Multinomial Naïve Bayes* dilakukan untuk mengetahui preferensi wisatawan dilihat dari tiga aspek terhadap objek wisata yang dikelola Perhutani. Ketiga aspek tersebut adalah *attractions*, *accessibilities*, dan *amenities*. Ulasan didapatkan dari TripAdvisor yang akan diklasifikasikan ke dalam tiga kelas, yaitu positif, negatif, dan netral. Tujuan dari penelitian ini, yaitu untuk mengetahui hasil implementasi metode *Multinomial Naïve Bayes* pada proses analisis sentimen, mengetahui nilai *precision*, *recall*, dan *f1*, dan terakhir mengetahui hasil usability testing dari dashboard. Hasil proses klasifikasi untuk aspek *attractions* berupa nilai *accuracy* sebesar 69,89%. Sedangkan untuk aspek *accessibilities* dan *amenities* didapatkan nilai *accuracy* sebesar 61,67% dan 63,01% untuk masing-masing aspek. Informasi berupa sentimen wisatawan tersebut ditampilkan ke dalam *dashboard* yang kemudian diujikan kepada pihak Perhutani menggunakan SUS (*System Usability Scale*) dan mendapatkan skor 65. Hasil skor tersebut menandakan dashboard memiliki rating “Good” dan termasuk ke dalam *range Marginal (High)* untuk penerimaan.

Kata kunci: analisis sentimen, multinomial naïve bayes, klasifikasi, objek wisata, dashboard

Abstract

Perum Perhutani is a BUMN (Badan Usaha Milik Negara) whose business portfolio is managing and developing 768 ecotourism in Java and Madura. Perhutani has not been able to develop the tourism objects optimally due to the lack of data and information regarding tourist preferences about tourism objects despite the effort they have done. Sentiment analysis using *Multinomial Naïve Bayes* was carried out to determine tourist preferences from three aspects of tourism objects managed by Perhutani. The three aspects are *attractions*, *accessibilities*, and *amenities*. Reviews obtained from TripAdvisor will be classified into three classes, namely positive, negative, and neutral. The purposes of this study are first, to find out the implementation results of *Multinomial Naïve Bayes* in sentiment analysis process, second, to find out the values of *precision*, *recall*, and *f1*, and lastly to know the results of usability testing from the dashboard. The results of the classification process for the *attractions* aspect is 69.89% of accuracy. As for *accessibilities* and *amenities*, the accuracy values are 61.67% and 63.01% for each aspect. Information about of tourist sentiment is displayed on a dashboard which is then tested by Perhutani using SUS (*System Usability Scale*) and gets a score of 65. That result means the dashboard has “Good” rating and in *range Marginal (High)* for acceptability.

Keywords: sentiment analysis, multinomial naïve bayes, classification, tourism objects, dashboard

1. PENDAHULUAN

Perum Perhutani atau biasa dikenal Perhutani adalah Badan Usaha Milik Negara (BUMN) yang berperan dalam pengelolaan hutan dan hasil hutan di Pulau Jawa dan Madura.

Selain itu, Perhutani bertugas untuk mendukung Kementerian Pariwisata dan Ekonomi Kreatif dalam mencapai sasaran strategis, yaitu diantaranya mengelola serta meningkatkan kualitas dan kuantitas 768 objek-objek wisata di

Pulau Jawa. Namun, Perhutani belum dapat mengelola dan mengembangkan objek-objek wisata tersebut dengan optimal dikarenakan kurangnya data dan informasi mengenai opini wisatawan terkait objek wisata.

Data ulasan wisatawan yang terdapat di media sosial dapat dimanfaatkan oleh Perhutani sebagai salah satu alat pendukung untuk mencapai sasaran strategis tersebut. Salah satu aplikasi web pada bidang pariwisata, yaitu TripAdvisor memfasilitasi para penggunanya untuk memberikan komentar serta foto terkait pengalaman mereka pada sebuah objek wisata (TripAdvisor, 2021). Akan tetapi, data yang ada pada TripAdvisor belum langsung bisa digunakan untuk pengambilan keputusan. Meskipun data yang tersedia banyak, namun data yang berasal dari pengguna media sosial tersebut memiliki banyak noise dan tidak terstruktur (Liu, 2012).

Salah satu metode yang bisa digunakan untuk mengatasi permasalahan tersebut adalah *data mining*. Tetapi, data ulasan ini belum menjadi informasi berguna bagi pihak Perhutani dalam pengambilan keputusan dikarenakan data berupa teks memiliki banyak *noise* dan tidak terstruktur (Liu, 2012). Penelitian ini bertujuan untuk mengetahui hasil analisis sentimen terhadap opini atau sentimen wisatawan yang didapat dari TripAdvisor sehingga data ulasan tersebut dapat menjadi informasi yang bisa digunakan oleh Perhutani dalam pengambilan keputusan.

Berdasarkan hasil wawancara dengan pihak Perum Perhutani di penelitian ini dibahas tiga aspek di bidang pariwisata, yaitu aspek *Attractions*, *Accessibilities*, dan *Amenities*. Metode analisis sentimen digunakan untuk mendapatkan distribusi opini wisatawan di ketiga aspek tersebut. Dengan menerapkan metode Multinomial Naïve Bayes dari seluruh data ulasan kita akan mengetahui apakah setiap aspek di atas dapat diklasifikasikan ke dalam kelas positif, netral, atau negatif dengan nilai keakuratan tertentu yang dihasilkan oleh metode ini secara simultan.

Hasil analisis sentimen di penelitian ini dapat digunakan pada business intelligence untuk membantu pihak manajemen Perum Perhutani dalam mengambil keputusan pengembangan objek wisata yang dikelolanya. Untuk kemanfaatan hasil ini penulis memberikan rancangan dari *dashboard* hasil analisis sentimen, yang telah diuji menggunakan SUS (System Usability Score). Hasil analisis

sentiment dan rancangan dashboard dinilai pihak Perhutani dapat diimplementasikan untuk kepentingan bisnis pengelolaan objek wisata yang dikelola oleh Perhutani.

Makalah ini ditulis dengan sistematika berikut ini. Setelah pendahuluan ini, pada Bagian Dua akan dibahas sumber-sumber pustaka yang digunakan atau dirujuk pada penelitian ini. Setelah itu di Bagian 3 akan dijelaskan metodologi yang digunakan di penelitian ini. Bagian 4 akan menampilkan hasil analisis sentimen dan juga rancangan dashboard. Bagian 5 menampilkan kesimpulan hasil penelitian dan diskusi untuk penyempurnaan penelitian serupa di masa datang. Makalah diakhiri daftar pustaka yang ditulis menjadi Bagian 6.

2. SUMBER PUSTAKA

Penelitian yang dilakukan mengacu kepada beberapa penelitian sebelumnya. Penelitian yang digunakan sebagai acuan penelitian ini oleh Geetha et al. (2017) menganalisis relasi antara *review* dan *rating* yang diberikan pengunjung hotel serta menganalisis konsistensi *review* dan *rating* terhadap kategori hotel. Penelitian ini menggunakan algoritme *Naïve Bayes* untuk mengklasifikasikan *review* pengunjung hotel menjadi positif, negatif, serta netral. Pada penelitian ini juga dicari kategori-kategori yang sering muncul untuk mengetahui preferensi pengunjung terhadap layanan hotel menggunakan *hierarchical clustering*. Hasil dari penelitian ini membuktikan bahwa *review* dan *rating* yang diberikan pengunjung dapat membantu pihak manajerial dalam melakukan perbaikan.

Sementara itu analisis sentimen di bidang pariwisata yang lainnya dilakukan oleh Prameswari et al. (2017) yang meneliti *review* pengunjung hotel-hotel di Bali pada TripAdvisor dengan lebih memperhatikan aspek layanan dan fasilitas hotel. Analisis sentimen dilakukan menggunakan algoritme *Recursive Neural Tensor Network* (RNTN). Hasil akurasi yang didapatkan, yaitu sebesar 85%.

Penelitian oleh Chang (2017) juga masih meneliti analisis sentimen untuk bidang pariwisata. Penelitian ini bertujuan untuk menganalisis keterkaitan antara *review* hotel dan *rating* hotel Hilton di Amerika Serikat pada TripAdvisor. Penelitian ini bertujuan untuk memvisualisasikan *rating* dan *review* hotel Hilton pada *dashboard Google Trends* sehingga

dapat membantu pihak pemasaran. Pada penelitian ini analisis sentimen dilakukan menggunakan metode *Convulsion Tree Kernel Classsification* untuk mengklasifikasikan *review* ke dalam label positif dan negatif. Hasil penelitian ini, yaitu akurasi sebesar 89,3%.

Penelitian selanjutnya oleh Bachtiar et al. (2020) bertujuan untuk melakukan analisis sentimen mengenai aspek *location*, *room*, *food*, *price*, dan *service* pada industri hotel. Analisis sentimen pada tingkat aspek dilakukan untuk mengetahui kepuasan pelanggan pada *guest house* sehingga dapat menghasilkan E-WOM (*Electronic Word of Mouth*) yang positif. Data yang digunakan merupakan ulasan berbahasa Indonesia yang didapat dari situs web yang berfokus pada biro perjalanan seperti Agoda.com, Expedia, dan Booking.com. Hasil dari *accuracy* dari analisis sentimen ini yang menggunakan metode SVM adalah sebesar 93% untuk aspek *location*, 80% untuk aspek *room*, 68% pada aspek *food*, 90% *accuracy* aspek *price*, dan 92% pada aspek *service*.

Penelitian terbaru tentang analisis sentimen di bidang pariwisata dapat dilihat di Paolanti, et al. (2021) yang memaparkan permasalahan informasi geografis terkait satu destinasi pariwisata yang sangat terkenal di Italia, yaitu Cilento. Metode yang digunakan pada penelitian ini adalah metode-metode *deep learning* dengan *dataset* yang diperoleh dari twitter. Penelitian ini bertujuan untuk membuat karakteristik dari pergerakan wisatawan secara spasial, temporal, dan demografis sepanjang daerah wisata pedesaan dan sepanjang pantai di Cilento.

Metode *Naïve Bayes* tidak hanya diterapkan untuk analisis sentimen di bidang pariwisata. Sebagai contoh penelitian oleh Arulselvi (2017) bertujuan untuk mengklasifikasikan sentimen mengenai buku pada topik-topik tertentu sehingga memudahkan pelajar untuk mengetahui buku-buku yang sering dibaca dari topik tertentu. Klasifikasi sentimen dilakukan menggunakan metode *Naïve Bayes Classifier* ke dalam tiga kelas (positif, negatif, dan netral). Pada penelitian ini juga dilakukan analisis tren untuk mengetahui judul buku apa saja yang populer pada rentang waktu tertentu. Akurasi klasifikasi *Naïve Bayes* pada penelitian ini sebesar 89.9%.

Penelitian lainnya dilakukan oleh Xu et al. (2020) di mana di penelitian ini klasifikasi menggunakan metode *Naïve Bayes* dilakukan untuk mengklasifikasikan beberapa *dataset*, yaitu *movie*, *network*, dan *health* di beberapa *e-*

commerce. Selain itu, penelitian ini juga membahas proses *continous learning* agar model mampu melakukan klasifikasi secara otomatis ketika *review* tentang produk terus bertambah.

Penelitian-penelitian tersebut dijadikan sebagai acuan pada penelitian yang dilakukan karena memiliki beberapa kesamaan. Untuk menggali sentimen dari wisatawan digunakan analisis sentimen dengan algoritme *Multinomial Naïve Bayes* untuk klasifikasi sama seperti dua penelitian yang sudah dipaparkan. Selain itu pada penelitian ini akan dicari kategori layanan objek wisata yang sering muncul.

3. METODOLOGI PENELITIAN

Untuk mendapatkan hasil yang diinginkan, beberapa langkah penelitian dibuat, seperti yang diperlihatkan oleh Gambar 1.



Gambar 1. Alur Metodologi Penelitian

Berikut penjelasan singkat langkah-langkah yang dilakukan di penelitian ini.

3.1 Identifikasi Masalah

Masalah yang diselesaikan di makalah ini didapat setelah adanya kontak dan diskusi dengan pihak manajemen Divisi Keuangan, Umum dan SDM Perum Perhutani. Dari beberapa diskusi terlihat kebutuhan Perum Perhutani untuk mengetahui sentimen dari para wisatawan yang telah mengunjungi objek-objek wisata yang dikelola oleh Perum Perhutani. Karena Perum Perhutani belum melakukan

assessment terhadap wisatawan secara langsung, maka digunakan analisis sentimen dengan data yang diperoleh dari media sosial dalam hal ini data ulasan dari TripAdvisor.

3.2 Studi Literatur

Setelah permasalahan teridentifikasi, selanjutnya dikumpulkan literatur-literatur yang digunakan sebagai rujukan penelitian. Literatur yang menjadi rujukan di penelitian ini di antaranya ditampilkan pada bagian sebelumnya.

3.3 Pengumpulan Data dan Informasi

Setelah melakukan identifikasi masalah untuk melanjutkan penelitian dibutuhkan data dan informasi yang mendukung jalannya penelitian. Dilakukan wawancara dengan cara menyusun pertanyaan-pertanyaan terstruktur yang ditanyakan kepada narasumber. Narasumber pada penelitian kali ini adalah Manajer Keuangan, SDM, dan Umum Perum Perhutani yang berperan dalam pengelolaan objek-objek wisata. Data dan informasi yang dikumpulkan terkait dengan kepentingan berjalannya penelitian dan dokumentasi penelitian.

Selain data dan informasi yang terkumpul melalui wawancara narasumber, terdapat data penting lain yang dibutuhkan. Data tersebut berupa komentar-komentar wisatawan yang diambil dari salah satu situs web yang mengulas objek-objek wisata yang dikelola oleh Perum Perhutani, yaitu TripAdvisor. Ulasan wisatawan tersebut diambil menggunakan cara yang disebut *web scraping* menggunakan *tools WebHarvy*. Data ini digunakan agar bisa melakukan tahapan selanjutnya, yaitu analisis sentimen.

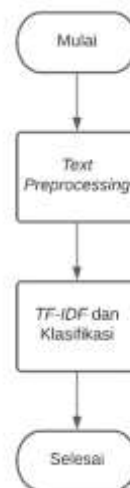
3.4 Analisis Sentimen

Pada tahap ini dibutuhkan masukan berupa data yang telah didapatkan pada tahap *web scraping* yang sudah terlabel ke dalam aspek (*Attractions*, *Accessibilities*, atau *Amenities*) dan sentimen (positif, negatif, atau netral) yang tepat. Tahap pertama dilakukan *text preprocessing* yang bertujuan untuk mempersiapkan data untuk kepentingan klasifikasi seperti pada Gambar 3. Dikarenakan data yang dimiliki berupa teks, maka sebelum dilakukan klasifikasi dilakukan beberapa sub tahapan termasuk konversi teks menjadi angka supaya data bisa diolah untuk proses klasifikasi atau lebih dikenal dengan sebutan *feature extraction* yang pada penelitian ini dilakukan menggunakan TF-IDF. Setelah itu dilakukan klasifikasi menggunakan metode

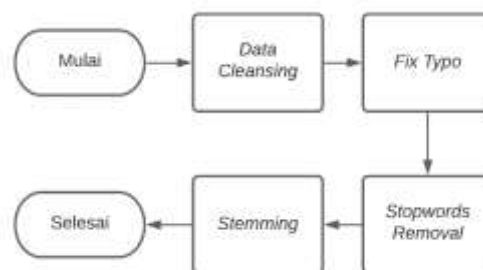
Multinomial Naive Bayes seperti tertera pada Persamaan (1) untuk menentukan sentimen wisatawan kepada kelas positif, netral, atau negatif.

$$P(w_i|c_j) = \frac{\text{count}(w_i, c_j) + 1}{(\sum_{w \in V} \text{count}(w, c_j)) + |V|} \quad (1)$$

Secara rinci proses analisis sentimen dapat dilihat di gambar berikut.



Gambar 2. Tahapan pada Analisis Sentimen di penelitian ini



Gambar 3. Tahapan pada text preprocessing

3.5 Analisis dan Pengujian Hasil Klasifikasi

Hasil klasifikasi mengenai ulasan wisatawan dianalisis untuk mengetahui performa model yang dibentuk. Terdapat beberapa pengukuran yang dilakukan, yaitu *accuracy*, *precision*, *recall*, dan *f1 score*. *Accuracy* pada menunjukkan keseluruhan performa model. Sedangkan *precision* dan *recall* dibutuhkan untuk memilih model terbaik berdasarkan tingginya jumlah *false positive* untuk *precision* atau *false negative* untuk *recall*. Terakhir, *f1 score* dibutuhkan untuk mencari keseimbangan antara *precision* dan *recall*.

$$\text{accuracy} = \frac{tp+tn}{tp+fp+fn+tn} \quad (2)$$

$$p = \frac{tp}{tp+fp} \quad (3)$$

$$r = \frac{tp}{tp+fn} \quad (4)$$

$$f_1 = \frac{2pr}{p+r} \quad (5)$$

3.6 Visualisasi Data

Data yang digunakan sebagai sumber penelitian dan data yang didapat dari penelitian disajikan ke dalam bentuk *dashboard*. Data-data tersebut disajikan ke dalam grafik-grafik yang membentuk informasi terkait objek-objek wisata yang dikelola oleh Perum Perhutani. *Dashboard* ini diperuntukkan oleh pihak manajemen Divisi SDM, Keuangan, dan Umum yang berperan dalam mengelola objek-objek wisata yang dikelola Perum Perhutani. Sehingga *dashboard* yang disusun sesuai dengan kebutuhan dengan pemegang kepentingan tersebut.

3.7 Penarikan Kesimpulan dan Saran

Tahap terakhir dari penelitian ini adalah penarikan kesimpulan dari hasil penelitian yang telah dilakukan. Penarikan kesimpulan dilakukan supaya bisa menjawab pertanyaan rumusan masalah yang sudah dikemukakan sebelumnya. Selain kesimpulan terdapat beberapa saran yang diberikan untuk penelitian serupa selanjutnya berdasarkan penelitian yang dilakukan kali ini.

4. HASIL DAN PEMBAHASAN

Setelah seluruh proses analisis sentimen dilakukan diperoleh nilai *accuracy*, *precision*, *recall*, dan *f1* untuk ketiga aspek yang diteliti sebagai berikut.

4.1 Aspek Attractions

Untuk aspek *attractions* diperoleh hasil *accuracy* 0,69898 dan nilai parameter lainnya dapat dilihat pada Tabel 1.

Tabel 1. Evaluasi hasil klasifikasi aspek *attractions*

Accuracy	0,69898		
	Precision	Recall	F1
Negatif	0,84	0,44	0,58
Netral	0,57	0,28	0,38
Positif	0,7	0,97	0,81

4.2 Aspek Accessibilities

Untuk aspek *Accessibilities* diperoleh hasil *accuracy* 0,61667 dan nilai parameter lainnya dapat dilihat pada Tabel 2.

Tabel 2. Evaluasi hasil klasifikasi aspek *accessibilities*

Accuracy	0,61667		
	Precision	Recall	F1
Negatif	0,5	0,14	0,22
Netral	0,62	0,94	0,75
Positif	0,71	0,16	0,26

4.3 Aspek Amenities

Untuk aspek *amenities* diperoleh hasil *accuracy* 0,63106 dan nilai parameter lainnya dapat dilihat pada Tabel 3.

Tabel 3. Evaluasi hasil klasifikasi aspek *amenities*

Accuracy	0,63106		
	Precision	Recall	F1
Negatif	1	0,21	0,34
Netral	0,68	0,65	0,67
Positif	0,57	0,87	0,69

Dari ketiga nilai *accuracy* yang diperoleh dapat disimpulkan bahwa metode yang digunakan dapat melakukan klasifikasi sentimen dengan cukup baik, walaupun nilai yang diperoleh masih dapat ditingkatkan beberapa cara yang sudah dipikirkan oleh penulis. Usulan cara perbaikan dapat dilihat di bagian Kesimpulan dan Saran.

4.4 Visualisasi Data

Data hasil analisis dipresentasikan ke dalam *dashboard* yang terdiri dari tiga grafik yang dapat dilihat pada Gambar 4. Grafik pertama berupa *bar chart* yang menampilkan jumlah sentimen pada ketiga aspek. Grafik kedua berupa *line chart* yang menampilkan tren aspek dari tahun ke tahun. Grafik terakhir berupa histogram menampilkan objek-objek wisata.



Gambar 4. *Dashboard*

Untuk mengetahui seberapa baik pihak Perhutani dalam memahami *dashboard*, dilakukan *usability testing*. *Usability testing* pada penelitian ini menggunakan alat ukur yang dikenal dengan *System Usability Score* (SUS) oleh Bangor et al. (2008) seperti pada Tabel 4. Pernyataan bernomor ganjil merupakan pernyataan positif sedangkan pernyataan bernomor ganjil berupa pernyataan negatif. Pihak Perhutani diminta untuk mengisi kuesioner dengan angka dengan interval 0 (sangat tidak setuju) hingga 4 (sangat setuju). Untuk mendapatkan skor SUS perlu dilakukan perhitungan dengan cara:

1. Untuk pernyataan bernomor ganjil, maka nilai - 1.
2. Untuk pernyataan bernomor genap, maka 5 - nilai.
3. Hasil perhitungan dari setiap pertanyaan kemudian dijumlahkan lalu dikalikan 2,5 untuk mendapat hasil akhir.

Tabel 4. Hasil *System Usability Score* (SUS)

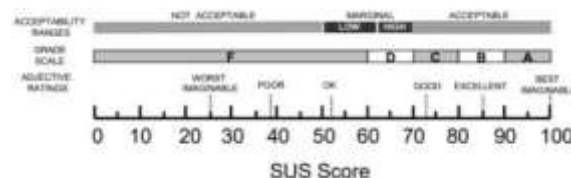
No.	Pertanyaan	Nilai
1.	Saya pikir akan sering menggunakan produk	3
2.	Saya beranggapan bahwa produk terlalu kompleks	3
3.	Saya pikir produk ini mudah digunakan	3
4.	Saya rasa saya akan membutuhkan bantuan dari teknisi untuk bisa menggunakan produk	1
5.	Saya menemukan fungsi-fungsi di dalam produk terintegrasi dengan baik	3
6.	Saya pikir terdapat terlalu banyak ketidakselarasan dalam produk	1
7.	Saya berpendapat kebanyakan orang dapat mempelajari penggunaan produk dengan cepat	3
8.	Saya menemukan bahwa produk sangat rumit ketika digunakan	1
9.	Saya sangat percaya diri dalam menggunakan produk	3
10.	Saya butuh belajar lebih banyak lagi sebelum bisa menggunakan produk	3

Berikut perhitungan dan hasil SUS.

$$\begin{aligned}
 \text{SUS} &= ((3-1) + (5-3) + (3-1) + (5-1) + (3-1) + (5-1) \\
 &\quad + (3-1) + (5-1) + (3-1) + (5-3)) * 2,5 \\
 &= (2+2+2+4+2+4+2+4+2+2) * 2,5 \\
 &= 65
 \end{aligned}$$

Berdasarkan perhitungan yang telah dilakukan, didapatkan nilai SUS sebesar 65. Hasil tersebut mengindikasikan bahwa

Dashboard tergolong ke dalam rating “Good” dan kategori *Marginal (high)* berdasarkan SUS Score pada Gambar 5. Artinya *dashboard* memiliki tingkat penerimaan dalam kategori marjinal atau belum memenuhi kategori *acceptable* oleh pihak Perum Perhutani seperti pada Gambar 5.



Gambar 5. SUS Score

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Pada penelitian ini telah diteliti analisis sentimen aspek terhadap tiga aspek pariwisata, yaitu *attraction*, *accessibilities*, dan *amenities* dari objek-objek wisata yang dikelola oleh Perum Perhutani. Metode yang digunakan adalah Multinomial Naive Bayes dengan data yang digunakan adalah data hasil *web scraping* ulasan dari TripAdvisor. Nilai-nilai *accuracy* untuk ketiga aspek tersebut adalah 0,69 untuk aspek *attractions*, 0,61 untuk aspek *accessibilities*, dan 0,63 untuk aspek *amenities*. Nilai-nilai ini dapat dikategorikan cukup baik dan masih bisa ditingkatkan dengan penyempurnaan pada hal pelabelan data mentah.

Hasil penerapan analisis sentimen ini ditampilkan dalam suatu *dashboard* yang dapat digunakan oleh Perum Perhutani sebagai informasi pelengkap dalam pengambilan keputusan strategis. Hasil *usability testing* menggunakan SUS yang dilakukan bernilai 65 yang berarti rancangan *dashboard* berada pada level marjinal (*high*) atau belum sampai pada level *acceptable*.

5.2 Saran

Nilai-nilai *accuracy* di skripsi ini masih pada kisaran 0,6. Penulis menyadari nilai ini masih dapat ditingkatkan dengan beberapa cara berikut ini.

1. Data yang diperoleh dari TripAdvisor adalah paragraf atau kalimat yang tertulis jauh dari tata bahasa baku. Proses melabeli kalimat-kalimat tersebut secara otomatis bukanlah perkara mudah, dengan kata lain memerlukan metode-metode tambahan di luar penggunaan *dictionary* seperti yang dilakukan di penelitian skripsi ini. Metode

yang dimaksud adalah metode dengan banyak tahapan, yang memuat tes keacakan, seperti yang disarankan oleh Vallantin (2019).

2. Dari data mentah yang diperoleh didapatkan kalimat-kalimat yang memuat aspek attraction jauh lebih banyak dibandingkan kalimat yang memuat kedua aspek lainnya. Kurang banyaknya data untuk kedua aspek ini penulis duga menjadi penyebab mengapa nilai *accuracy* untuk aspek *accessibilities* dan *amenities* lebih kecil dibandingkan aspek attraction. Untuk penelitian berikutnya perlu dipertimbangkan keseimbangan banyaknya data untuk setiap aspek.
3. Mengingat terdapat banyak metode lain untuk analisis sentimen, penulis menyarankan penerapan metode-metode lainnya selain metode yang digunakan di skripsi ini agar mendapatkan perbandingan nilai *accuracy* untuk ketiga aspek tersebut.
4. Sebenarnya ada dua aspek lain yang sering dipertimbangkan di bidang pariwisata, yaitu aspek *activities* dan *accommodation*. Untuk mendapatkan informasi yang lengkap Perum Perhutani dapat menerapkan analisis sentimen untuk menilai kedua aspek tersebut.

6. DAFTAR PUSTAKA

- Alaeri, A., Becken, S., & Stantic, B., 2017. Sentiment Analysis in Tourism: Capitalizing Big Data. *Journal of Travel Research*, [e-journal] 58(2), p.175-191. Tersedia melalui: Perpustakaan Universitas BX <<http://perpustakaan.ubx.ac.id>> [Diakses 14 Juni 2021]
- Alpaydin, E. (2010). *Introduction to Machine Learning*. 2nd ed. Cambridge: MIT Press.
- Arulselvi, A. C., Sendhilkumar, S., & Mahalakshmi, S., 2018. Classification of tweets for sentiment and trend analysis. *Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems*, pp.566–573.
- Aydin, O., 2018. *R Web Scraping Quick Start Guide: Techniques and tools to crawl and scrape data from websites*. Brimingham: Packt Publishing.
- Bachtiar, F. A., Paulina, W., & Rusydi, A. N., 2020. Text Mining for Aspect Based Sentiment Analysis on Customer Review : a Case Study in the Hotel Industry. *5th International Workshop on Innovations in Information and Communication Science and Technology*.
- Bangor, A., Staff, T., Kortum, P., Miller, J., & Staff, T. (2008). Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies*, 4(3), p.114–123.
- Berry, M. W., & Kogan, J., 2010. *Text Mining: Applications and Theory*. John Wiley & Sons, Inc.
- Chang, Y. C., Ku, C. H., & Chen, C. H., 2017. Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *International Journal of Information Management*.
- Cooper, C., Fletcher, J., Gilbert D. & Wanhill, S., 1993. *Tourism: Principles and Practice*. Harlow: Longman Group Limited.
- Feldman, R., & Sanger, J., 2007. *The Text Mining Handbook*. Cambridge: Cambridge University Press.
- Ferdiana, R., 2019. *Dataset Indonesia untuk Analisis Sentimen*. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 8(4).
- Few, S. 2006. *Information Dashboard Design (The Effect)*. Italy: O'Reilly Media, Inc.
- Geetha, M., Singha, P., & Sinha, S., 2017. Relationship between Customer Sentiment and Online Customer Ratings for Hotels - An Empirical Analysis. *Tourism Management*, 61, p.43-54.
- Han, J., Kamber, M. & Pei, J., 2012. *Data Mining Concepts and Techniques*. 3rd ed. Massachussets: Morgan Kaufmann.
- Liu, B., 2012. *Sentiment Analysis and Opinion Mining*. Massachussets: Morgan & Claypool Publishers.
- Manning, C.D., Raghavan, P., & Schutze, H., 2009. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

- Mirjalili, V., & Raschka, S., 2018. *Python Machine Learning*. 3rd ed. Brimingham: Packt Publishing.
- Müller, A. C., & Guido, S., 2015. *Introduction to Machine Learning with Python and Scikit-Learn*. California: O'Reilly Media.
- Navlani, A., 2018. *Naive Bayes Classification Using Scikit-learn*. Tersedia di: <<https://www.datacamp.com/community/tutorials/naive-bayes-scikit-learn>> [Diakses tanggal 23 Juni 2021]
- NLP Stanford, 2009. *Extracting Common Sentiments from Review*. [pdf] NLP Stanford. Tersedia di: <<https://nlp.stanford.edu/courses/cs224n/2009/fp/14.pdf>> [Diakses 13 Juni 2021].
- Paolanti, M., Mancini, A., Frontoni, E., Felicetti, A., Marinelli, L., Marcheggiani, E., & Pierdicca, R., 2021. Tourism destination management using sentiment analysis and geo-location information: a deep learning approach. *Information Technology & Tourism*, 23, p.241–264.
- Perhutani, 2021. Visi, Misi, dan Tata Nilai. [online] Tersedia di: <<https://perhutani.co.id/tentang-kami/visi-misi-dan-tata-nilai/>> [Diakses 14 Juni 2021]
- Prameswari, P., Surjandari, I., & Laoh, E., 2017. Opinion mining from online reviews in Bali tourist area. *International Conference on Science in Information Technology: Theory and Application of IT for Education, Industry and Society in Big Data Era*, pp.226–230.
- TripAdvisor, 2021. *About TripAdvisor*. [online] Tersedia di: <<https://tripadvisor.mediaroom.com/us-about-us>> [Diakses 14 Juni 2021]
- Turban, E., Sharda, S., Delen, D. & King, D., 2010. *Business Intelligence A Managerial Approach*. 2nd ed. New Jersey: Prentice Hall.
- Vallantin, L., 2019. *How to label text for sentiment analysis-good practices*. [online] Tersedia di: <<https://towardsdatascience.com/how-to-label-text-for-sentiment-analysis-good-practises-2dce9e470708>> [Diakses 14 Juni 2021]
- WEF, 2019. *The Travel & Tourism Competitiveness Report 2019*. [pdf] World Economic Forum. Tersedia di: <http://www3.weforum.org/docs/WEF_TTCR_2019.pdf> [Diakses 14 April 2021]
- Xu, F., Pan, Z., & Xia, R., 2020. E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. *Information Processing and Management*, 57.