
Analisa Kepuasan Tamu Hotel Menggunakan Algoritma RegEx dan VADER

Abdullah Rustamaji¹⁾, Roy Rudolf Huizen²⁾, Dandy Pramana Hostiadi³⁾

Program Studi Magister Sistem Informasi
Institut Teknologi dan Bisnis STIKOM Bali
Denpasar, Indonesia

e-mail: 222012027@stikom-bali.ac.id¹⁾, roy@stikom-bali.ac.id²⁾,
dandy@stikom-bali.ac.id³⁾,

Abstrak

Peranan pariwisata semakin penting karena memberikan kontribusi besar terhadap penerimaan decisa dan pendapatan daerah. Kepuasan tamu hotel menjadi salah satu faktor yang sangat penting yang dapat mempengaruhi keberhasilan dan keberlangsungan bisnis perhotelan dan pariwisata. Namun saat ini banyak hotel masih melakukan analisa terhadap komentar tamu yang masuk dengan cara manual, sehingga hasil analisa tersebut sering kali tidak akurat karena adanya pandangan atau perspektif yang subjektif. Sehingga dibutuhkan model analisis yang tepat untuk menganalisis komentar tamu hotel sebagai dasar pengambilan keputusan bagi pelaku bisnis perhotelan. Penelitian ini mengusulkan model sentimen analisis berdasarkan komentar tamu hotel menggunakan penggabungan metode pengolahan data teks yaitu Regex dan VADER. Tujuan penelitian adalah untuk mendapatkan analisis sentimen dari tamu hotel yang akurat, sehingga dapat menjadi dasar dalam pengambilan keputusan. Dalam penelitian ini komentar dari tamu di proses dalam dua tahap yaitu pre-process dan analisis sentimen. Pada tahap pre-processing dilakukan analisis teks cleaning, teks grouping dan tokenisasi. Hasil penelitian menunjukkan bahwa model mampu menganalisis komentar tamu yang bersifat positif dan negatif dari 597 komentar yang masuk. Hasil pemodelan dapat digunakan dalam pengambilan keputusan bagi pelaku bisnis perhotelan dalam perbaikan layanan yang tepat sasaran sehingga tingkat kepuasan tamu hotel akan semakin meningkat.

Kata kunci: RegEx, VADER, Machine Learning, Analisa Kepuasan, Tamu Hotel.

1. Pendahuluan

Peranan sektor pariwisata nasional semakin penting dilihat dari perkembangan dan kontribusi yang diberikan sektor pariwisata melalui penerimaan devisa, pendapatan daerah, pengembangan wilayah, maupun dalam penyerapan investasi dan tenaga kerja serta pengembangan usaha yang tersebar di berbagai pelosok wilayah di Indonesia. Sektor pariwisata mempunyai pengaruh terhadap perekonomian suatu negara melalui beberapa faktor. Pertama, sektor pariwisata sebagai penghasil devisa untuk memperoleh barang modal yang dipakai dalam proses produksi. Kedua, pengembangan pariwisata menstimulus pembangunan infrastruktur. Ketiga, pengembangan sektor pariwisata mendorong perkembangan sektor-sektor ekonomi lainnya. Keempat, pariwisata ikut berkontribusi terhadap peningkatan kesempatan kerja dan pendapatan. Kelima, pariwisata dapat menyebabkan *positive economies of scale* (Yakup, 2019) [1].

Bali termasuk destinasi wisata terpopuler di dunia. Pesona memikat terpancar dari keindahan pulau Bali. Banyak pesona yang memanjakan mata yang dapat dinikmati di Bali. Keindahan pantainya, hingga budaya di pulau Bali. Perkembangan Pariwisata Bali dari tahun ke tahun sangat dipengaruhi oleh faktor keragaman budaya yang dimiliki oleh masyarakat Bali. Adat, seni, dan budaya Bali sebagai potensi dasar yang dominan di dalamnya tersirat satu cita-cita akan adanya hubungan timbal balik antara pariwisata dan kebudayaan. Hal ini menjadi salah satu keunikan yang khas dibandingkan dengan destinasi-destinasi lainnya yang ada di Indonesia [2].

Hotel merupakan suatu organisasi yang bergerak dibidang jasa dengan fokus kegiatannya adalah untuk melayani tamu hotel melalui pelayanan yang berkualitas. Hotel diharuskan memiliki bentuk pelayanan yang terbaik untuk ditawarkan kepada tamu hotelnya agar tamu hotelnya memilih jasanya. Pemberian pelayanan yang terbaik tentu setiap hotel akan membuat adanya kreatifitas ataupun inovasi untuk memperluas pasar tamu hotelnya maka kompetisi sebagai penyedia pelayanan terbaik dan terjangkau akan terlihat dalam kemudahan memperoleh informasi [3]. Oleh karena itu kritik atau saran dari pelanggan atau tamu hotel sangat penting bagi pihak penyedia layanan hotel untuk meningkatkan layanan yang sesuai dengan keinginan para pelanggan atau tamu hotel sehingga tamu akan selalu kembali menginap di hotel dan keuntungan bagi pihak hotel akan semakin meningkat.

Dalam penelitian ini diusulkan model analisis komentar tamu yang diistilahkan dengan *sentiment analysis*. Tujuannya adalah untuk menganalisis dan mengukur kepuasan tamu hotel menggunakan pendekatan *machine learning* dengan algoritma RegEx dan VADER (*Valence Aware Dictionary and sEntiment Reasoner*). *Regular Expression* (RegEx) adalah alat yang sangat kuat yang digunakan dalam pemrosesan teks dan tugas pencocokan pola, sering disingkat sebagai RegEx atau RegExp [4]. Tools ini memberikan cara yang ringkas untuk mencari, mengekstrak, dan memanipulasi teks berdasarkan pola yang ditentukan. Dengan berperan sebagai bahasa mini dalam bahasa pemrograman yang lebih besar, mereka memungkinkan pengguna untuk mendefinisikan pola-pola karakter yang perlu mereka cocokkan dalam sebuah teks yang diberikan. Sebuah metode untuk menemukan rangkaian karakter tertentu yang tertanam dalam teks karakter dijelaskan, dan penerapannya dalam bentuk sebuah kompuler dibahas. Kompuler tersebut menerima ekspresi reguler sebagai bahasa sumber dan menghasilkan program IBM 7094 sebagai bahasa objek. Program objek tersebut kemudian menerima teks yang akan dicari sebagai input dan menghasilkan sinyal setiap kali sebuah rangkaian yang tertanam dalam teks cocok dengan ekspresi reguler yang diberikan. Contoh, masalah, dan solusi juga disajikan [4]. VADER (*Valence Aware Dictionary and sEntiment Reasoner*) adalah algoritma analisis sentimen yang dikembangkan khusus untuk mendeteksi polaritas sentimen (positif, negatif, atau netral) dari teks. Algoritma ini dirancang untuk menangani beberapa aspek khusus dalam analisis sentimen, termasuk pengenalan intensitas sentimen dan penanganan konteks sentimen dalam kalimat yang kompleks. VADER mencapai beberapa skor yang mencolok untuk berbagai domain seperti twit, ulasan film, atau ulasan produk [5]. Dalam penelitian ini VADER akan digunakan untuk menganalisa sentimen dari tamu hotel. Penggunaan dua pendekatan analisis teks yaitu RegEx dan VADER dapat dimanfaatkan bagi pelaku bisnis perhotelan di Bali atau pihak penyedia layanan hotel sebagai dasar melakukan perbaikan layanan berdasarkan komentar tamu yang dianalisis sehingga dapat menambah tingkat kepuasan para tamu hotel.

2. Metode Penelitian

Pada bagian ini akan membahas metodologi penelitian *machine learning* menggunakan algoritma RegEx dan VADER untuk menganalisa data komentar dari tamu hotel. Dalam penelitian ini menggunakan aplikasi Orange Data Mining versi 3.35 untuk mengolah data dan melakukan analisa.

Beberapa tahapan yang dilakukan dalam process ini dimulai dengan menyiapkan *dataset* kemudian dilanjutkan dengan pre-proses untuk melakukan pengecekan disetiap komentar atau data yang masuk, mulai dari pengecekan karakter spesial, *emoticon*, kesalahan ejaan atau kata gaul, dan lain lain untuk memastikan agar data diproses lebih lanjut oleh *machine learning*. Selanjutnya dilakukan proses RegEx yaitu untuk menghilangkan karakter yang tidak bisa diproses oleh *machine learning*. Proses terakhir dalam tahapan ini adalah analisa menggunakan VADER untuk mengetahui bobot dari setiap kalimat data komentar tamu, sehingga kita akan mengetahui apakah kalimat atau komentar tamu tersebut mengandung sentimen positif, negatif, atau netral.

2.1 Dataset (Data Komentar Tamu)

Langkah awal dari penelitian ini adalah menyiapkan *dataset* sebagai bahan analisa dalam *machine learning*. Dalam penelitian ini kami menggunakan *dataset* yang didapat dari komentar tamu dari media sosial dan *questionare*. Saat tamu *check-out* akan diundang untuk mengisi *questionare* tentang pengalaman mereka saat menginap dihotel, undangan *questionare* berupa *link*. Data komentar tamu juga diambil dari social media, yaitu halaman *fanpage* Instagram dan facebook.

Data komentar yang masuk berupa teks, dalam penelitian ini data yang terkumpul sebanyak 597 komentar, data diambil dalam kurun waktu mulai Oktober 2022 sampai dengan September 2023. Data tersebut bisa kita lihat dalam potongan data dalam tabel berikut:

Tabel 1. Data komentar tamu

Response Date	Overall Comment
29/09/2023 16:12	The room keys had errors 3 times, and every day I had to go back and forth for a total of 1 hour just to sort out the key issues. The sales team, especially "Ruly," was not respectful in marketing the membership product. It felt like they assumed I couldn't afford it. Even before I declined, I was treated dismissively with "It's okay, there's no pressure to take the membership, as long as I finish my presentation." I didn't receive a clear offer from the sales team.

29/09/2023 19:12	Overall, it's good
	Amazing hotels High standard Great service Friendly people Great atmosphere and amenities Great food options
29/09/2023 22:32	Amazing pool, gym and spa Overall best hotel ever stayed in
29/09/2023 23:07	Great experience staying here
29/09/2023 22:24	The hotel facility is new. Staff service was great and very friendly. Food taste great at the resturant. A minor improvement if there can be more selection of pasta and steak at the seafood resturant would be great. Overall was an excellent experience staying at this hotel.
30/09/2023 13:23	All good all new. Service in high level
30/09/2023 20:36	The room is nice and aesthetic, the breakfast variety is great, the taste is amazing, the service is very friendly, the pool is beautiful, and the lounge is spacious and cool.
26/09/2023 18:32	Very beautiful resort, amazing staff always with a smile and ready to serve. We had a very comfortable stay, amazing breakfast buffet with best quality and great tasting food A special thanks to your Chef Alit, he made special gluten free breakfast for my daughter every morning. His enthusiasm and welcoming smile made my daughter's holiday one of the best. He made sure my daughter's food requests are safely fulfilled, never said no to anything. Another thanks to your all Backstage Crew , they all are amazing amazing. THANK YOU FOR MAKING OUR STAY A WONDERFUL EXPERIENCE.

Detail tabel data diatas dapat dilihat berisi tanggal komentar masuk dan kolom kedua berisi isi komentar atau kritik yang ditulis oleh tamu.

2.2 Pre-Proses Data

Preprocess data dilakukan agar data dapat diproses dengan *machine learning* menggunakan VADER. Dalam proses ini setiap data komentar yang masuk dilakukan pengecekan dengan proses seperti berikut:

- Menghapus komentar spam atau komentar palsu yang biasanya masuk dari media sosial.
- Menghapus *emoticon* yang tidak dapat diproses oleh *machine learning*, contohnya 😊, 😞, dan lain lain.
- Menghapus karakter spesial yang tidak dapat diproses oleh *machine learning*, contohnya tanda !, tanda @, dan lain lain.
- Koreksi kesalahan ejaan dan kata slang atau kata gaul, contohnya *cheers*, *dude*, dan lain lain.
- Penghapusan *duplicate* data.

2.3 RegEx (Regular Expression)

Selanjutnya data diproses menggunakan algoritma RegEx untuk menghilangkan karakter spesial yang tidak dapat diproses oleh machine learning dari sebuah kata atau kalimat.

RegEx tidak terbatas pada bahasa pemrograman atau platform tertentu. RegEx juga didukung oleh berbagai bahasa pemrograman dan alat, menjadikannya alat yang serbaguna untuk tugas pemrosesan teks di berbagai *environments*.

Untuk lebih memahami signifikansi praktis dari RegEx, dapat kita lihat beberapa contoh berikut:

- Validasi Data.
RegEx dapat digunakan untuk memvalidasi input pengguna, seperti memastikan bahwa alamat email atau nomor telepon mengikuti format tertentu.
- Pencarian dan Manipulasi Teks.
RegExp dapat digunakan untuk mencari pola dalam dokumen teks, mengekstrak informasi tertentu, atau melakukan operasi substitusi teks.
- Analisis Log.
RegEx sering digunakan dalam analisis log untuk mengekstrak data relevan seperti *timestamp*, alamat IP, dan kode kesalahan dari file log.

Proses RegEx ini diperlukan agar teks dapat diolah dengan lebih efisien dan memudahkan dalam menganalisa teks.

2.4 VADER (*Valence Aware Dictionary and sEntiment Reasoner*)

Proses selanjutnya adalah menganalisa sentimen menggunakan VADER dengan cara menggabungkan metode kualitatif dan kuantitatif untuk menghasilkan kamus sentimen standar yang berkualitas tinggi. Sebuah kamus sentimen adalah daftar kemampuan leksikal (misalnya, kata-kata) yang umumnya dapat dikategorikan dengan orientasi semantik mereka sebagai positif, negatif, atau netral. Dari hasil analisa sentimen menggunakan VADER tersebut akan diketahui bobot dari tiap kata atau kalimat apakah mengandung sentimen positif, negatif, atau netral. Pada bagian ini akan membahas metodologi penelitian machine learning menggunakan algoritma RegEx dan VADER untuk menganalisa data komentar dari tamu hotel. Dalam penelitian ini menggunakan aplikasi Orange Data Mining versi 3.35 untuk mengolah data dan melakukan analisa.

Rumus perhitungan analisa menggunakan VADER dapat kita lihat seperti berikut:

$$\text{Skor Komposit/Compound} = \frac{\text{Jumlah Skor Valence untuk Setiap Token}}{\text{Faktor Normalisasi}}$$

Faktor normalisasi adalah jumlah dari nilai absolut skor valence untuk setiap token dalam teks. Skor valence memiliki nilai sebagai berikut:

- **Skor Positif:**

Skor antara 0 dan 1 yang mewakili proporsi teks yang bersifat positif.

- **Skor Negatif:**

Skor antara 0 dan -1 yang mewakili proporsi teks yang bersifat negatif.

- **Skor Netral:**

Skor antara -1 dan 1 yang mewakili proporsi teks yang bersifat netral.

- **Skor komposit**

Kemudian skor komposit digunakan untuk mengklasifikasikan sentimen keseluruhan dari teks:

$$\begin{aligned} \text{Positif} &= \text{Skor Komposit} \geq 0.05 \\ \text{Negatif} &= \text{Skor Komposit} \leq -0.05 \\ \text{Netral} &= -0.05 < \text{Skor Komposit} < 0.05 \end{aligned}$$

3. Hasil dan Pembahasan

Setelah melakukan beberapa tahapan yang dimulai dari pre-proses sampai dengan analisa sentimen menggunakan VADER, maka hasilnya dapat kita lihat apakah kalimat mengandung sentimen positif, negatif atau netral. Dan secara rinci dapat kita lihat dalam tabel berikut:

3.1 Hasil Analisa Vader

Hasil analisa menggunakan VADER dalam membobotkan setiap kata dalam kalimat yang hasilnya berupa nilai sentimen positif, negatif, dan netral. Secara rinci dapat kita lihat dalam tabel dibawah:

Tabel 2. Hasil Analisa Menggunakan Vader

No	Overall Comment	pos	neg	neu	compound
1	The room keys had errors 3 times, and every day I had to go back and forth for a total of 1 hour just to sort out the key issues. The sales team, especially "Ruly," was not respectful in marketing the membership product. It felt like they assumed I couldn't afford it. Even before I declined, I was treated dismissively with "It's okay, there's no pressure to take the membership, as long as I finish my presentation." I didn't receive a clear offer from the sales team.	0.051	0.134	0.815	-0.7239
2	Overall, it's good	0.592	0	0.408	0.4404
3	Amazing hotels High standard Great service Friendly people Great atmosphere and amenities Great food options Amazing pool, gym and spa Overall best hotel ever stayed in	0.59	0	0.41	0.9823
4	Great experience staying here	0.577	0	0.423	0.6249
5	The hotel facility is new. Staff service was great and very friendly. Food taste great at the resturant. A minor improvement if there can be more selection of pasta and steak at the seafood resturant would be great. Overall was an excellent experience staying at this hotel.	0.36	0	0.64	0.9735

Dalam tabel diatas dapat kita lihat hasil analisa menggunakan VADER, penjelasan secara lebih detail sebagai berikut:

- **Kolom No.** (Nomor):

Pada kolom ini berisi nomor urut data komentar tamu yang masuk.

- **Kolom Pos** (Sentimen Positif):

Pada kolom ini berisi nilai bobot untuk sentimen positif dari kata atau kalimat komentar tamu.

- **Kolom Neg** (Sentimen Negatif):

Pada kolom ini berisi nilai bobot untuk sentimen negatif dari kata atau kalimat komentar tamu.

- **Kolom Neu** (Sentimen Netral):

Pada kolom ini berisi nilai bobot untuk sentimen netral dari kata atau kalimat komentar tamu.

- **Kolom Compound** (Komposit):

Pada kolom ini berisi nilai bobot keseluruhan dari kata atau kalimat komentar tamu, sehingga kita dapat melihat nilai bobot dari keseluruhan kalimat apakah mengandung sentiment positif, negatif, atau netral.

4. Kesimpulan

Dalam penelitian ini, kami membahas pendekatan *machine learning* untuk menganalisa data komentar tamu dengan menggabungkan algoritma RegEx dan VADER sebagai solusi atas hasil analisa secara manual yang saat ini masih dilakukan oleh banyak hotel yang hasilnya tidak akurat karena adanya pandangan dan perspektif yang subjektif.

Penelitian kami diatas menunjukkan hasil yang akurat dalam menganalisa data dari komentar para tamu, dari jumlah total 597 komentar yang masuk sebagian besar berisi sentimen positif dan hanya sebagian kecil saja yang berisi sentimen negatif. dari hasil analisa tersebut bisa dipakai oleh pihak hotel sebagai pendukung dalam pengambilan keputusan untuk memperbaiki layanan hotel sehingga tingkat kepuasan tamu hotel akan semakin meningkat.

Kami juga berharap penelitian kami dapat berkontribusi dan digunakan sebagai pendukung dalam penelitian selanjutnya untuk mencari metode lain yang lebih sempurna untuk memajukan industri pariwisata.

Daftar Pustaka

- [1] A.A.A Ribeka Martha Purwahita, Putu Bagus Wisnu Wardhana, I Ketut Ardiasa, & I Made Winia. (2021). Dampak Covid-19 terhadap Pariwisata Bali Ditinjau dari Sektor Sosial, Ekonomi, dan Lingkungan (Sebuah Tinjauan Pustaka). *Jurnal Kajian Dan Terapan Pariwisata*, 1(2), 68–80. <https://doi.org/10.53356/diparojs.v1i2.29>.
- [2] Ida Bagus Gede Paramital, I. G. G. P. A. P. 2. (2020). NEW NORMAL BAGI PARIWISATA BALI DI MASA PANDEMI COVID 19. *EISSN 2614-5340*. <http://ejournal.ihdn.ac.id/index.php/PB/index>.
- [3] Susepti, A., Hamid, D., & Kusumawati, A. (2017). PENGARUH KUALITAS PELAYANAN TERHADAP KEPUASAN DAN LOYALITAS TAMU HOTEL (Studi tentang Persepsi Tamu Hotel Mahkota Plengkung Kabupaten Banyuwangi). In *Jurnal Administrasi Bisnis (JAB)/Vol* (Vol. 50, Issue 5).
- [4] Thompson, I. (1968). *Regular Expression Search Algorithm The Algorithm* (Vol. 11, Issue 6).
- [5] Kamoltep M., Wararat S., (2020). "Emotion Analysis and Classification of Movie Reviews Using Data Mining". *IEEE Xplore*, 978-1-7281-9792-0/20 <https://doi.org/10.1109/DATABIA50434.2020.9190363>.
- [6] Borg, A., & Boldt, M. (2020). Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Systems with Applications*, 162. <https://doi.org/10.1016/j.eswa.2020.113746>
- [7] Es-Sabery, F., Es-Sabery, I., Hair, A., Sainz-de-Abajo, B., & Garcia-Zapirain, B. (2022). Emotion processing by applying a fuzzy-based Vader lexicon and a parallel deep belief network over massive data. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2022.3200389>
- [8] Mardjo, A., & Choksuchat, C. (2022). HyVADR: Hybrid VADER-Random Forest and GWO for Bitcoin Tweet Sentiment Analysis. *IEEE Access*, 10, 101889–101897. <https://doi.org/10.1109/ACCESS.2022.3209662>
- [9] Shihab Elbagir & Jing Yang (2019). Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment : IMECS 2019 : 13-15 March, 2019.
- [10] Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *AAAI Conference*, <http://sentiment.net/>
- [11] Setiowati, Y., & Setyorini, F. (n.d.). Service Extraction and Sentiment Analysis to Indicate Hotel Service Quality in Yogyakarta based on User Opinion. *ISRITI*, 978-1-5386-7422-2/18/\$31.00 ©2018 IEEE.
- [12] Al Amin, Imran Hossain, Aysha Akther*, & Kazi Masudul Alam (2019). Bengali VADER: A Sentiment Analysis Approach Using Modified VADER. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.
- [13] Kastrati, Z., Imran, A. S., Kurti, A. (2020). "Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs". *IEEE Access*, 8, 106799–106810. <https://doi.org/10.1109/ACCESS.2020.3000739>.
- [14] Carvalho, I., Oliveira, H. G., Silva, C. (2023). "The Importance of Context for Sentiment Analysis in Dialogues". *IEEE Access*, 11, 86088–86103. <https://doi.org/10.1109/ACCESS.2023.3304633>.
- [15] Saad, S. E., Yang, J. (2019). "Twitter Sentiment Analysis Based on Ordinal Regression". *IEEE Access*, 7, 163677–163685. <https://doi.org/10.1109/ACCESS.2019.2952127>.
- [16] Karsten M. T., Matthias L., Patrick P., & Carsten G., (2019). "GerVADER - A German Adaptation of the VADER Sentiment Analysis Tool for Social Media Texts". *Creative Commons License Attribution 4.0 International (CC BY 4.0)*.
- [17] Made, N., Kristina, R., Perjalanan, I., Duta, D., Negeri, H., Sugriwa, G. B., Ratna, J., 51 Tatanan, N., & Bali, D. (n.d.). Pemulihan Ekonomi Pariwisata Bali di Era New Normal. <https://infocorona.baliprov.go.id/>
- [18] Badan Pusat Statistik Provinsi Bali. 2023. Tingkat Penghunian Kamar (TPK) Hotel Bintang Menurut Kelas di Provinsi Bali (Persen) 2023. <https://bali.bps.go.id/indicator/16/106/1/banyaknya-wisatawan-mancanegara-bulanan-ke-bali-menurut-pintu-masuk.html>