

Analisis Sentimen Review Pelanggan Hotel Menggunakan Metode K-Nearest Neighbor (K-NN) (Studi Kasus: Hotels.Com, Booking.Com, Agoda.Com)

Rahma Nimas Safitri¹⁾ Vivine Nurcahyawati²⁾ Julianto Lemantara³⁾

Program Studi/Jurusan Sistem Informasi
Universitas Dinamika

Jl. Raya Kedung Baruk 98 Surabaya, 60298

Email : 1)rahmanimas.s@gmail.com, 2)vivine@dinamika.ac.id, 3)julianto@dinamika.ac.id

Abstract: *With the increasing number of gadgets and other online media, it is possible for consumers to provide reviews of services in the form of comments and opinions. In Appgrooves only assess by rating, but sometimes the rating is not enough to show consumers' responses to the services they get. From these problems, sentiment analysis is needed to classify user reviews based on positive and negative sentiments. The method used in this research was K-Nearest Neighbor. The K value used in this study is 9. The selection of K value based on testing the accuracy value by testing the K value 3,5,6,7 and 9. From the tests that have been done, the K value with the highest accuracy result was K = 9 . The results of studies that have been done for Hotels.com the percentage of positive comments by 32%, for Booking.com the percentage of positive sentiment is 29% , for Agoda.com the percentage of positive comments is 86%. Based on the validity test using 10-Fold Cross Validation, the accuracy for Hotels.com is 94.55%, for Booking.com is 87.58%, and for Agoda.com is 98.83%.*

Keywords: Hotels.com, Booking.com, Agoda.com, Sentiment Analysis, K-Nearest Neighbor

Dilansir dari AppGrooves (Appgrooves, 2019) dalam *Best 10 Hotel Booking Apps*, didapatkan informasi Best 10 Hotel Booking Apps dengan tiga peringkat utama diduduki oleh Hotels.com, Booking.com, Agoda.com, disusul oleh HotelTonight, Trivago, Expedia, Priceline Travel, Choice Hotel, Tripadvisor, dan KAYAK. Dengan makin meningkatnya gadget dan media online lainnya memungkinkan konsumen untuk memberikan review terhadap pelayanan berupa komentar maupun opini. Pada Appgrooves hanya menilai dengan rating, akan tetapi terkadang rating tidak cukup menunjukkan tanggapan konsumen atas pelayanan yang mereka dapatkan.

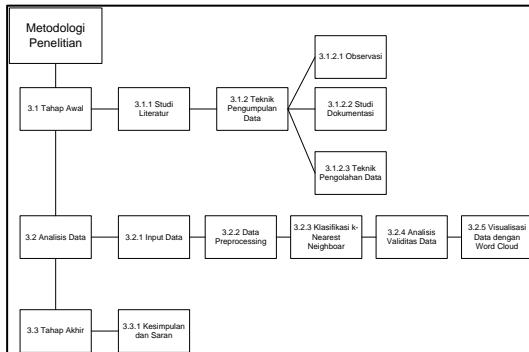
Dari permasalahan tersebut maka diperlukan analisis sentimen untuk mengklasifikasikan ulasan pengguna berdasarkan sentimen positif dan negatif. Dalam penelitian ini penulis menggunakan metode K-Nearest Neighbor untuk mengklasifikasikan ulasan negatif dan positif. Alasan penggunaan algoritma k-Nearest Neighbor digunakan dalam penelitian ini dikarenakan tingkat akurasi yang baik dan efektif jika digunakan pada data latih (training) yang berjumlah besar dan mengandung informasi yang kurang atau tidak berarti (noisy). Pengujian Validasi menggunakan

10-Fold Cross Validation untuk menentukan nilai akurasi dari data hasil penelitian yang telah dilakukan.

Terdapat beberapa penelitian terdahulu dalam hal klasifikasi sentimen analisis diantaranya Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor (Deviyanto & R. Wahyudi , 2018), Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel (Ernawati & Wati, 2018), dan Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor (Riany, Fajar, & Lukman, 2016).

METODE

Metodologi yang digunakan untuk melakukan analisis sentimen untuk pemeringkatan popularitas situs pemesanan hotel menggunakan metode k-nearest neighbor (k-NN) dibagi menjadi tiga tahapan yaitu tahap awal, analisis data, dan tahap akhir.



Gambar 1 Metodologi Penelitian

1. Pengumpulan Data

Data awal yang dibutuhkan pada penelitian ini adalah link dari situs pemesanan hotel online yang dijadikan sebagai objek. Selanjutnya, data link yang sudah ada diolah dalam aplikasi Data Miner untuk mengambil data-data pengguna yang ada pada situs Trustpilot.com

2. Data Preprocessing

Berdasarkan ketidak teraturan struktur data teks, maka proses sistem temu kembali informasi ataupun text mining memerlukan beberapa tahap awal yang pada intinya adalah mempersiapkan agar teks dapat diubah menjadi lebih terstruktur. Salah satu implementasi dari text mining adalah tahap text preprocessing. (Informatikologi, 2016). Tahap *Preprocessing* diantara:

a. Case Folding

Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu, peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil atau lowercase). Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.

b. Tokenizing

Tahap *Tokenizing* adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya.

Tokenisasi secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata, bagaimana membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan.

c. Filtering

Tahap *Filtering* adalah tahap mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist*

(membuang kata kurang penting) atau *wordlist* (menyimpan kata penting).

d. Stemming

Pembuatan indeks dilakukan karena suatu dokumen tidak dapat dikenali langsung oleh suatu Sistem Temu Kembali Informasi atau *Information Retrieval System (IRS)*. Oleh karena itu, dokumen tersebut terlebih dahulu perlu dipetakan ke dalam suatu representasi dengan menggunakan teks yang berada di dalamnya.

Proses *stemming* pada teks berbahasa Indonesia berbeda dengan stemming pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan sufiks. Sedangkan pada teks berbahasa Indonesia semua kata imbuhan baik itu sufiks dan prefiks juga dihilangkan.

3. K-Nearest Neighbor

Algoritma k-Nearest Neighbor adalah algoritma *supervised learning* dimana hasil dari instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori k-tetangga terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan obyek baru berdasarkan atribut dan *sample-sample* dari *training data*.

Algoritma k-Nearest Neighbor menggunakan *Neighborhood Classification* sebagai nilai prediksi dari nilai *instance* yang baru (Ismail, Cara Kerja Algoritma k-Nearest Neighbor (k-NN), 2018). Tahapan algoritma K-NN), diantaranya:

a. Menentukan parameter k (jumlah tetangga paling dekat).

b. Tentukan bobot untuk setiap *term* dengan menggunakan *Term Weighting TF-IDF*.

c. Hitung kemiripan antar dokumen dengan menggunakan *cosine similarity*

$$\text{Cos}(i,j) = \frac{\sum_k (d_{ik}d_{jk})}{\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}} \dots\dots\dots(1)$$

Gambar 2 Rumus Cosine Similarity

d. Urutkan hasil perhitungan *cosine similarity* dari besar ke kecil

e. Ambil sebanyak K yang paling tinggi kemiripannya dengan dokumen yang diklasifikasikan, tentukan kelasnya.

4. 10-Fold Cross Validation

Cross-validation (CV) adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi /

evaluasi. Model atau algoritma dilatih oleh subset pembelajaran dan divalidasi oleh subset validasi. Selanjutnya pemilihan jenis CV dapat didasarkan pada ukuran dataset. Biasanya CV K-fold digunakan karena dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi (Wibowo, 2017). Pemilihan nilai K dalam metode K-NN akan mempengaruhi hasil akurasi. Oleh sebab itu, perlu dilakukan uji coba nilai K 3, 5, 7, dan 9 untuk mengetahui nilai K yang memiliki tingkat akurasi tertinggi.

5. Visualisasi Data

Visualisasi data dengan *wordcloud* menghasilkan kata yang sering muncul dalam text. Penggunaan *font-size* huruf yang berbeda menggambarkan frekuensi kemunculan kata pada data yang dianalisis. Visualisasi data menggunakan Diagram Pie menggambarkan perbandingan persentase sentimen negatif dan positif dari masing-masing studi kasus yang dianalisis.

HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Data yang diambil adalah komentar pengguna yang memuat baik komentar positif maupun negatif dari pengguna dengan data yang didapatkan bersumber dari:

- 1. <https://www.trustpilot.com/review/www.hotels.com>
- 2. <https://www.trustpilot.com/review/www.booking.com>
- 3. <https://www.trustpilot.com/review/www.agoda.com>

Jumlah data yang diambil dari situs Trustpilot.com mengenai tiga situs diatas kemudian difilter dan diambil hanya yang menggunakan bahasa inggris dengan total data yang diambil sebanyak 26761 ulasan dengan rincian sebagai berikut:

- 1. Hotels.com : 1474 ulasan dari tahun 2011 sampai tahun 2019
- 2. Booking.com : 5418 ulasan dari tahun 2009 sampai tahun 2019
- 3. Agoda.com : 19869 ulasan dari tahun 2010 sampai tahun 2019

2. Data Preprocessing

Pada data preprocessing telah dilakukan proses *case folding*, *tokenizing*, *filtering*, dan *stemming*.

a. Case Folding

Proses *case folding* adalah mengubah data teks menjadi *lowercase* dengan tujuan agar kata yang sama namun berbeda secara penulisan huruf kapital dan tidak, tidak dianggap kata yang berbeda. Contoh data sebelum dan sesudah melalui tahap *lowercase* ditampilkan pada tabel berikut ini.

Tabel 1. *Case Folding*

Sebelum <i>Case Folding</i>	Sesudah <i>Case Folding</i>
HOTELS.COM --- They are horrible and hangs up on you! I booked online for SPAIN and when i arrived there they gave me a room that was not that i reserved for and they said there were no other rooms available ??? the hotels.com gave me a lower rate each night, then i get to hotel i get charge higher rate. SO i called, hotel.com and they say they the hotel can change their own rate??? really?? so then what is the advantage of hotels.com. Learn lesson, to make hotel reservations, book it directly with hotels. DO NOT EVER USE 3RD or 2ND party!! They are NEVER guarranteed!!	hotels.com --- they are horrible and hangs up on you! i booked online for spain and when i arrived there they gave me a room that was not that i reserved for and they said there were no other rooms available ??? the hotels.com gave me a lower rate each night, then i get to hotel i get charge higher rate. so i called, hotel.com and they say they the hotel can change their own rate??? really?? so then what is the advantage of hotels.com. learn lesson, to make hotel reservations, book it directly with hotels. do not ever use 3rd or 2nd party!! they are never guarranteed!!

b. Tokenizing

Pada proses *tokenizing* akan terjadi pemecahan kata. Contoh data sebelum dan sesudah melalui tahap *tokenizing* ditampilkan pada tabel berikut ini.

Tabel 2. *Tokenizing*

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
hotels.com --- they are horrible and hangs up on you! i booked online for spain and when i arrived there they gave me a	hotels com they are horrible and hangs up on you i booked online for spain and when i arrived there

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
room that was not that i reserved for and they said there were no other rooms available ??? the hotels.com gave me a lower rate each night, then i get to hotel i get charge higher rate. so i called, hotel.com and they say they the hotel can change their own rate??? really?? so then what is the advantage of hotels.com. learn lesson, to make hotel reservations, book it directly with hotels. do not ever use 3rd or 2nd party!! they are never guarranteed!!	they gave me a room that was not that i reserved for and they said there were no other rooms available the hotels com gave me a lower rate each night then i get to hotel i get charge higher rate so i called hotel com and they say they the hotel can change their own rate really so then what is the advantage of hotels com learn lesson to make hotel reservations book it directly with hotels do not ever use rd or nd party they are never guarranteed

c. Filtering

Filtering pada data komentar merupakan proses membersihkan komentar dari kata yang tidak diperlukan untuk mengurangi *noise*. Contoh data sebelum dan sesudah melalui tahap *tokenizing* ditampilkan pada tabel berikut ini.

Tabel 3. *Filtering*

Sebelum <i>Filtering</i>	Sesudah <i>Filtering</i>
hotels com they are horrible and hangs up on you i booked online for spain and when i arrived there they gave me a room that was not that i reserved for and they said there were no other rooms available the hotels com gave me a lower rate each night then i get to hotel i get charge higher rate so i called hotel com and	hotels com horrible hangs i booked online spain i arrived gave room i reserved said rooms available hotels com gave lower rate night i get hotel i get charge higher rate i called hotel com say hotel change rate advantage hotels com learn lesson make hotel reservations book hotels use rd nd party guarranteed

Sebelum <i>Filtering</i>	Sesudah <i>Filtering</i>
they say they the hotel can change their own rate really so then what is the advantage of hotels com learn lesson to make hotel reservations book it directly with hotels do not ever use rd or nd party they are never guarranteed	

d. Stemming

Proses *stemming* dilakukan untuk mendapatkan kata dasar dengan cara menghilangkan awalan, akhiran, dan sisipan. Pada penelitian ini kamus yang digunakan adalah Porter. Contoh data sebelum dan sesudah melalui tahap *tokenizing* ditampilkan pada tabel berikut ini.

Tabel 4. *Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
hotels com horrible hangs i booked online spain i arrived gave room i reserved said rooms available hotels com gave lower rate night i get hotel i get charge higher rate i called hotel com say hotel change rate advantage hotels com learn lesson make hotel reservations book hotels use rd nd party guarranteed	hotel com horribl hang i book onlin spain i arriv gave room i reserv said room avail hotel com gave lower rate night i get hotel i get charg higher rate i call hotel com sai hotel chang rate advantag hotel com learn lesson make hotel reserv book hotel us rd nd parti guarrante

3. K-Nearest Neighbor

Klasifikasi K-NN dalam penelitian ini terdiri dari 5 langkah dimana langkah pertama adalah menentukan parameter k (jumlah tetangga paling dekat). Pada penentuan nilai parameter k, dibutuhkan pengujian mulai dari k=3 sampai dengan k=10 dimana nilai k yang diambil adalah bilangan ganjil dengan sample uji adalah Hotels.com.

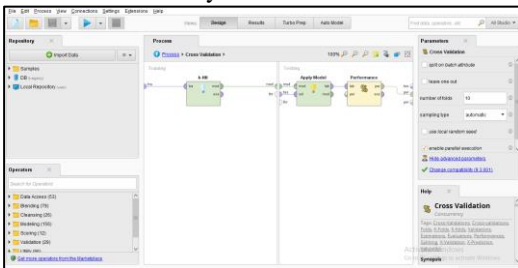
Tabel 1 Menentukan nilai K yang paling optimal

K	Accuracy
3	81,27%
5	83,85%

K	Accuracy
7	84,53%
9	84,93%

Berdasarkan hasil uji coba nilai K dan tingkat akurasi tertinggi maka pada penelitian ini nilai K yang diambil adalah 9. Tahap kedua adalah menentukan bobot untuk setiap *term* dengan menggunakan *Term Weighting* TF-IDF. Nilai bobot yang telah didapatkan digunakan untuk menghitung kemiripan antar dokumen dengan menggunakan *cosine similarity*.

Setelah didapatkan nilai *cosine similarity*, selanjutnya mengurutkan hasil perhitungan *cosine similarity* dari besar ke kecil (*Descending*), Lalu Ambil sebanyak K yang paling tinggi kemiripannya dengan dokumen yang diklasifikasikan. Langkah terakhir adalah menentukan kelasnya.



Gambar 3 Tampilan saat memasukkan metode K-NN

Bobot(W)=tf*idf						
D1	D2	D3	D4	D5	D6	
0,845098	0	0	0	0	0	0
0	0,845098	0	0	0	0	0
0	0	0,845098	0	0	0	0
0	0	0	0,544068	0	0,544068	
0	0	0	0	0,544068	0,544068	

Gambar 4 Menghitung nilai TF-IDF

TF-IDF dihitung dengan mengalikan nilai tf dan idf. Tf didapatkan dari banyaknya kata yang muncul pada satu dokumen. Sedangkan idf adalah banyaknya kata yang muncul pada keseluruhan dokumen.

WD6 * WDn				
WD6 * WD1	WD6 * WD2	WD6 * WD3	WD6 * WD4	WD6 * WD5
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0,2960100 37	
0	0	0	0	0,2960100 37

WD6 * WDn				
WD6 * WD1	WD6 * WD2	WD6 * WD3	WD6 * WD4	WD6 * WD5
0	0	0	0,2960100 37	0,2960100 37

Gambar 5 Menghitung nilai Cosine Similarity(1)

Pada gambar 5, nilai *similarity* didapatkan dari perhitungan bobot dari W6 dengan W *training*, dimana w6 adalah kata yang dijadikan *testing*. Selanjutnya, nilai bobot kata dikuadratkan sehingga menghasilkan nilai pada Gambar 4.6

D1	D2	D3	D4	D5	D6
0,714 19069 7	0,71 4190 7	0,714 19069 7	0,296 01003 7	0,296 01003 7	0,5 920 2

Gambar 6 Menghitung nilai Cosine Similarity(2)

Setelah didapatkan hasil kuadrat dari bobot kata, selanjutnya hasil kuadrat diakarkan. Kemudian, hasil perkalian pada *cosine similarity*(1) dibagi dengan nilai data uji pada *cosine similarity*(2) dikali dengan *data testing* pada *cosine similarity*(2). Hasil dari perhitungan manual seperti pada gambar 7 sesuai dengan perhitungan menggunakan *tools* rapidminer

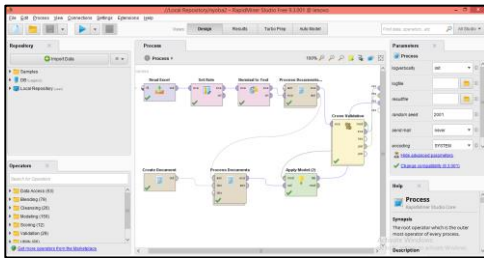
Hasil Cosine Similarity				
Cos(D6, D1)	Cos(D6, D2)	Cos(D6, D3)	Cos(D6, D4)	Cos(D6, D5)
0	0	0	0,70710 6781	0,70710 6781

Gambar 7 Hasil Cosine Similarity(2)

Dipilih nilai K untuk K-NN adalah 9 maka akan dipilih 9 nilai *similarity* yang tertinggi.

4. Cross Validation

Dari analisis validitas akan menghasilkan suatu nilai akurasi yang menggambarkan seberapa besar tingkat keakuratan analisis yang telah dilakukan.



Gambar 8. Tampilan *cross validation*

Berikut adalah hasil uji validitas data menggunakan *10-Fold Cross Validation* pada Hotels.com, Booking.com, dan Agoda.

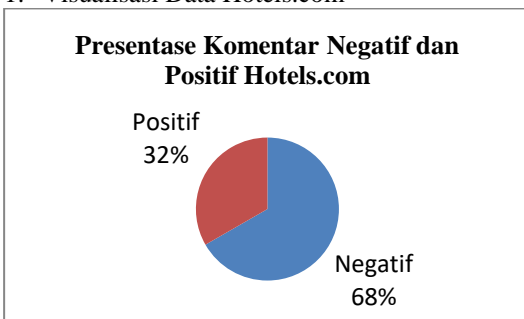
Tabel 5. Hasil *10-Fold Cross Validation*

Fold ke-	Akurasi		
	Hotels.com	Booking.com	Agoda
2	85,18%	89,93%	92,77%
3	84,77%	89,30%	92,81%
4	84,84%	89,37%	92,82%
5	84,23%	89,66%	92,89%
6	84,29%	89,63%	92,81%
7	84,09%	89,72%	92,85%
8	84,77%	89,52%	92,85%
9	84,36%	89,48%	92,85%
10	84,93%	89,68%	92,90%
Rata-rata	84,60%	89,58%	92,84%

5. Visualisasi Data

Tahap visualisasi data telah dilakukan dengan memvisualkan data dengan Worldcloud dan Diagram Pie. Visualisasi data dengan wordcloud menghasilkan kata yang sering muncul dalam teks yang telah dianalisis sebelumnya, penggunaan *font-size* huruf yang berbeda menggambarkan frekuensi kemunculan kata pada data yang dianalisis. Visualisasi data menggunakan Diagram Pie menggambarkan perbandingan persentase sentimen negatif dan positif dari masing-masing studi kasus yang telah dianalisis.

1. Visualisasi Data Hotels.com



Gambar 9 Persentase Komentar Positif dan Negatif Hotels.com

Pada gambar diatas, data yang digunakan dalam klasifikasi sebanyak 1472 ulasan. Dalam hal ini, presentase komentar negatif pada situs Hotels.com lebih banyak daripada komentar negatifnya. Untuk presentase komentar positif sebesar 32%, dan untuk presentase komentar negatif sebanyak 68%.

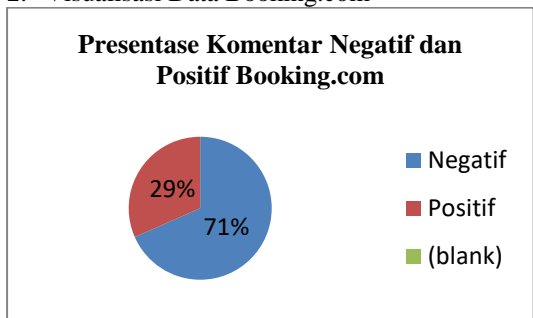


Gambar 10 Wordcloud Ulasan Positif Hotels.Com



Gambar 11 Wordcloud Ulasan Negatif Hotels.Com

2. Visualisasi Data Booking.com



Gambar 12 Persentase Komentar Positif dan Negatif Booking.com

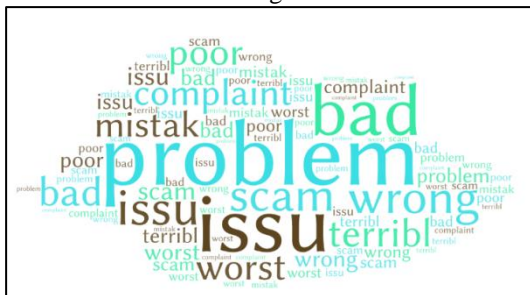
Pada gambar 5, data yang digunakan dalam klasifikasi sebanyak 5418 ulasan. Dalam hal ini, persentase komentar negatif pada situs Booking.com lebih banyak daripada komentar negatifnya. Untuk persentase komentar positif sebesar 29%, dan untuk presentase komentar negatif sebanyak 71%.



Gambar 13 Wordcloud Ulasan Positif Booking.Com

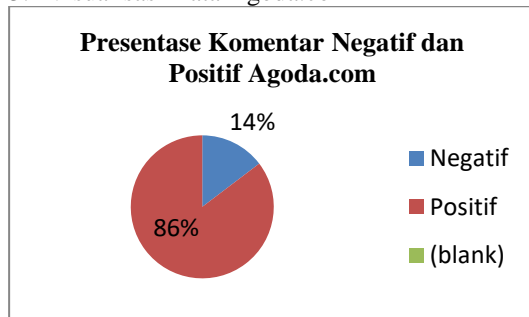


Gambar 17 Wordcloud Ulasan Negatif Agoda.Com



Gambar 14 Wordcloud Ulasan Negatif Booking.Com

3. Visualisasi Data Agoda.com



Gambar 15 Presentase Komentar Positif dan Negatif Agoda.com

Pada gambar diatas, data yang digunakan dalam klasifikasi sebanyak 19869 ulasan. Dalam hal ini, presentase komentar positif pada situs Agoda.com lebih banyak daripada komentar negatifnya. Untuk presentase komentar positif sebesar 86%, dan untuk presentase komentar negatif sebanyak 14%.



Gambar 16 Wordcloud Ulasan Positif Agoda.Com

KESIMPULAN

Penelitian ini telah berhasil melakukan analisis sentimen dengan menggunakan metode K-Nearest Neighbor pada tiga studi kasus yaitu Hotels.com, Booking.com, dan Agoda.com. Berdasarkan implementasi dan hasil evaluasi yang telah dilakukan dengan menggunakan nilai K=9 , maka dapat disimpulkan bahwa:

1. Hotels.com

Hotels.com memiliki dominasi ulasan negatif lebih banyak daripada ulasan positifnya. Perbandingan antara komentar positif dan negatif pada hotels.com sebesar 32% untuk sentimen positif, dan 68% untuk sentimen negatif. Beberapa hal yang sering dikeluhkan seperti kesulitan saat ingin reservasi melalui Hotels.com sering dikeluhkan, sehingga pada saat frekuensi kata dimunculkan kata *problem* muncul dominan.

2. Booking.com

Booking.com juga memiliki dominasi ulasan negatif lebih banyak daripada ulasan positifnya. Booking.com memiliki perbandingan antara komentar positif sebesar 29% , dan 71% untuk sentimen negatif. Beberapa hal yang sering dikeluhkan seperti kesulitan saat ingin reservasi melalui Booking.com sering dikeluhkan, sehingga pada saat frekuensi kata dimunculkan kata *problem* muncul dominan.

3. Agoda.com

Kemudahan saat melakukan reservasi dan *refund* sering disebut dalam ulasan. Agoda.com memiliki perbandingan antara sentimen positif sebesar 86%, dan 14% untuk sentimen negatif

RUJUKAN

Appgrooves. (2019, 9 26). *Best 10 Hotel Booking Apps*. Dipetik 9 26, 2019, dari Appgrooves: https://appgrooves.com/rank/travel_and_local/hotel-booking/best-hotel-booking-apps

- Deviyanto, A., & R. Wahyudi, M. D. (2018, 5).
PENERAPAN ANALISIS SENTIMEN
PADA PENGGUNA TWITTER. *JISKa*
(*Jurnal Informatika Sunan Kalijaga*),
3(1), 1-13.
- Ernawati, S., & Wati, R. (2018, 6 1). Penerapan
Algoritma K-Nearest Neighbors Pada
Analisis. *Jurnal Khatulistiwa*
Informatika, VI(1), 64-69.
- Ismail, A. M. (2018, 8 17). Cara Kerja
Algoritma k-Nearest Neighbor (k-NN).
Dipetik 9 17, 2019, dari Medium:
[https://medium.com/bee-solution-
partners/cara-kerja-algoritma-k-nearest-
neighbor-k-nn-389297de543e](https://medium.com/bee-solution-partners/cara-kerja-algoritma-k-nearest-neighbor-k-nn-389297de543e)
- Informatikalogi. (2016, 11 27). *Text*
Preprocessing. Diambil kembali dari
Informatikalogi.com:
[https://informatikalogi.com/text-
preprocessing/](https://informatikalogi.com/text-preprocessing/)
- Riany, J., Fajar, M., & Lukman, M. P. (2016).
Penerapan Deep Sentiment Analysis
pada Angket Penilaian. *Jurnal Sisfo*,
06(01), 147-156.
- Wibowo, A. (2017, 11 24). *10 Fold-Cross*
Validation. Diambil kembali dari Binus:
[https://mti.binus.ac.id/2017/11/24/10-
fold-cross-validation/](https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/)