

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/342589974>

Analisis Sentimen Berbasis Aspek Ulasan Pelanggan Terhadap Kertanegara Premium Guest House Menggunakan Support Vector Machine

Article · June 2020

CITATIONS

0

READS

185

3 authors, including:



Wirdhayanti Paulina

2 PUBLICATIONS 0 CITATIONS

[SEE PROFILE](#)



Alfi Nur Rusydi

Brawijaya University

19 PUBLICATIONS 9 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Mobile-based Sleepiness Detection [View project](#)



Food Labeling on Android using TensorFlow [View project](#)

Analisis Sentimen Berbasis Aspek Ulasan Pelanggan Terhadap Kertanegara Premium Guest House Menggunakan Support Vector Machine

Wirdhayanti Paulina¹, Fitra Abdurrachman Bachtiar², Alfi Nur Rusydi³

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹wirdhayantipaulina@gmail.com, ²fitra.bachtiar@ub.ac.id, ³alfi.nurruasydi@ub.ac.id

Abstrak

Kertanegara *Premium Guest House* adalah salah satu penginapan di Kota Malang yang terletak di Jl. Semeru No.59. Kertanegara sangat menyadari pentingnya keberadaan E-WOM (Electronic Word Of Mouth) demi kelangsungan bisnisnya. E-WOM dapat disebarkan melalui ulasan pelanggan. Kertanegara memiliki dua sumber ulasan pelanggan yaitu *Guest Review* dan ulasan *online* di situs OTA (*Online Travel Agent*). Saat ini, proses pengolahan ulasan pelanggan masih berfokus hanya pada *Guest Review*. Namun di sisi lain, Kertanegara juga memiliki ulasan dalam situs OTA yang perlu untuk diolah karena 90 persen proses *booking* berasal dari situs OTA. Salah satu cara yang dapat digunakan untuk menganalisis dan mengolah teks ulasan/*review* tersebut adalah analisis sentimen. Analisis sentimen dilakukan dalam tingkat aspek untuk menentukan layanan dan aspek yang memiliki polaritas negatif atau positif menggunakan metode *Support Vector Machine* (SVM) dan *Term Weighting* (TF-IDF). Data teks ulasan yang digunakan berbahasa Indonesia berasal dari situs Agoda.com, Expedia, Pegi-Pegi, Booking.Com, TripAdvisor dan berjangka waktu dari tahun 2012 sampai tahun 2019. Pengujian hasil klasifikasi menghasilkan rata-rata nilai *Accuracy* diatas 70%. Hasil analisis kemudian divisualisasikan melalui *dashboard* dengan menampilkan 6 komponen penting sehingga dapat membantu Kertanegara dalam pengambilan langkah strategis untuk membenahi, memperbaiki dan meningkatkan layanan yang memiliki polaritas negatif.

Kata kunci: *Guest House, Online Travel Agent, Analisis Sentimen, Klasifikasi, SVM, Dashboard*

Abstract

Kertanegara *Premium Guest House* is one of the inns in the city of Malang, located on Jl. Semeru No.59. Kertanegara is very aware of the importance of the existence of E-WOM (Electronic Word Of Mouth) for the continuity of its business. E-WOM can be disseminated through customer reviews. Kertanegara has two sources of customer reviews namely *Guest Reviews* and online reviews on the OTA (*Online Travel Agent*) site. At present, the process of processing customer reviews is still focused only on *Guest Review*. But on the other hand, Kertanegara also has a review on the OTA website that needs to be processed because 90 percent of the booking process comes from the OTA website. One method that can be used to analyze and process the review text is sentiment analysis. Sentiment analysis is carried out at the aspect level to determine services and aspects that have negative or positive polarity using the *Support Vector Machine* (SVM) and *Term Weighting* (TF-IDF) methods. The review text data used in the Indonesian language comes from the sites of Agoda.com, Expedia, Pegi-Pegi, Booking.Com, TripAdvisor and has a timeline from 2012 to 2019. Testing the classification results produce an average *Accuracy* value above 70%. The results of the analysis are then visualized through a dashboard by displaying 6 important components so that it can assist Kertanegara in taking strategic steps to fix, improve and improve services that have negative polarity.

Keywords: *Guest House, Online Travel Agent, Sentiment Analysis, Classification, SVM, Dashboard*

1. PENDAHULUAN

Saat ini, situs pemesanan perjalanan wisata atau yang biasa disebut *Online Travel Agent*

(OTA) bukan hanya berfungsi untuk reservasi akomodasi wisata tetapi juga mempunyai peran baru sebagai media *Electronic Word of Mouth* (E-WOM). Menurut Westbrook (1987) E-

WOM adalah semua komunikasi yang sifatnya informal yang ditujukan untuk konsumen melalui teknologi berbasis internet terkait penggunaan atau karakteristik dari suatu produk, layanan, atau penyediaanya. Melalui situs tersebut, pelanggan dapat menyampaikan pengalaman pasca penggunaan suatu produk atau jasa melalui teks ulasan atau *review*.

Kertanegara *Premium Guest House* merupakan salah satu penginapan di kota Malang yang berada di Jl. Semeru No.59 dan telah beroperasi sejak tahun 2012. Kertanegara telah menyadari pentingnya peran teknologi demi kelangsungan bisnisnya. Saat ini, Kertanegara telah mendaftarkan propertinya ke dalam beberapa situs OTA seperti *Tripadvisor*, *Booking.com*, *Pegi-Pegi*, dsb. Melalui penuturan Donny selaku *Sales & Marketing Manager*, 90 persen proses *booking* berasal dari situs OTA. Fakta tersebut patutnya menjadi pertimbangan bagi pihak manajemen dalam mengolah konten di situs OTA untuk membangun *image* Kertanegara yang baik sehingga mampu menghasilkan E-WOM yang positif. Sebab, E-WOM yang positif memiliki pengaruh yang signifikan terhadap konsumen karena dapat mendorong mereka untuk memilih akomodasi yang sesuai dengan apa yang mereka harapkan, meningkatkan keinginan konsumen untuk membeli dan merekomendasikannya ke kerabat mereka serta meningkatkan tingkat keputusan konsumen untuk menggunakan layanan tersebut (Ouyang 2017). E-WOM positif dapat dihasilkan melalui ulasan pelanggan yang baik, sementara itu ulasan yang baik dapat dihasilkan dari pengalaman pelanggan yang memuaskan terhadap layanan dan akomodasi yang disediakan.

Kertanegara mempunyai dua sumber untuk mendapatkan *feedback* pelanggan yaitu *guest review* (berbentuk kertas) dan ulasan online di situs OTA. Donny menyebutkan bahwa selama ini proses pengolahan ulasan pelanggan masih berfokus hanya pada *Guest Review*. *Guest Review* yang telah dikumpulkan akan dibahas langsung satu persatu secara verbal oleh pihak manajemen dan proses ini dilakukan rutin setiap bulan oleh Kertanegara. Pihak manajemen secara aktif mengulas hal-hal negatif yang diutarakan pelanggan dan mencari solusi yang mungkin dilakukan. Walaupun dilakukan secara aktif, proses ini dinilai tidak efektif karena jumlah *guest review* yang sedikit dan informasi yang tersedia juga terbatas. Disisi lain, ulasan

online di situs OTA belum dimanfaatkan secara maksimal oleh pihak Kertanegara. Hal ini disebabkan karena jumlah ulasan sangat banyak dan tersebar di berbagai situs OTA sehingga memiliki cakupan yang sangat luas serta butuh waktu 2-3 hari untuk mendapatkannya setelah pelanggan menulis ulasan lewat email. Donny berharap agar ulasan online dapat diolah untuk mengambil informasi tersembunyi yang dapat membantu proses evaluasi manajemen Kertanegara.

Salah satu cara yang dapat digunakan untuk menganalisis dan mengolah teks ulasan/*review* tersebut adalah analisis sentimen. Analisis sentimen dilakukan untuk mengelompokkan polaritas dari ulasan pelanggan untuk mengetahui opini pada kalimat tersebut termasuk positif atau negatif. Analisis sentimen dilakukan pada tingkat aspek untuk mengelompokkan hal-hal yang dikeluhkan oleh pelanggan ke dalam aspek-aspek tertentu kemudian divisualisasikan melalui *dashboard* sehingga informasi yang dibutuhkan oleh pihak manajemen dalam menentukan solusi yang tepat menjadi lebih rinci dan efektif.

Tsujii (2015) melakukan penelitian untuk mengambil informasi tersembunyi dari komentar *HotelClub.com*, *Agoda.com*, *Booking.com*, dan *Expedia.com* menggunakan teknik *text mining* untuk mengevaluasi ungkapan yang ada di dalam komentar yang berdasarkan label positif dan negatif. Selain sentimen juga terdapat beberapa kategori yang diuji dimana aspek lokasi, kamar, hotel, dan layanan menjadi topik yang paling banyak dibahas oleh para wisatawan. Pratama et. al., (2018) menggunakan metode TF-IDF sebagai *Term Weighting* dan *Support Vector Machine (SVM)* secara *multiclass* dengan teknik *one vs all* untuk analisis sentimen data opini pelanggan pada aspek pariwisata pantai malang. Rata-rata akurasi yang dihasilkan dari penggunaan metode ini sebesar 80%. Han Xiao Shi (2011) menggunakan SVM untuk klasifikasi *review online* tentang hotel dengan alasan SVM mempunyai efektifitas terbaik diantara algoritme *supervised machine learning* lainnya. Penelitian ini menghasilkan bahwa penerapan SVM dengan *Term Weighting* TF-IDF mempunyai nilai *Recall* sebesar 89,2%, *Precision* 85,2% dan *F-Score* 87,2%. Rodrigues (2017) juga menyebutkan dalam hasil penelitiannya bahwa diantara *classifier machine learning Support Vector*

Machine memiliki akurasi terbaik dibandingkan dengan *classifier* lainnya.

Berdasarkan paparan diatas, Analisis Sentimen akan dilakukan pada tingkat aspek terhadap ulasan pelanggan pada situs OTA Kertanegara *Premium Guest House* menggunakan algoritme *Support Vector Machine* dengan aspek berupa kamar, harga dan layanan. Hasil dari penelitian ini akan berupa klasifikasi opini pelanggan berdasarkan beberapa aspek tersebut dengan tujuan agar Kertanegara dapat membenahi layanan yang kurang memuaskan sehingga dapat mempertahankan kualitas dan menghasilkan e-WOM yang positif.

2. LANDASAN KEPUSTAKAAN

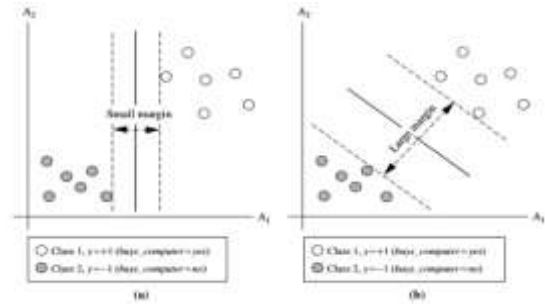
2.1 Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan payung dari cabang studi seperti *opinion extraction*, *sentiment mining*, *subjectivity analysis*, *affect analysis*, *emotion analysis*, *review mining*, dll. Analisis sentimen adalah suatu bidang studi yang menganalisis opini, sentimen pujian, emosi seseorang terhadap entitas seperti produk, layanan, organisasi, kejadian, masalah, serta atribut dari entitas (Liu 2012). Analisis sentimen terbagi menjadi 3 tingkat yaitu : *Document Level*, *Sentence Level*, dan *Entity Level*

2.2 Support Vector Machine

Algoritma SVM bertujuan untuk mencari *Maximum Marginal Hyperplane (MMH)* menggunakan *support vector* dan *margin*. MMH merupakan sebuah *hyperplane* terbaik dengan jarak margin terbesar yang digunakan untuk memisahkan data secara maksimal dan akurat terhadap setiap kelas.

Margin bisa di definisikan sebagai jarak terpendek dari sebuah *hyperplane* terhadap satu sisi dari margin itu sama dengan jarak *hyperplane* dengan sisi margin lainnya, dengan catatan kedua margin tersebut dalam posisi paralel dengan *hyperplane* (Han et al. 2012)



Gambar 1. Small Hyperplane vs Optimal Hyperplane

Jika terdapat sebuah *dataset* dalam bentuk $(X_1, y_1), (X_2, y_2), (X_3, y_3), \dots, (X_i, y_i)$ dimana X_i adalah *training tuple* dan y_i adalah label kelas dengan $i = 1 \dots N, X_i \in R^d$ dan $y_i \in \{-1, 1\}$. Setiap y_i dapat memilih salah satu dari dua nilai baik +1 atau -1 SVM akan membentuk *classifier* seperti persamaan (1) berikut :

$$f(x_i) = \begin{cases} \geq 0, y_i = +1 \\ < 0, y_i = -1 \end{cases} \quad (1)$$

Dalam SVM sebuah *hyperplane* akan digambarkan dalam persamaan berikut:

$$W \cdot X + b = 0 \quad (2)$$

Keterangan

- W = Bobot Scalar
- n = banyak atribut
- b = Nilai scalar atau disebut bias
- X = set data latih atau training tuples

2.3 TF-IDF

Term Weighting TI-IDF adalah salah satu pembobotan yang sering digunakan dan merupakan gabungan dari *Term Frequency* dan *Inverse Document Frequency*. *TF-IDF* terdiri dari frekuensi term dan inverse dokumen yang didapatkan dari membagi seluruh jumlah dokumen terhadap jumlah dokumen yang memiliki term tersebut. (Feldman and James 2007).

$$n_i(d) = df_i * \log D/df_i \quad (3)$$

Keterangan :

- $n_i(d)$ = bobot dari istilah t dalam dokumen d
- df_i = frekuensi munculnya istilah t dalam dokumen d
- $\log \frac{D}{df_i}$ = nilai *inverse document frequency* dari istilah t

2.4 Confusion Matrix

Confusion Matrix berisi informasi mengenai performa dari suatu sistem klasifikasi yang dievaluasi menggunakan data atau metrik yang

ada di dalam *Confusion matrix*. *Confusion Matrix* menganalisis seberapa baik klasifikasi yang telah dilakukan terhadap kelas aktual dan kelas hasil prediksi.

Tabel 1. *Confusion Matrix*

		Prediksi	
		Positif	Negatif
Aktual	Positif	TP	FN
	Negatif	FP	TN

Confusion matrix merepresentasikan tingkat akurasi dari proses klasifikasi yang telah dilakukan. Tingkat akurasi menunjukkan proporsi jumlah prediksi benar.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

Precision adalah proporsi dari pelabelan yang teridentifikasi dengan benar (*Completeness*), rumus mencari *Precision*:

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall adalah proporsi dari informasi yang dapat ditemukan dari label, rumus mencari *Recall*:

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

Precision dan *Recall* dapat digunakan untuk mendapatkan proporsi pengukuran lain yaitu *F1-Score*. *F1-Score* merupakan *harmonic mean* dari perhitungan *Precision* dan *Recall*, rumus mencari *F1-Score*:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (7)$$

2.5 Dashboard

Dashboard adalah sebuah *tools* yang secara diagnostik memberikan gambaran singkat tentang kinerja perusahaan untuk pihak *top management*. Terdapat fitur visual dan fungsional yang dapat membantu meningkatkan dan pengetahuan (Pauwels 2009). Penelitian ini menggunakan *Metabase* yang merupakan sebuah *tools Business Intelligence open source* untuk memvisualisasikan data melalui grafik yang telah dianalisis melalui *dashboard* yang interaktif.

3. METODE PENELITIAN



Gambar 2. Metodologi Penelitian

Tahap penelitian dimulai dengan mengidentifikasi masalah melalui proses wawancara dengan salah satu pihak Kertanegara. Selanjutnya dilakukan tahap studi literatur yang bertujuan untuk mengumpulkan teori yang dapat menunjang penelitian yang akan dilakukan untuk menyelesaikan permasalahan. Pengumpulan data dilakukan pada situs OTA (*Online Travel Agent*) menggunakan *tools webscraping* untuk mengekstrak data ulasan pelanggan yang akan digunakan pada proses Analisis Sentimen. Selanjutnya dilakukan perancangan data yang meliputi proses pengkategorian data dan pelabelan data secara manual serta perancangan *dashboard*. Hasil pelabelan data akan diberikan kembali pada pihak Kertanegara untuk validasi data.

Tahap berikutnya adalah *text preprocessing* yang bertujuan untuk mempersiapkan data sebelum masuk ke tahap *term weighting* dengan menggunakan modul NLTK dan Sastrawi pada Python. Beberapa tahapan yang ada dalam *text preprocessing* adalah *formalisasi* dan *translasi*, *cleansing*, *tokenizing*, *case folding*, *stemming*, dan *stopword removal*. Selanjutnya adalah tahap pembobotan kata atau *term weighting* yang akan menghasilkan data ulasan yang memiliki bobot kata dengan menggunakan modul *Sckit-learn* pada Python. Tahap klasifikasi sentimen membutuhkan dua data yaitu data ulasan yang telah memiliki label dan bobot kata. Tahap ini dibagi menjadi dua bagian, dimana tahap pertama bertujuan untuk memilih model terbaik yang akan digunakan pada tahap kedua yaitu klasifikasi sentimen terhadap setiap aspek. Pemilihan model dilakukan dengan menerapkan *Stratified KFold Cross Validation* dan klasifikasi dilakukan dengan algoritme *Support Vector*

Machine.

Tahap analisis dan pengujian hasil kalsifikasi dilakukan dengan menggunakan metrik *F1-Score*, *Precision*, dan *Recall* serta penghitungan *Accuracy*. Pengujian dilakukan dengan memanfaatkan modul Scikit-learn pada Bahasa Pemrograman *Python*. Tahap terakhir adalah menyajikan visualisasi melalui *dashboard* dari hasil analisis klasifikasi sentimen sesuai dengan rancangan *dashboard*.

4. IMPLEMENTASI

4.1 Pengumpulan Data

Pengumpulan dilakukan dengan metode *Web Scraping* melalui tools *webscraper.io*. Data yang akan dikumpulkan berupa ulasan pelanggan yang berasal dari beberapa situs OTA. Metode *Web Scraping* bertujuan untuk mengambil variabel dari data ulasan pada situs untuk dapat diolah dalam Analisis Sentimen. Variabel yang diambil antara lain adalah nama pembuat ulasan, waktu membuat ulasan, dan teks ulasan terhadap akomodasi yang akan dianalisis. Proses *Scraping* menghasilkan data sebanyak 1561 data teks ulasan yang terdiri dari 435 teks ulasan dari situs Agoda.com, 37 teks ulasan dari situs Expedia, 27 teks ulasan dari situs Pegi-Pegi, 235 teks ulasan dari situs Booking.Com, dan 827 teks ulasan dari situs TripAdvisor. Data teks ulasan berbahasa Indonesia dan berjangka waktu dari tahun 2012 sampai tahun 2019.

4.2 Perancangan Dataset

Pemilihan Aspek yang digunakan didasarkan pada hasil penelitian oleh Dolnicar (2003) untuk mencari atribut hotel yang paling penting dan di perhatikan oleh konsumen, lalu membuat urutan tingkatannya. Proses pemilihan aspek juga melibatkan pihak Kertanegara melalui proses wawancara sehingga didapatkan tiga aspek yang akan digunakan dalam Analisis Sentimen yaitu aspek harga, kamar dan layanan.

Setelah mendapatkan aspek yang akan digunakan maka dilakukan perancangan dataset dengan tujuan untuk manualisasi pelabelan data ulasan pelanggan terhadap aspek yang telah ditentukan dan sentimen dari setiap aspek. Label data akan dibagi menjadi 2 kelas berdasarkan sentimen positif dan negatif.

Tabel 2. Polaritas sentimen aspek

Nama Aspek	Positif	Negatif
Kamar	hotel dapat memenuhi hal-hal terkait kondisi kamar hotel	hotel tidak dapat memenuhi hal-hal terkait kondisi kamar hotel
Harga	hotel dapat menetapkan harga sesuai dengan nilai yang didapatkan	hotel tidak dapat menetapkan harga sesuai dengan nilai yang didapatkan
Layanan	hotel memenuhi ekspektasi pelanggan mengenai layanan yang dimiliki	hotel tidak memenuhi ekspektasi pelanggan mengenai layanan yang dimiliki

4.3 Text Preprocessing

Berikut adalah tahapan yang dilakukan dalam proses text preprocessing.

a. Formalisasi dan Translasi

Formalisasi merupakan tahap mengubah kata menjadi bentuk baku yang sesuai dengan KBBI. Translasi merupakan tahap untuk menerjemahkan kata dari bahasa asing ke dalam bahasa Indonesia

Tabel 3. Formalisasi dan Translasi

Isi Teks	Bentuk Baku KBBI
Tdk	tidak
Sdh	sudah
Thanks	Terima Kasih
Friendly	Ramah

b. Cleansing

Tahap *Cleansing* bertujuan untuk menghilangkan elemen elemen yang tidak diperlukan dalam proses analisis sentimen. Elemen tersebut terdiri tanda baca, angka, tag html.

Tabel 4. Cleansing

Isi Teks	Cleansing
semua hal penting mudah diakses. akomodasi yang sangat rapi, staf yang baik.	semua hal penting mudah diakses akomodasi yang sangat rapi staf yang baik
tempat menginap yang cukup nyaman ... harga baik ... pelayanan baik	tempat menginap yang cukup nyaman harga baik pelayanan baik

c. Tokenizing

Tahap *Tokenizing* bertujuan untuk memisahkan teks atau kalimat ulasan menjadi potongan kata.

Tabel 5. Tokenizing

Isi Teks	Tokenizing
----------	------------

semua hal penting mudah diakses akomodasi yang sangat rapi staf yang baik	['semua', 'hal', 'penting', 'mudah', 'diakses', 'akomodasi', 'yang', 'sangat', 'rapi', 'staf', 'yang', 'baik']
tempat menginap yang cukup nyaman harga baik pelayanan baik	['tempat', 'menginap', 'yang', 'cukup', 'nyaman', 'harga', 'baik', 'pelayanan', 'baik']

d. Case Folding

Tahap *Case Folding* bertujuan untuk mengubah kata yang dihasilkan pada proses tokenizing menjadi karakter huruf kecil.

Tabel 6. *Case Folding*

Isi Teks	Case Folding
['semua', 'hal', 'penting', 'mudah', 'diakses', 'akomodasi', 'yang', 'sangat', 'rapi', 'staf', 'yang', 'baik']	['semua', 'hal', 'penting', 'mudah', 'diakses', 'akomodasi', 'yang', 'sangat', 'rapi', 'staf', 'yang', 'baik']
['tempat', 'menginap', 'yang', 'cukup', 'nyaman', 'harga', 'baik', 'pelayanan', 'baik']	['tempat', 'menginap', 'yang', 'cukup', 'nyaman', 'harga', 'baik', 'pelayanan', 'baik']

e. Stemming

Tahap *Stemming* bertujuan untuk menghilangkan imbuhan kata pada teks ulasan pelanggan sehingga didapatkan kata dasar.

Tabel 7. *Stemming*

Isi Teks	Stemming
['semua', 'hal', 'penting', 'mudah', 'diakses', 'akomodasi', 'yang', 'sangat', 'rapi', 'staf', 'yang', 'baik']	['semua', 'hal', 'penting', 'mudah', 'akses', 'akomodasi', 'yang', 'sangat', 'rapi', 'staf', 'yang', 'baik']
['tempat', 'menginap', 'yang', 'cukup', 'nyaman', 'harga', 'baik', 'pelayanan', 'baik']	['tempat', 'inap', 'yang', 'cukup', 'nyaman', 'harga', 'baik', 'layan', 'baik']

f. Stopword Removal

Tahap *Stopword Removal* bertujuan untuk menghilangkan kata-kata yang tidak terlalu berpengaruh pada analisis sentimen atau kata yang terdapat di dalam *stopword list*.

Tabel 8. *Stopword Removal*

Isi Teks	Stopword Removal
['semua', 'hal', 'penting', 'mudah', 'diakses', 'akomodasi', 'yang', 'sangat', 'rapi', 'staf', 'yang', 'baik']	['semua', 'penting', 'mudah', 'akses', 'akomodasi', 'sangat', 'rapi', 'staf', 'baik']
['tempat', 'menginap', 'yang', 'cukup', 'nyaman', 'harga', 'baik', 'pelayanan', 'baik']	['tempat', 'inap', 'cukup', 'nyaman', 'harga', 'baik', 'layan', 'baik']

4.4 Term Weighting

Perhitungan bobot kata (*term weighting*)

dilakukan menggunakan TF-IDF. Hasil dari nilai TF-IDF akan menampilkan bobot kata yang meningkat sesuai dengan kemuculan kata dan perhitungan IDF menghasilkan bobot yang sesuai dengan tingkat keunikan dari kata tersebut dalam indeks kata. Berikut adalah manualisasi hasil perhitungan *Term Frequency* (TF), *Document Frequency* (DF) dan dilanjutkan dengan *Inverse Document Frequency* (IDF).

Tabel 9. Perhitungan TF, DF, IDF

Term	TF					DF	IDF
	D1	D2	D3	D4	D5		
kamar	1	2	5	0	2	4	0,0969
harga	1	0	1	0	0	2	0,3979
hotel	0	1	1	0	0	2	0,3979
staf	0	0	1	0	0	1	0,6989

Selanjutnya adalah contoh perhitungan bobot TF-IDF untuk kata 'kamar' pada D2:

$$Tf-idf_{t,d} = TF * IDF = 2 * 0,09691 = 0,1938$$

Tabel 10. Hasil TF-IDF

Term	TF				
	D1	D2	D3	D4	D5
kamar	0,09691	0,19382	0,48455	0	0,19382
harga	0,39794	0	0,39794	0	0
hotel	0	0,39794	0,39794	0	0
staf	0	0	0,69897	0	0

4.5 Klasifikasi Sentimen

Klasifikasi sentimen dibagi menjadi dua tahap. Tahap pertama adalah pemilihan model bertujuan untuk mendapatkan model terbaik yang akan digunakan untuk klasifikasi sentimen dalam rangka menghasilkan akurasi yang tinggi untuk setiap data. Pemilihan model dilakukan menggunakan *K-Fold Cross Validation*. Tujuan dari penggunaan *K-Fold Cross Validation* untuk menyusun dataset menjadi data *training* dan data *testing* dalam rangka untuk membatasi masalah *overfitting* data, adanya variansi tunggal atau tumpah tindih pada pembagian data training dan data testing sehingga model tidak kehilangan data yang signifikan untuk proses pemodelan dan pengujian. *Stratified K-fold* memastikan bahwa setiap fold yang dipilih memiliki persentase proporsi sampel yang sama untuk

setiap kelas.

Nilai k yang digunakan adalah 6 sehingga dataset akan dibagi menjadi 6 bagian dengan rasio data latih dan data uji sebesar 83 : 17. Kombinasi data latih dan data uji akan disimpan ke dalam 4 variabel yaitu X_{Train} , X_{Test} , y_{train} , y_{test} . Variabel X_{Train} berisi data bobot ulasan yang digunakan sebagai data latih. Variabel X_{Test} berisi data ulasan yang digunakan sebagai data uji. Variabel y_{train} berisi data label yang akan dilatih. Variabel y_{test} berisi data label yang akan diuji. Seluruh proses dalam tahap ini akan diterapkan pada data masing-masing aspek dengan tujuan mendapatkan model terbaik untuk proses klasifikasi sentimen.

Tahap kedua adalah klasifikasi sentimen menggunakan *Support Vector Machine* kernel Linear dengan classifier *LinearSVC*. Langkah awal klasifikasi sentimen dilakukan dengan fungsi *fit* untuk menyiapkan dan menyesuaikan model dengan memanfaatkan variabel X_{Train} dan y_{train} . Langkah selanjutnya, klasifikasi sentiment menggunakan fungsi *predict* untuk memprediksi label kelas dengan memanfaatkan variabel X_{Test} yang bertindak sebagai data yang akan diklasifikasi. Hasil dari proses ini berupa array yang berisi hasil klasifikasi sentiment dari variabel X_{Test} .

5. HASIL DAN PEMBAHASAN

5.1 Analisis Hasil Klasifikasi

Berikut adalah hasil klasifikasi sentimen terhadap 3 aspek yaitu :

a. Kamar

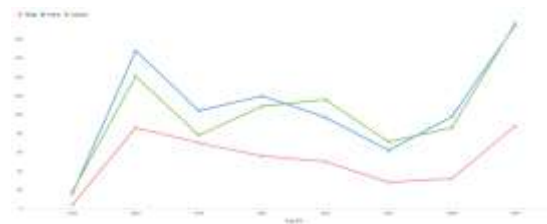
Klasifikasi terhadap ulasan aspek kamar menghasilkan 1070 ulasan dengan nilai 860 ulasan untuk kelas sentimen positif dan 210 ulasan untuk kelas sentimen negatif.

b. Harga

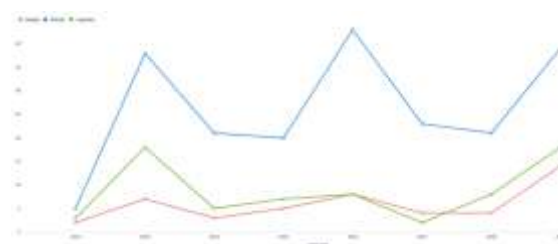
Klasifikasi terhadap ulasan aspek harga menghasilkan 461 ulasan dengan nilai 414 ulasan untuk kelas sentimen positif dan 47 ulasan untuk kelas sentimen negatif.

c. Layanan

Klasifikasi terhadap ulasan aspek layanan menghasilkan 885 ulasan dengan nilai 816 ulasan untuk kelas sentimen positif dan 69 ulasan untuk kelas sentimen negatif.



Gambar 3. Tren sentimen positif



Gambar 4. Tren sentimen negatif

Melalui gambar 3 dan 4 menunjukkan tren sentimen dari ketiga aspek mengalami pergerakan naik turun. Kelas sentimen positif dari ketiga aspek memiliki nilai rata-rata tertinggi pada tahun 2019 sedangkan. Kelas sentimen negatif dari ketiga aspek memiliki nilai rata-rata tertinggi pada tahun 2013 dan 2019. Hal ini menunjukkan bahwa Kertanegara terus melakukan perbaikan pada ketiga aspek tersebut namun perbaikan yang dilakukan belum tepat sasaran sehingga menurunkan kepuasan pelanggan.

Berdasarkan analisis hasil klasifikasi dan analisis tren dapat diketahui bahwa aspek kamar memiliki urgensi untuk segera ditangani oleh pihak Kertanegara. Hal ini dapat dibuktikan melalui hasil klasifikasi yang menunjukkan bahwa aspek kamar memiliki tingkat ulasan negatif yang paling besar selama kurun waktu 2012-2019. Analisis tren juga menunjukkan pada tahun 2019 terjadi kenaikan tingkat ulasan negatif terhadap ketiga aspek. Namun, aspek kamar memiliki tingkat kenaikan tertinggi diantara ketiganya.

5.2 Pengujian Klasifikasi

Pengujian hasil klasifikasi sentimen dilakukan menggunakan modul *Sckit-learn* pada Python dengan menggunakan 4 parameter pengujian yaitu *Accuracy*, *Precision*, *Recall* dan *F1-Score*.

Tabel 11. Pengujian Hasil Klasifikasi

Aspek	Sentimen	Accuracy	Precision	Recall	F1-Score
	Positif	0,79	0,82	0,75	0,78

Kamar	Negatif	0,79	0,77	0,83	0,80
	Positif	0,70	0,65	0,92	0,76
Harga	Negatif	0,70	0,86	0,50	0,63
	Positif	0,83	0,79	0,92	0,85
Layanan	Negatif	0,83	0,90	0,75	0,82

Tabel.11 menampilkan hasil pengujian hasil klasifikasi terhadap setiap aspek menggunakan metrik pengukuran yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Aspek layanan memiliki nilai akurasi tertinggi sebesar 0,83, sedangkan nilai akurasi terendah terdapat pada aspek harga sebesar 0,70. Aspek kamar memiliki nilai *precision* tertinggi pada sentimen positif sebesar 0,82, sedangkan nilai *precision* terendah pada sentimen positif terdapat pada aspek harga sebesar 0,65. Aspek layanan dan harga memiliki nilai *recall* tertinggi pada sentimen positif sebesar 0,92, sedangkan nilai *recall* terendah pada sentimen positif terdapat pada aspek kamar sebesar 0,75. Aspek layanan memiliki nilai *F1-Score* tertinggi pada sentimen positif sebesar 0,85, sedangkan nilai *F1-Score* terendah pada sentimen positif terdapat pada aspek harga sebesar 0,63.

5.3 Hasil Visualisasi Dashboard

Visualisasi dari hasil klasifikasi sentimen ditampilkan melalui *dashboard* dengan menggunakan *tool Metabase* yang menyajikan 6 komponen informasi diantaranya, informasi mengenai jumlah ulasan yang digunakan untuk analisis sentimen, *bar chart* yang menunjukkan jumlah ulasan pelanggan berdasarkan terhadap semua aspek yang telah ditentukan, *line chart* untuk menunjukkan tren jumlah ulasan pelanggan terhadap semua aspek yang telah ditentukan berdasarkan kelas sentimen pada kurun waktu tertentu, *pie chart* menunjukkan jumlah sentimen pada setiap aspek yang dipilih, yaitu *line chart* menunjukkan tren jumlah sentimen untuk setiap aspek yang dipilih pada kurun waktu tertentu, tabel yang menampilkan nama, tanggal, dan isi ulasan berdasarkan aspek dan sentimen yang dipilih.

Selanjutnya dilakukan pengujian Usability terhadap *dashboard* yang menghasilkan *SUS Score* sebesar 67,5 dimana *dashboard* memiliki tingkat *Acceptable* dengan rating *Good* yang berarti *dashboard* dapat diterima dengan baik oleh pihak Kertanegara Premium Guest House.



Gambar 5. Halaman *Dashboard* Analisis Sentimen

6. KESIMPULAN

1. Penggunaan metode *Support Vector Machine* yang digunakan dengan pembobotan TF-IDF dapat menjadi salah satu cara untuk menyelesaikan permasalahan klasifikasi dalam analisis sentimen. Hal tersebut dapat dibuktikan dengan penggunaan hasil pembobotan kata TF-IDF data ulasan dengan nilai label terhadap masing-masing aspek untuk proses klasifikasi sentimen menggunakan metode SVM.
2. Analisis hasil klasifikasi menunjukkan bahwa aspek kamar memiliki nilai ulasan negatif terbesar diantara kelima aspek sehingga aspek ini memiliki tingkat urgensi untuk segera ditangani oleh pihak Kertanegara.
3. Pengujian hasil klasifikasi sentimen memiliki rata-rata yang baik dengan nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* diatas 70%.
4. Hasil visualisasi *dashboard* dengan *tools Metabase* menghasilkan nilai 67,5 pada *Usability Testing* menggunakan metode *System Usability Scale (SUS)*. *Dashboard* memiliki tingkat *Acceptable* dengan rating *Good* yang berarti *dashboard* dapat diterima dengan baik oleh pihak Kertanegara Premium Guest House.

7. DAFTAR PUSTAKA

- Dolnicar, Sara. 2003. "Which Hotel attributes Matter ? A review of previous and a framework for future research." *Proceedings of the 9th Annual Conference of the Asia Pacific Tourism Association (APTA)*. 176-188.
- Feldman, Ronen, dan Sanger James. 2007. *The*

Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data. New York: Cambridge University Press.

Han, J., Kamber, M., & Pei, J., 2012. *Data Mining: concepts and techniques.* Amsterdam : Elsevier/Morgan

Han Xiao Shi, Xiao Jun Li. 2011. "A Sentiment Analysis Model for Hotel Reviews Based on Supervised Learning." *Proceedings of the 2011 International Conference on Machine Learning and Cybernetics, Guilin, 10-13 July, 2011* 10-13.

Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publishers.

Ouyang, Yu. 2017. "A Study of the Tourism Consumers ' Acceptance of Electronic Word- of- Mouth – A Case of Travel Accommodation." *Proceedings of the 2017 IEEE International Conference on Information, Communication and Engineering IEEE-ICICE 2017 - Lam, Meen & Prior* 258-261.

Pratama, Y., Bachtiar, F., & Setiawan, N. 2018. "Analisis Sentimen Opini Pelanggan Terhadap Aspek Pariwisata Pantai Malang Selatan Menggunakan TF-IDF dan Support Vector Machine." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya* 2 (12): 6244-6252.

Rodrigues, Anisha P. 2017. "Comparative Study of Machine Learning Techniques in Sentimental Analysis." *International Conference on Inventive Communication and Computational Technologies (ICICCT 2017)* 216-221.

Tsujii, Koichi. 2015. "Towards Extracting the Hotel Evaluations from the Comments by the Foreign Tourists with Text mining." *2015 IIAI 4th International Congress on Advanced Applied Informatics* 46-49.