

INTISARI

ANALISIS SENTIMEN BERDASARKAN ASPEK PADA ULASAN HOTEL BERBAHASA INDONESIA

Oleh:

Maulisa Dewi Mahenda

16/398521/PA/17482

Ulasan dapat digunakan sebagai rekomendasi sebelum konsumen menentukan hotel yang akan dipilih. Pemesan diminta untuk memberikan penilaian per aspek dalam bentuk bintang yang disediakan oleh TripAdvisor.com dan menyebabkan timbulnya bias diantara ulasan yang sudah ditulis dan penilaian per aspek. Oleh karena itu, dibutuhkan analisis sentimen berdasarkan aspek dari ulasan yang sudah dituliskan oleh konsumen. Selain itu, dibutuhkan pula klastering *aspect term* untuk menentukan kata yang termasuk pada suatu aspek. Aspek yang digunakan pada penelitian ini adalah aspek lokasi, layanan, *value*, dan kebersihan.

Proses klastering *aspect term* menggunakan *Fuzzy C-Means* dan proses klasifikasi sentimen menggunakan algoritma *Bernoulli Naive Bayes* dan *Multinomial Naive Bayes*. Data yang digunakan merupakan data yang berasal dari TripAdvisor.com dan menggunakan 129,256 ulasan dari 580 hotel.

Proses klastering *aspect term* menggunakan *Fuzzy C-Means* dan menghasilkan nilai koefisien *fuzzy partition* tertinggi sebesar 0.81. Proses klasifikasi sentimen menggunakan representasi fitur *CountVectorizer* dan *TfidfVectorizer*. Berdasarkan hasil pengujian klasifikasi sentimen setiap klaster aspek, rata-rata akurasi tertinggi sebesar 74.48% dan rata-rata *f-measure* tertinggi sebesar 84.58% dihasilkan oleh algoritma *Multinomial Naive Bayes* menggunakan *TfidfVectorizer*. Hasil akurasi yang diperoleh dari perhitungan *rating* sebesar 79.78%

Kata Kunci: *Fuzzy C-Means clustering*, analisis sentimen berdasarkan aspek, *naive bayes*

ABSTRACT

ASPECT BASED SENTIMENT ANALYSIS IN INDONESIAN HOTEL REVIEW

By:

Maulisa Dewi Mahenda

16/398521/PA/17482

Reviews can be used as recommendations before choosing hotel. The customer is asked to provide a per-aspect *rating* in the form of a star provided by TripAdvisor.com and causes a bias between the reviews that have been written and the *ratings* per aspect. Therefore, aspect based sentiment analysis is needed from the reviews written by customer. In addition, it also requires clustering of aspect *terms* to determine the words included in which aspect. Aspect that used in this experiment is location, service, value, and cleanliness.

The clustering of aspect *terms* process uses *Fuzzy C-Means* and the sentiment classification process uses *Bernoulli Naive Bayes* algorithm and *Multinomial Naive Bayes* algorithm. The data used is data from TripAdvisor.com and uses reviews from 580 hotels that contains 129,256 reviews.

The clustering aspects *terms* process uses *Fuzzy C-Means* and produces the highest fuzzy partition coefficient value of 0.81. The sentiment classification process uses the *CountVectorizer* and *TfidfVectorizer* feature representations. Based on the results of the sentiment classification test for each aspect cluster, the highest average accuracy of 74.48% and highest average f-measure of 84.58% is generated by the *Multinomial Naive Bayes* algorithm using a *TfidfVectorizer*. Accuracy result obtained from the calculation of rating prediction is 79.78%

Keyword: *Fuzzy C-Means clustering*, aspect based sentiment analysis, *naive bayes*