

**ANALISIS SENTIMEN PADA ULASAN WISATAWAN OBJEK WISATA BUKIT
PELALANGAN AROSBAYA DI *GOOGLE MAPS* MENGGUNAKAN ALGORITMA
*NAIVE BAYES***

Endah Widya Palupi

(190441100089)

Pembimbing 1 : Dr. Wahyudi Setiawan, S.Kom, M.Kom

Pembimbing 2 : Yudha Dwi Putra Negara, S.Kom.,M.Kom

ABSTRAK

Layanan bagi wisatawan dapat ditingkatkan dan dikembangkan secara berkelanjutan oleh petugas atau penyelenggara tempat wisata. Untuk mengetahui opini positif dan negatif pada ulasan tersebut maka diperlukan analisis sentimen. *Dataset* diambil dengan menggunakan *tools* webharvy. Data diambil pada tanggal 6 April 2023 dan berhasil diambil sebanyak 1470 data berupa ulasan. Data yang digunakan merupakan data berbahasa Indonesia. Kemudian data tersebut dilakukan pelabelan oleh pakar yaitu dosen Bahasa Indonesia. *Preprocessing* yang digunakan pada penelitian ini yaitu *data cleansing*, *case folding*, konversi *slangword*, *tokenizing*, *filtering*, dan *stemming*. Tujuan penelitian adalah untuk mengetahui metode yang lebih efektif dengan membandingkan teknik pembobotan kata *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Term Frequency-Relevance Frequency* (TF-RF) yang dikombinasikan dengan teknik klasifikasi *Naive Bayes Classifier* dengan menggunakan teknik validasi silang yaitu *k-cross validation* dengan $k=10$. *Dataset* pada penelitian ini menghasilkan sentimen positif lebih banyak dibandingkan dengan sentimen negatifnya. Hasil rata-rata akurasi *naïve bayes* dengan pembobotan TF-IDF paling besar pada komposisi *rate* 80:20 yaitu 87.94% dengan rata-rata nilai presisi 77.62%, *recall* 87.94% , *F1-score* 82.38% sedangkan dengan pembobotan TF-RF paling besar pada komposisi *rate* 70:30 yaitu 85.95% dengan rata-rata nilai presisi 74.39%, *recall* 85.95% , *F1-score* 79.62%. Nilai akurasi merupakan seberapa akurat model memprediksi dengan benar, nilai presisi menentukan seberapa baik model dalam mengidentifikasi kelas positif dengan benar, nilai *recall* menentukan seberapa baik model menemukan semua kelas negatif yang benar, dan nilai *F1-score* untuk mengukur seberapa baik model klasifikasi dapat mengklasifikasikan data dengan benar, dengan memperhatikan baik presisi maupun *recall*.

Kata kunci: Analisis Sentimen, Bukit Pelalangan Arosbaya, *Google Maps*, *Naïve Bayes*, TF-IDF, TF-RF, Wisata

SENTIMENT ANALYSIS OF TOURIST REVIEWS OF BUKIT PELALANGAN AROSBAYA TOURISM AT GOOGLE MAPS USING NAIVE BAYES ALGORITHM

Endah Widya Palupi

(190441100089)

Thesis Adviser 1 : Dr. Wahyudi Setiawan, S.Kom, M.Kom

Thesis Adviser 2 : Yudha Dwi Putra Negara, S.Kom.,M.Kom

ABSTRACT

Services for tourists can be enhanced and developed sustainably by officers or organizers of tourist attractions. To determine positive and negative opinions in these reviews, sentiment analysis is required. The dataset was obtained using web scraping tools called WebHarvy. The data was collected on April 6, 2023, and a total of 1470 reviews were successfully gathered. The data used in this study is in the Indonesian language. Subsequently, the data underwent labeling by an expert, who is an Indonesian language lecturer. The preprocessing techniques employed in this research include data cleansing, case folding, slang word conversion, tokenizing, filtering, and stemming.

The aim of this study is to determine the more effective method by comparing the term weighting techniques of Term Frequency-Inverse Document Frequency (TF-IDF) and Term Frequency-Relevance Frequency (TF-RF) combined with the Naive Bayes Classifier classification technique, using k-cross validation with k=10. The dataset used in this study yielded a greater number of positive sentiments compared to negative sentiments. The average accuracy of Naive Bayes with TF-IDF weighting was highest at an 80:20 rate composition, which reached 87.94%, with an average precision of 77.62%, recall of 87.94%, and F1-score of 82.38%. Meanwhile, with TF-RF weighting, the highest accuracy was achieved at a 70:30 rate composition, which reached 85.95%, with an average precision of 74.39%, recall of 85.95%, and F1-score of 79.62%. Accuracy represents how accurately the model predicts correctly, precision determines how well the model identifies the positive class correctly, recall determines how well the model finds all the true negative classes, and the F1-score measures how well the classification model can correctly classify the data, considering both precision and recall.

Keywords: *Bukit Pelalangan Arosbaya, Google Maps, Naïve Bayes, Sentiment Analysis, TF-IDF, TF-RF, Tourism*