

ABSTRAK

Dalam era internet dan media baru yang berkembang pesat, berbagai jenis *platform* telah menjadi sarana utama untuk menyebarkan informasi, termasuk informasi pariwisata. Salah satu alat penting dalam industri pariwisata adalah aplikasi *Google Maps*. Dalam era *Big Data*, *Google Maps* memainkan peran kunci dalam mengumpulkan jejak digital para wisatawan yang mengunjungi suatu lokasi. Jumlah ulasan dan pendapat yang diberikan oleh pengunjung pada aplikasi ini menjadi faktor krusial karena calon wisatawan cenderung mencari informasi terlebih dahulu di internet. Lamongan merupakan kabupaten di Jawa Timur yang memiliki banyak destinasi wisata, termasuk Wisata Bahari Lamongan(WBL) dan Goa Maharani & Zoo. Banyaknya pengunjung yang datang, tidak sedikit yang memberikan opini salah satunya pada aplikasi *goggle maps*. Penelitian ini dilakukan untuk menganalisis sentimen ulasan wisata pada *google maps* dengan menggunakan metode *Naïve Bayes Multinomial*. Kelebihan metode *Naïve Bayes* adalah dapat mengelolah data dengan jumlah yang besar, serta lebih kecil untuk terjadinya *overfitting*. Namun,metode *Naïve Bayes* memiliki kelemahan ketika bekerja dengan jumlah fitur yang banyak. Pengambilan data dilakukan dengan teknik *scraping* menggunakan *Instant data scrapper*. Data yang digunakan berjumlah 380 ulasan lalu dilakukan *ramdom oversampling* menjadi 207 data. Selanjutnya dilakukan beberapa *text preprocessing*, yang meliputi, *cleaning*, *tokenizing*, *casefolding*, *Stopword removal*, *stemming*, *normalization*, serta pembobotan kata menggunakan TF-IDF. Hasil terbaik dari pengujian model klasifikasi *Naïve bayest* yaitu dengan menggunakan *5-fold stratified cross validation*, yaitu mencapai *akurasi* sebesar 91,8% *AUC* sebesar 0,940, *precision* sebesar 92,9% *recall* sebesar 91,8% *F1-score* sebesar 0,916 *MCC* sebesar 0,884. Klasifikasi Naive Bayes yang dihasilkan terbukti cukup akurat dalam mengklasifikasikan berita pariwisata ke dalam Tiga kategori, yaitu *positif*, *negatif* dan *netral*.

Kata kunci: Pariwisata Lamongan, Analisis sentimen, *Naïve bayes*, *ramdom oversampling*

ABSTRACT

In the era of the internet and rapidly developing new media, various types of platforms have become the main means of disseminating information, including tourism information. One of the important tools in the tourism industry is the Google Maps application. In the era of Big Data, Google Maps plays a key role in collecting digital footprints of tourists visiting a location. The number of reviews and opinions given by visitors on this application is a crucial factor because prospective tourists tend to search for information on the internet first. Lamongan is a district in East Java that has many tourist destinations, including Wisata Bahari Lamongan (WBL) and Goa Maharani & Zoo. The large number of visitors who come, not a few of whom provide opinions, one of which is on the Google Maps application. This study was conducted to analyze the sentiment of tourist reviews on Google Maps using the Naïve Bayes Multinomial method. The advantage of the Naïve Bayes method is that it can process large amounts of data, and is less likely to overfit. However, the Naïve Bayes method has weaknesses when working with a large number of features. Data retrieval is carried out using the scraping technique using Instant data scrapper. The data used amounted to 380 reviews, then random oversampling was carried out to 207 data. Furthermore, several text preprocessing were carried out, including cleaning, tokenizing, casefolding, Stopword removal, stemming, normalization, and word weighting using TF-IDF. The best results from testing the Naïve Bayes classification model were using 5-fold stratified cross validation, which achieved an accuracy of 91.8% AUC of 0.940, precision of 92.9% recall of 91.8% F1-score of 0.916 MCC of 0.884. The resulting Naive Bayes classification proved to be quite accurate in classifying tourism news into three categories, namely positive, negative and neutral.

Keywords: *Lamongan Tourism, Sentiment analysis, Naïve Bayes, random oversampling.*