

SENTIMENT ANALYSIS OF BALI'S RAPID TOURISM GROWTH: COMPARING INDOBERT PERFORMANCE WITH INSET LEXICON AND MANUAL LABEL INTEGRATION

By

Ni Ketut Artini Artalia, NIM 2215091072

**Department of Informatics Engineering
Information Systems Program**

ABSTRACT

The rapid growth of tourism in Bali has sparked diverse public opinions, particularly on TikTok, where comment sections often reflect intense debates that may influence Bali's image as a sustainable tourism destination. To better understand this phenomenon, this study conducts sentiment analysis on public opinions on TikTok using Bidirectional Encoder Representations from Transformers, specifically IndoBERT, which is designed for Indonesian-language data. The data labeling process is carried out using two approaches: manual labeling by experts and automatic lexicon-based labeling using the InSet Lexicon, aiming to evaluate the impact of labeling quality on model performance as well as time and resource efficiency. Four experimental scenarios are implemented: (1) manual labeling with IndoBERT, (2) InSet Lexicon labeling with IndoBERT, (3) manual labeling with Random Over Sampling (ROS) integrated with IndoBERT, and (4) InSet Lexicon labeling with ROS integrated with IndoBERT. The results indicate that the fourth scenario achieves the best performance with an accuracy of 75.06%, which further improves to 81.83% after a proof-of-concept experiment involving the removal of slang words. Additionally, the sentiment analysis results show that public opinion tends to be neutral. These findings suggest that lexicon-based automatic labeling using InSet Lexicon, combined with data balancing techniques, is effective in improving model performance while also enhancing annotation efficiency.

Keywords: Sentiment Analysis, IndoBERT, InSet Lexicon, Random Over Sampling, Bali, TikTok.

ANALISIS SENTIMEN TERHADAP PERTUMBUHAN PARIWISATA PESAT DI BALI: PERBANDINGAN KINERJA INDOBERT DENGAN INTEGRASI KAMUS INSET DAN LABEL MANUAL

Oleh

Ni Ketut Artini Artalia, NIM 2215091072

Jurusan Teknik Informatika

Program Studi Sistem Informasi

ABSTRAK

Pertumbuhan pesat pariwisata di Bali memicu beragam opini publik, khususnya di TikTok, yang sering kali menampilkan perdebatan intens dan berpotensi memengaruhi citra Bali sebagai destinasi pariwisata berkelanjutan. Untuk memahami fenomena ini, penelitian ini melakukan analisis sentimen terhadap opini publik di TikTok dengan menggunakan metode Bidirectional Encoder Representations from Transformers, khususnya IndoBERT yang dirancang untuk data berbahasa Indonesia. Proses pelabelan data dilakukan melalui dua pendekatan, yaitu pelabelan manual oleh ahli dan pelabelan otomatis berbasis leksikon menggunakan InSet Lexicon, guna mengevaluasi pengaruh kualitas label terhadap performa model serta efisiensi waktu dan sumber daya. Penelitian ini menerapkan empat skenario pengujian, yaitu: (1) pelabelan manual dengan IndoBERT, (2) pelabelan InSet Lexicon dengan IndoBERT, (3) pelabelan manual dengan Random Over Sampling (ROS) dan IndoBERT, serta (4) pelabelan InSet Lexicon dengan ROS dan IndoBERT. Hasil penelitian menunjukkan bahwa skenario keempat menghasilkan performa terbaik dengan akurasi sebesar 75,06%, dan meningkat menjadi 81,83% setelah dilakukan proof of concept dengan menghapus kata slang. Selain itu, hasil analisis sentimen menunjukkan bahwa opini masyarakat cenderung bersifat netral. Temuan ini mengindikasikan bahwa pelabelan otomatis berbasis InSet Lexicon yang dikombinasikan dengan teknik penyeimbangan data efektif dalam meningkatkan kinerja model sekaligus efisien dalam proses anotasi.

Kata Kunci: Analisis Sentiment, IndoBERT, InSet Lexicon, Random Over Sampling, Bali, TikTok.