

PEMBANGUNAN MODEL KLASIFIKASI SENTIMEN PADA TWITTER TERHADAP PARIWISATA BALI MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

Oleh

Ni Luh Risma Pradnyadari, NIM 1915101052

Program Studi Ilmu Komputer

Jurusan Teknik Informatika

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model klasifikasi emosi dalam ulasan pariwisata tentang Bali, yang diperoleh dari Twitter. Informasi yang diberikan didasarkan pada tweet berbahasa Indonesia yang dikumpulkan secara otomatis melalui teknik perayapan menggunakan API Twitter, dari Januari hingga Desember 2023. Setelah pengumpulan, mahasiswa tingkat akhir Program Studi Pendidikan Bahasa Indonesia secara manual memberi label data dengan sentimen, seperti positif, negatif, dan netral. Proses pembersihan kemudian dilakukan, meliputi normalisasi teks, konversi huruf besar/kecil, tokenisasi, penghapusan kata-kata yang tidak informatif, dan lematisasi, untuk mendapatkan dataset bersih yang sesuai untuk analisis. Ekstraksi fitur dilakukan menggunakan metode TF-IDF, dengan tujuan mengubah teks menjadi format numerik. Untuk klasifikasi, algoritma SVM digunakan, memanfaatkan pengklasifikasi SVC dari pustaka Scikit-learn. Pemilihan parameter optimal dilakukan secara otomatis menggunakan GridSearchCV, mengeksplorasi berbagai jenis kernel (linier, polinomial, sigmoid) dan berbagai nilai parameter C. Untuk memvalidasi model, teknik K-Fold Cross Validation diterapkan dengan lima split ($n_splits = 5$) untuk meningkatkan kapasitas generalisasinya dan mencegah overfitting. Temuan menunjukkan bahwa model SVM mampu mengklasifikasikan emosi dengan kinerja yang memuaskan. Menurut evaluasi menggunakan matriks kebingungan, model yang paling efisien mencapai presisi 88%, akurasi 87%, recall 86%, dan skor F1 86%. Model tersebut secara efektif menunjukkan kemampuannya untuk mengidentifikasi pola emosional dalam tweet terkait pariwisata, terutama dengan membedakan antara komentar positif dan negatif. Hasil ini menunjukkan bahwa penggunaan SVM dalam kombinasi dengan TF-IDF cocok untuk mendeteksi opini publik terhadap pariwisata di Bali. Diharapkan penelitian ini dapat menjadi referensi untuk menerapkan analisis sentimen ke media sosial, membantu pengambilan keputusan dalam sektor pariwisata..

Kata Kunci: Sentimen, Twitter, Pariwisata Bali, Support Vector Machine (SVM), TF-IDF, K-Fold Cross Validation

**DEVELOPMENT OF A SENTIMENT CLASSIFICATION MODEL ON
TWITTER ON BALI TOURISM USING THE SUPPORT VECTOR
MACHINE (SVM) METHOD**

By

Ni Luh Risma Pradnyadari, NIM 1915101052

Computer Science Study Program Informatics Engineering

ABSTRACT

This study aims to develop an emotion classification model for tourism reviews about Bali, obtained from Twitter. The information was based on Indonesian-language tweets automatically collected through web crawling techniques using the Twitter API, from January to December 2023. After collection, final-year students from the Indonesian Language Education Study Program manually labeled the data with sentiments such as positive, negative, and neutral. A cleaning process was then carried out, including text normalization, case conversion, tokenization, removal of non-informative words, and lemmatization, to obtain a clean dataset suitable for analysis. Feature extraction was performed using the TF-IDF method, aiming to transform the text into a numerical format. For classification, the SVM algorithm was used, employing the SVC classifier from the Scikit-learn library. Optimal parameter selection was conducted automatically using GridSearchCV, exploring various kernel types (linear, polynomial, sigmoid) and different values of the parameter C. To validate the model, the K-Fold Cross Validation technique was applied with five splits ($n_splits = 5$) to enhance its generalization capacity and prevent overfitting. The findings indicate that the SVM model was able to classify emotions with satisfactory performance. Based on the evaluation using a confusion matrix, the most efficient model achieved a precision of 88%, accuracy of 87%, recall of 86%, and an F1 score of 86%. The model effectively demonstrated its ability to identify emotional patterns in tweets related to tourism, particularly in distinguishing between positive and negative comments. These results suggest that the use of SVM in combination with TF-IDF is suitable for detecting public opinion on tourism in Bali. It is expected that this research can serve as a reference for applying sentiment analysis to social media, assisting decision-making in the tourism sector.

Keywords: *Sentiment, Twitter, Bali Tourism, Support Vector Machine (SVM), TF-IDF, K-Fold Cross Validation.*