

ABSTRAK

Nama Mahasiswa (1) : Vincent Lius
NIM (1) : 41820010037
Nama Mahasiswa (2) : Huzaiyah Qahar Djauhari
NIM (2) : 41820010070
Nama Mahasiswa (3) : Ahmad Harits Ramadhani
NIM (3) : 41820010051
Dosen TA : Andi Nugroho, ST, M.Kom
Judul Tugas Akhir : Implementasi Naive Bayes Dengan Pendekatan Leksikon
Dan TF-IDF Dalam Analisis Sentimen

Di era digital dengan peningkatan penggunaan media sosial sebagai platform ekspresi publik, analisis sentimen yang akurat menjadi semakin penting, terutama untuk evaluasi ulasan aplikasi. Penelitian ini menekankan pentingnya klasifikasi sentimen yang akurat, terutama karena banyaknya ulasan aplikasi yang negatif. Kami mengusulkan pendekatan kombinasi yang mengintegrasikan algoritma Naive Bayes dengan pelabelan sentimen berbasis leksikon dan pembobotan TF-IDF untuk pelatihan model. Melalui penggunaan dataset yang terdiri dari 5000 ulasan diambil dari Kaggle mengenai ulasan aplikasi di Indonesia, penelitian ini mengeksplorasi leksikon bahasa Indonesia, khususnya InSet dan SentiStrengthID. Evaluasi efektivitas penggabungan Naive Bayes dengan TF-IDF dan metode berbasis leksikon menghasilkan kontribusi signifikan dalam pemahaman yang lebih mendalam tentang analisis sentimen. Dari evaluasi yang kami lakukan, kami mencapai kesimpulan bahwa leksikon SentiStrength menunjukkan kinerja yang lebih baik di semua parameter. Model yang menggunakan leksikon SentiStrengthID mencapai akurasi sebesar 86%, dengan nilai presisi 85%, recall 86%, dan F1-score 85%, yang lebih baik dibandingkan dengan penggunaan leksikon InSet, yang memiliki akurasi 70%, presisi 80%, recall 70%, dan F1-score 71%.

Kata kunci:

Analisis Sentimen, Leksikon InSet, Leksikon SentiStrengthID, Naive bayes, Ulasan Aplikasi.

ABSTRACT

Name (1) : Vincent Lius
NIM (1) : 41820010037
Name (2) : Huzairah Qahar Djauhari
NIM (2) : 41820010070
Name (3) : Ahmad Harits Ramadhani
NIM (3) : 41820010051
Counsellor : Andi Nugroho, ST, M.Kom
Title : Implementation Naive Bayes Algorithm using Lexicon
Approach and TF-IDF for Sentiment Analysis of Reviews

In the digital era marked by the increased use of social media as a platform for public expression, accurate sentiment analysis has become crucial, particularly in evaluating app reviews. This research highlights the need for precise sentiment classification, driven by the prevalence of negative feedback in several app reviews. To address this challenge, we propose a combined approach that integrates the Naive Bayes algorithm with lexicon-based sentiment labeling and TF-IDF weighting for model training. Utilizing a dataset of 5000 reviews from Kaggle regarding app reviews in Indonesia, this study explores Indonesian lexicons, specifically InSet and SentiStrengthID, for sentiment labeling. Evaluating the effectiveness of combining Naive Bayes with TF-IDF and lexicon-based methods provides significant insights into sentiment analysis in the context of app reviews. Based on our evaluation, we concluded that the SentiStrength lexicon performs better on all parameters. The model using the SentiStrengthID lexicon achieved an accuracy of 86%, with a precision of 85%, recall of 86%, and F1-score of 85%, which is better compared to the InSet lexicon, which had an accuracy of 70%, precision of 80%, recall of 70%, and F1-score of 71%.

Keywords:

Sentiment Analysis, Lexicon InSet, Lexicon SentiStrengthID, Naive Bayes, App Review.