

PENERAPAN SELEKSI FITUR *INFORMATION GAIN* DALAM KLASIFIKASI ULASAN *TRAVELOKA* DI *GOOGLE PLAY* STORE DENGAN *NAÏVE BAYES*

Putri Noer Kumalasari¹, Moh Dasuki², Miftahur Rahman³

Program Studi Teknik Informatika, Fakultas Teknik Universitas Muhammadiyah

Jember

Putrinoerkumalasarii@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk membandingkan kinerja analisis sentimen terhadap ulasan pengguna aplikasi Traveloka menggunakan algoritma Multinomial Naïve Bayes pada dua kondisi, yaitu penerapan seleksi fitur Information Gain dan tanpa seleksi fitur. Data penelitian dikumpulkan melalui proses web scraping dari Google Play Store dengan memanfaatkan library *google-play-scraper*, sehingga diperoleh sebanyak 500 ulasan berbahasa Indonesia. Pelabelan sentimen dilakukan dengan dua metode, yakni pelabelan manual oleh pakar bahasa serta pelabelan otomatis yang dilakukan oleh peneliti. Seluruh data kemudian diproses melalui tahapan *preprocessing* yang meliputi pembersihan data, tokenisasi, *case folding*, penghapusan *stopword*, dan *stemming*. Pada skenario pertama, seleksi fitur menggunakan metode Information Gain diterapkan untuk menentukan kata-kata yang memiliki pengaruh signifikan dalam proses klasifikasi sentimen, sedangkan pada skenario kedua digunakan pendekatan Bag of Words tanpa proses seleksi fitur. Evaluasi performa model dilakukan dengan metode *5-Fold Cross Validation* menggunakan metrik akurasi, *precision*, *recall*, *f1-score*, serta *confusion matrix* dengan membandingkan hasil klasifikasi terhadap label yang diberikan oleh pakar bahasa. Hasil penelitian menunjukkan bahwa model Naïve Bayes yang menerapkan seleksi fitur Information Gain menghasilkan rata-rata akurasi sebesar 87,4%, sementara model tanpa seleksi fitur memperoleh akurasi sebesar 86,0%. Temuan ini mengindikasikan bahwa penggunaan seleksi fitur Information Gain mampu meningkatkan kinerja serta efisiensi model dalam analisis sentimen ulasan pengguna.

Kata kunci: *Information Gain, Bag of Word, Google Play Scraper, Confusion Matrix, Naïve Bayes*

Implementation of Information Gain Feature Selection in Traveloka Review Classification on Google Play Store Using Naïve Bayes

Putri Noer Kumalasari¹, Moh Dasuki², Miftahur Rahman³

Informatic Engineering Study Program, Faculty of Engineering

University of Muhammadiyah Jember

Putrinoerkumalasarii@gmail.com

ABSTRACT

This study aims to compare the performance of sentiment analysis on user reviews of the Traveloka application using the Multinomial Naïve Bayes algorithm under two scenarios, namely with Information Gain feature selection and without feature selection. The research data were collected through a web scraping process from the Google Play Store using the google-play-scraper library, resulting in a total of 500 reviews written in Indonesian. Sentiment labeling was conducted using two approaches: manual labeling by language experts and automatic labeling performed by the researcher. The collected data were then processed through several preprocessing stages, including data cleaning, tokenization, case folding, stopword removal, and stemming. In the first scenario, feature selection was performed using the Information Gain method to identify words that have the most significant influence on sentiment classification, while in the second scenario a Bag of Words approach without feature selection was applied. Model evaluation was carried out using the 5-Fold Cross Validation method with performance metrics including accuracy, precision, recall, f1-score, and confusion matrix, by comparing the model predictions with labels provided by language experts. The results indicate that the Naïve Bayes model with Information Gain feature selection achieved an average accuracy of 87.4%, whereas the model without feature selection achieved an accuracy of 86.0%. These findings demonstrate that the application of Information Gain feature selection can improve the performance and efficiency of sentiment analysis models for user reviews.

Keywords: *Information Gain, Bag of Word, Google Play Scraper, Confusion Matrix, Naïve Bayes*