

ABSTRAK

Penerapan Metode *Multinomial Naive Bayes* dengan Optimasi TF-IDF dan Seleksi Fitur *Chi-Square* untuk Analisis Sentimen pada Ulasan Aplikasi *Getcontact* di *Google Play Store*

Moh. Reza Alfi Ali¹⁾, Hardian Oktavianto²⁾, Ginanjar Abdurrahman³⁾,
Nur Qodariyah Fitriyah⁴⁾, Luluk Handayani⁵⁾

Program Studi Teknik Informatika, Fakultas Teknik
Universitas Muhammadiyah Jember

Penelitian ini membahas analisis sentimen terhadap ulasan pengguna aplikasi *Getcontact* pada *Google Play Store* yang bersifat tidak terstruktur dan sulit dianalisis secara manual. Tujuan penelitian ini adalah mengklasifikasikan sentimen pengguna ke dalam kategori positif dan negatif menggunakan metode *Multinomial Naive Bayes* yang dioptimasi dengan pembobotan TF-IDF dan seleksi fitur *Chi-square*. Dataset yang digunakan terdiri dari 1000 ulasan yang telah melalui tahap *preprocessing*, kemudian dibagi menjadi 80% data pelatihan dan 20% data pengujian, serta dievaluasi menggunakan 200 *data unseen*. Representasi teks dilakukan menggunakan TF-IDF, kemudian dilanjutkan dengan seleksi fitur *Chi-square* untuk menentukan fitur yang paling relevan. Eksperimen dilakukan dengan variasi jumlah fitur yang digunakan. Hasil evaluasi menunjukkan akurasi sebesar 85% pada data pengujian dengan nilai presisi, *recall*, dan *F1-score* sebesar 85%. Pada *data unseen*, diperoleh akurasi sebesar 80% dengan nilai presisi, *recall*, dan *F1-score* sekitar 80%. Penurunan akurasi sebesar 5% dari data pengujian ke *data unseen* mengindikasikan bahwa model tidak mengalami *overfitting* yang signifikan dan memiliki kemampuan generalisasi yang memadai terhadap data baru.

Kata Kunci: *multinomial naive bayes*; TF-IDF; *chi-square*

ABSTRACT

Application of the Multinomial Naive Bayes Method with TF-IDF Optimization and Chi-Square Feature Selection for Sentiment Analysis of Getcontact Application Reviews on Google Play Store

Moh. Reza Alfi Ali¹⁾, Hardian Oktavianto²⁾, Ginanjar Abdurrahman³⁾,
Nur Qodariyah Fitriyah⁴⁾, Luluk Handayani⁵⁾

*Informatics Engineering Study Program, Faculty of Engineering
Universitas Muhammadiyah Jember*

This study addresses sentiment analysis of user reviews of the Getcontact application on the Google Play Store, which are unstructured and difficult to analyze manually. The objective of this research is to classify user sentiment into positive and negative categories using the Multinomial Naive Bayes method optimized with TF-IDF weighting and Chi-square feature selection. The dataset consists of 1,000 reviews that have undergone preprocessing, divided into 80% training data and 20% testing data, and evaluated further using 200 unseen reviews. Text representation is performed using TF-IDF, followed by Chi-square feature selection to identify the most relevant features. Experiments are conducted with variations in the number of selected features. The evaluation results show an accuracy of 85% on testing data with balanced precision, recall, and F1-score of 85%. Evaluation on unseen data produces an accuracy of 80% with precision, recall, and F1-score around 80%. The 5% decrease in accuracy from testing data to unseen data indicates that the model does not experience significant overfitting and demonstrates adequate generalization capability when applied to new data.

Keywords: multinomial naive bayes; TF-IDF; chi-square