

ABSTRAK

Analisis sentimen pada ulasan aplikasi *google play store* dapat digunakan untuk memahami persepsi pengguna terhadap kualitas aplikasi. Namun, performa model analisis sentimen tidak hanya dipengaruhi oleh arsitektur model, tetapi juga oleh teknik pelabelan yang digunakan dalam pembentukan *dataset* latih. Pelabelan berbasis *rating* sering digunakan karena praktis, tetapi berpotensi menimbulkan *label noise* karena skor pengguna tidak selalu sesuai dengan isi teks ulasan. Penelitian ini bertujuan untuk menganalisis pengaruh teknik pelabelan otomatis berbasis *rating*, InSet, dan SentiStrength_id terhadap performa model *deep learning* pada analisis sentimen ulasan Pokémon GO.

Dataset yang digunakan terdiri dari 8.806 ulasan aplikasi Pokémon GO berbahasa Indonesia dari *google play store* yang telah melalui tahap penyaringan dan pembersihan data. Model yang digunakan adalah CNN, LSTM, dan GRU sebagai model pembanding. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, uji *One Way ANOVA* digunakan untuk menguji pengaruh perbedaan teknik pelabelan terhadap performa model.

Hasil penelitian menunjukkan bahwa teknik pelabelan berpengaruh terhadap performa model. Pada CNN, akurasi pelabelan *rating*, InSet, dan SentiStrength_id masing-masing sebesar 73,95%, 87,68%, dan 90,14%. Pada LSTM, akurasi yang diperoleh sebesar 73,35%, 87,43%, dan 83,59%. Pada GRU, akurasi yang diperoleh sebesar 69,32%, 86,15%, dan 84,23%. Akurasi tertinggi sebesar 90,14% diperoleh pada kombinasi pelabelan SentiStrength_id dengan model CNN. Perbedaan akurasi yang cukup besar antara pelabelan *rating* dan pelabelan berbasis leksikon menunjukkan bahwa ketidaksesuaian antara *rating* dan isi teks ulasan berdampak pada kualitas pembelajaran model. Penelitian ini berkontribusi dalam memberikan bukti empiris bahwa teknik pelabelan otomatis merupakan faktor yang berpengaruh terhadap performa model *deep learning* sehingga dapat menjadi rekomendasi metodologis dalam pemilihan teknik pelabelan pada analisis sentimen ulasan aplikasi berbahasa Indonesia.

Kata kunci— Analisis Sentimen, Teknik Pelabelan Otomatis, *Deep Learning*, InSet, SentiStrength_id

ABSTRACT

Sentiment analysis of Google Play Store app reviews can be used to understand user perceptions of app quality. However, the performance of a sentiment analysis model is influenced not only by the model architecture but also by the labeling technique used to generate the training dataset. Rating-based labeling is often used due to its practicality, but it has the potential to introduce label noise because user scores do not always correspond to the review text. This study aims to analyze the effect of rating-based automatic labeling techniques, InSet, and SentiStrength_id on the performance of deep learning models in sentiment analysis of Pokémon GO reviews.

The dataset used consists of 8,806 Indonesian-language Pokémon GO app reviews from the Google Play Store that have undergone data filtering and cleaning. The models used were CNN, LSTM, and GRU as comparison models. Evaluation was conducted using accuracy, precision, recall, and F1-score metrics. Furthermore, a one-way ANOVA test was used to examine the effect of different labeling techniques on model performance.

The results showed that labeling techniques influenced model performance. On CNN, the accuracy of rating, InSet, and SentiStrength_id labeling was 73.95%, 87.68%, and 90.14%, respectively. On LSTM, the accuracy obtained was 73.35%, 87.43%, and 83.59%. On GRU, the accuracy obtained was 69.32%, 86.15%, and 84.23%. The highest accuracy of 90.14% was obtained by combining SentiStrength_id labeling with the CNN model. The significant difference in accuracy between rating labeling and lexicon-based labeling indicates that the mismatch between ratings and review text content impacts the quality of model learning. This study contributes to providing empirical evidence that automatic labeling techniques are an influential factor in the performance of deep learning models, thus providing methodological recommendations in selecting labeling techniques for sentiment analysis of Indonesian-language app reviews.

Keywords— Sentiment Analysis, Automatic Labeling Techniques, *Deep Learning*, InSet, SentiStrength_id