

INTISARI

PENGGUNAAN *EASY DATA AUGMENTATION* UNTUK KLASIFIKASI EMOSI PADA DATA TEKS TWITTER

oleh

ALGAMMA PARAMAYUDHA
20/459259/PA/19920

Perkembangan media sosial telah meningkat pesat dalam beberapa tahun terakhir, dengan Twitter menjadi salah satu platform yang populer di Indonesia. Twitter sering digunakan untuk diskusi dan percakapan dalam bentuk *tweet* serta sebagai sumber informasi terkini. Twitter juga menjadi ruang untuk mengekspresikan pandangan dan opini, yang sering kali mengandung elemen emosional yang dapat diolah lebih lanjut untuk memahami reaksi dan respon masyarakat. Klasifikasi emosi pada data teks Twitter membutuhkan dataset yang besar dan beragam untuk mencapai performa optimal. Namun, pengumpulan dan anotasi dataset secara manual dapat memakan waktu dan biaya yang besar.

Penelitian ini menerapkan dan mengevaluasi metode *Easy Data Augmentation* (EDA) untuk memperbesar ukuran dataset dalam klasifikasi emosi pada data Twitter berbahasa Indonesia. Penelitian ini bertujuan untuk melihat apakah peningkatan ukuran dataset melalui EDA dapat meningkatkan kinerja model klasifikasi *Support Vector Machine* (SVM) dan *Long Short Term Memory* (LSTM).

Setelah dilakukannya penelitian ini, diperoleh kesimpulan bahwa secara keseluruhan, penerapan metode *Easy Data Augmentation* (EDA) berhasil meningkatkan kinerja klasifikasi emosi pada data Twitter berbahasa Indonesia. Pada model SVM, skenario augmentasi SR + RS + RI meningkatkan F1-Score dari 65,12% pada data asli menjadi 68,12%, menunjukkan peningkatan sebesar 3%. Sementara itu, pada model LSTM, skenario augmentasi *Random Insertion* meningkatkan F1-Score dari 61,08% menjadi 65,31%, dengan peningkatan sebesar 4,23%. Secara keseluruhan, hasil terbaik dari eksperimen pada penelitian ini adalah model SVM dengan skenario augmentasi SR + RS + RI yang mendapatkan F1-Score tertinggi sebesar 68,12%.

Kata kunci: klasifikasi emosi, augmentasi data, *Easy Data Augmentation*, SVM, LSTM, TF-IDF, FastText, Twitter.

ABSTRACT

APPLICATION OF EASY DATA AUGMENTATION FOR EMOTION CLASSIFICATION ON TWITTER TEXT DATA

by

ALGAMMA PARAMAYUDHA
20/459259/PA/19920

The development of social media has rapidly increased in recent years, with Twitter emerging as one of the popular platforms in Indonesia. Twitter is frequently used for discussions and conversations through tweets, as well as a source of up-to-date information. Additionally, Twitter serves as a space for users to express their views and opinions, which often contain emotional elements that can be further analyzed to understand public reactions and responses. Emotion classification in Twitter text data requires a large and diverse dataset to achieve optimal performance. However, manually collecting and annotating datasets can be time-consuming and costly.

This study applies and evaluates the Easy Data Augmentation (EDA) method to expand the dataset size for emotion classification on Indonesian Twitter data. The study aims to determine whether increasing the dataset size through EDA can enhance the performance of classification models, specifically Support Vector Machine (SVM) and Long Short Term Memory (LSTM).

The results of this study indicate that, overall, the application of Easy Data Augmentation (EDA) successfully enhanced emotion classification performance for Indonesian Twitter data. In the SVM model, the SR + RS + RI augmentation scenario improved the F1-Score from 65.12% on the original data to 68.12%, reflecting a 3% increase. Meanwhile, in the LSTM model, the Random Insertion augmentation scenario improved the F1-Score from 61.08% to 65.31%, representing a 4.23% increase. Overall, the best result from the experiments in this study was achieved by the SVM model with the SR + RS + RI augmentation scenario, which achieved the highest F1-Score of 68.12%.

Keywords: emotion classification, data augmentation, Easy Data Augmentation, SVM, LSTM, TF-IDF, FastText, Twitter.