

INTISARI

PERBANDINGAN METODE AUGMENTASI DATA EDA PADA DETEKSI *HUMAN VALUES* DALAM DISKUSI ISU DI *GITHUB* DENGAN SVM

Oleh

Ahmad Fauzan

20/459257/PA/19918

Penelitian ini membahas tentang pendeteksian *human values* dalam pengembangan perangkat lunak yang sedang mendapatkan banyak perhatian khususnya dalam bidang industri perangkat lunak dan para penggunanya, contohnya pada kasus pelanggaran privasi Facebook-Cambridge Analytica yang populer. Studi menunjukkan bahwa mengabaikan *human values* seperti privasi, keadilan, dan kesetaraan dapat menyebabkan ketidakpuasan pengguna dan dampak sosial-ekonomi yang negatif. Beberapa metode yang sudah diusul salah satunya deteksi *human values* menggunakan machine learning. Akan tetapi, performa yang dihasilkan belum maksimal karena data yang tidak seimbang dan berjumlah relatif sedikit.

Penelitian ini berfokus pada penggunaan augmentasi data untuk meningkatkan performa deteksi *human values* dalam pengembangan perangkat lunak. Penelitian menggunakan metode EDA (*Easy Data Augmentation*) untuk memperbesar dataset kemudian membandingkan kinerja model hasil penggunaan metode augmentasi dengan model yang tidak menggunakan augmentasi data.

Berdasarkan hasil augmentasi dan pengujian, metode Random Insertion (RI) dan Synonym Replacement (SR) menunjukkan performa terbaik dengan peningkatan F1 Score masing-masing sebesar 4.15% dan 1.14% dibandingkan model tanpa augmentasi, membuktikan efektivitas pemilihan metode yang tepat. Sebaliknya, metode Random Deletion (RD) dan Random Swap (RS) menyebabkan bias, dengan RD mengurangi informasi penting pada label positif dan RS meningkatkan asosiasi kata positif, sehingga tidak optimal jika hanya data positif yang diaugmentasi. Selain itu, metode kombinasi yang mencakup RD dan RS cenderung memiliki performa lebih rendah dibandingkan dengan kombinasi metode lainnya.

Kata-kata kunci : *Human Values*, Augmentasi Data, Isu Github, SVM.

ABSTRACT

COMPARISON OF EDA DATA AUGMENTATION METHODS IN HUMAN VALUES DETECTION IN GITHUB'S ISSUE DISCUSSION USING SVM

By

Ahmad Fauzan

20/459257/PA/19918

This study discusses the detection of human values in software development, which has garnered significant attention, especially in the software industry and among its users, exemplified by the popular case of Facebook-Cambridge Analytica privacy violations. Studies show that neglecting human values such as privacy, fairness, and equality can lead to user dissatisfaction and negative socio-economic impacts. Several methods have been proposed, one of which is detecting human values using machine learning. However, the performance achieved has not been optimal due to imbalanced and relatively small datasets.

This research focuses on the use of data augmentation to improve the performance of human values detection in software development. The study employs the Easy Data Augmentation (EDA) method to enlarge the dataset and then compares the performance of models using augmentation techniques with those that do not use data augmentation.

Based on the results of augmentation and testing, the Random Insertion (RI) and Synonym Replacement (SR) methods demonstrate the best performance, with F1 Score increases of 4.15% and 1.14%, respectively, compared to the model without augmentation, highlighting the effectiveness of selecting appropriate methods. Conversely, Random Deletion (RD) and Random Swap (RS) introduce biases, with RD reducing crucial information for positive labels and RS amplifying the association with positive words, making them less effective when only positive data is augmented. Additionally, combinations including RD and RS generally perform worse compared to other method combinations

Keyword : *Human Values, Data Augmentation, Github Issue, SVM.*