

INTISARI

DETEKSI *REVIEW* DUPLIKAT PADA *REVIEWS* APLIKASI *GOOGLE PLAY* BERDASARKAN TINGKAT KEMIRIPAN MENGGUNAKAN ALGORITMA JARO WINKLER

Oleh

Ilham Dwi Tara Hidayat

14/364161/PA/15922

Salah satu permasalahan dalam menguji performa aplikasi pada Google Play Store karena adanya *review SPAM* duplikat. Keberadaan *review SPAM* sendiri merugikan pengguna dan mengganggu para pengembang dalam melakukan evaluasi sebuah aplikasi. Hal ini dikarenakan *review SPAM* tidak memberikan informasi valid terkait fitur aplikasi yang diulas. Penelitian ini bertujuan membangun dan merancang sistem untuk mendeteksi *SPAM* duplikat pada ulasan *Google Play* dengan mengukur dan menganalisis tingkat kemiripan antar ulasan.

Dalam penelitian ini dilakukan penambahan sistem deteksi *SPAM* duplikat pada sistem klasifikasi yang menggunakan algoritma *Support Vector Machine (SVM)*. Algoritma yang digunakan untuk mendeteksi *SPAM* duplikat adalah algoritma Jaro Winkler. Tingkat kemiripan antar reviews dapat dihitung menggunakan algoritma Jaro Winkler. Selanjutnya dipilih nilai *threshhold* yang paling optimal untuk diterapkan pada sistem.

Hasil akhir dari penelitian ini menunjukkan terjadi peningkatan akurasi klasifikasi *review* sebesar 2,3 % setelah dilakukan penambahan sistem deteksi *SPAM* duplikat terhadap sistem klasifikasi *SVM*.

Kata kunci: Kemiripan, Jaro Winkler, *SPAM*, *HAM*, *SPAM Detection*, Data Duplikat.

ABSTRACT

DUPLICATE REVIEW DETECTION ON GOOGLE PLAY APP REVIEWS BASE ON SIMILARITY LEVELS USING JARO WINKLER ALGORITHM

By

Ilham Dwi Tara Hidayat
14/364161/PA/15922

One of the problems in testing application performance on the Google Play Store is because of a duplicate SPAM review. The existence of SPAM review harms users and interferes the developers in evaluating an application. This is because the SPAM review does not provide valid information regarding the application features reviewed. This study aims to build and design a system for detecting duplicate SPAM on Google Play reviews by measuring and analyzing the degree of similarity between reviews.

In this research, the addition of a duplicate SPAM detection system was added to the classification system which has been used the Support Vector Machine (SVM) algorithm. The algorithm used to detect duplicate SPAM is the Jaro Winkler algorithm. The level of similarity between reviews can be calculated using the Jaro Winkler algorithm. Then the most optimal threshold value is chosen to be applied to the system.

The final results of this study indicate an increase in the review classification accuracy of 2,3% after the addition of a duplicate SPAM detection system to the SVM classification system.

Keyword: Similarity, Jaro Winkler, SPAM, HAM, SPAM Detection, Duplication Data.