

INTISARI

OPTIMALISASI DETEKSI SPAM PADA THREAD TWITTER MENGUNAKAN LIGHTGBM DAN XGBOOST MELALUI INTEGRASI MULTI-FITUR TERHADAP DATA TIDAK SEIMBANG

Oleh

Aulia Al-Jihad Safhadi

23/525903/PPA/06625

Twitter merupakan salah satu platform media sosial yang menjadi sasaran empuk bagi pelaku spam untuk menyebarkan konten tidak relevan, termasuk tautan afiliasi atau referral yang berbahaya. Spam di Twitter dapat muncul dalam bentuk serangkaian *tweet (thread)* yang terlihat seperti pesan normal namun sebenarnya bertujuan untuk mengelabui pengguna dan sistem deteksi spam. Deteksi spam pada Twitter menghadapi tantangan signifikan, terutama terkait dengan ketidakseimbangan dataset (*imbalanced dataset*) dan pendekatan yang hanya fokus pada salah satu fitur, seperti fitur berbasis *text*, *user*, atau *content*. Oleh karena itu, dibutuhkan pengembangan metode yang efektif pada kasus serangkaian *tweet (thread)* dalam mengatasi masalah *imbalanced dataset* serta dapat mengintegrasikan fitur berbasis *text*, *user*, dan *content* dalam sistem deteksi spam di Twitter. Penelitian ini mengusulkan pendekatan menggunakan serangkaian *tweet (thread)* dengan penggunaan bobot sebagai cara mengatasi *imbalanced dataset* dan pengelompokan fitur berbasis *text*, *user*, atau *content* pada algoritma Gradient Boosting Machine (GBM) seperti LightGBM dan XGBoost untuk mendeteksi pesan spam di Twitter. Setiap hasil akurasi pada *single tweet* ataupun serangkaian *tweet (thread)* dengan penggunaan bobot atau tanpa bobot pada setiap kelompok fitur dikomparasikan untuk setiap metode. Pengintegrasian fitur berbasis *text*, *user* dan *content* pada serangkaian *tweet (thread)* dengan penggunaan bobot menggunakan XGBoost memberikan akurasi terbaik sebesar 97.90%.

Kata Kunci: NLP, Spam, Klasifikasi, Twitter, LightGBM, XGBoost, Thread, *Imbalanced Dataset*

ABSTRACT

OPTIMIZING TWITTER SPAM DETECTION THROUGH TWITTER THREADS USING LIGHTGBM AND XGBOOST VIA MULTI-FEATURE INTEGRATION ON IMBALANCED DATA

By

Aulia Al-Jihad Safhadi

23/525903/PPA/06625

Twitter is one of the social media platforms that is an easy target for spammers to spread irrelevant content, including malicious affiliate or referral links. Spam on Twitter can appear in the form of a series of tweets (threads) that look like normal messages but are actually intended to trick users and spam detection systems. Spam detection on Twitter faces significant challenges, especially related to imbalanced datasets and approaches that only focus on one feature, such as text-based, user-based, or content-based features. Therefore, it is necessary to develop an effective method in the case of a series of tweets (threads) to overcome the problem of imbalanced datasets and can integrate text-based, user-based, and content-based features in a spam detection system on Twitter. This study proposes an approach using a series of tweets (threads) with the use of weights as a way to overcome imbalanced datasets and grouping text-based, user-based, or content-based features in the Gradient Boosting Machine (GBM) algorithm such as LightGBM and XGBoost to detect spam messages on Twitter. Each accuracy result on a single tweet or a series of tweets (threads) with the use of weights or without weights on each feature group is compared for each method. Integration of text, user and content-based features in a series of tweets (threads) with the use of weights using XGBoost provides the best accuracy of 97.90%.

Keywords: NLP, Spam, Classification, Twitter, LightGBM, XGBoost, Thread, Imbalanced Dataset