

INTISARI

DETEKSI KOMENTAR SPAM PADA INSTAGRAM MENGGUNAKAN COMPLEMENTARY NAÏVE BAYES

Oleh:

NUR AZIZUL HAQIMI

16/403703/PPA/05220

Instagram (IG) merupakan salah satu aplikasi media sosial berbasis web dan *mobile* dimana pengguna dapat membagikan foto atau video dengan fitur yang tersedia. Mengunggah foto atau video dengan *caption* yang berisi penjelasan tentang foto atau video tersebut dapat menuai komentar *spam*. Komentar *spam* berisi komentar yang tidak relevan dengan *caption* dan foto. Keberadaan *spam* menjadi penghalang untuk mendapatkan informasi dengan tingkat kesesuaian konten tinggi.

Ketika mengidentifikasi komentar *spam* dan *nonspam*, masalah yang menantang adalah komentar *spam* lebih sedikit daripada komentar *nonspam* sehingga mengarah pada masalah ketidakseimbangan data (*imbalance dataset*). *Imbalanced dataset* dapat memberikan pengaruh terhadap performa suatu algoritma klasifikasi. Metode *Complementary Naïve Bayes* (CNB) diketahui dapat menghitung ketidakmungkinan data pada suatu kelas tertentu dengan mengidentifikasi bahwa data tersebut berada di kelas lain. Hal tersebut menjadi fokus penelitian terkait dengan implementasi metode CNB dalam menangani *imbalance dataset* pada deteksi komentar *spam* Instagram. Penelitian menggunakan pembobotan TF-IDF dengan metode klasifikasi *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC) sebagai klasifikasi pembandingan.

Berdasarkan hasil pengujian dengan data latih sejumlah 2500 komentar dan data uji sejumlah 100 komentar pada kasus *imbalanced dataset* (25% *spam* dan 75% *nonspam*) maka didapatkan akurasi CNB sebesar 92%, *precision* sebesar 86%, *recall* sebesar 100% dan *f-measure* sebesar 93%. Metode SVM menghasilkan akurasi sebesar 87%, *precision* sebesar 79%, *recall* sebesar 100% dan *f-measure* sebesar 88%. NBC memiliki akurasi sebesar 92%, *precision* sebesar 90%, *recall* sebesar 94% dan *f-measure* sebesar 93%. Kesimpulannya metode CNB dan NBC lebih cocok untuk mendeteksi komentar spam dengan kasus *imbalanced dataset*.

Kata kunci: Instagram, *Spam*, *Complementary Naïve Bayes*, *Naïve Bayes Classifier*, *Support Vector Machine*.

ABSTRACT

DETECTION OF SPAM COMMENTS ON INSTAGRAM USING COMPLEMENTARY NAÏVE BAYES

By:

NUR AZIZUL HAQIMI
16/403703/PPA/05220

Instagram (IG) is a web-based and mobile social media application where users can share photos or videos with available features. Upload photos or videos with captions that contain an explanation of the photo or video that can reap spam comments. Comments on spam containing comments that are not relevant to the caption and photos. The existence of spam is a barrier to getting information with a high level of content suitability.

When identifying spam and nonspam comments, a challenging problem is that spam comments are less than nonspam comments, leading to problems with imbalanced datasets. Imbalanced datasets can influence the performance of a classification algorithm. The Complementary Naïve Bayes (CNB) method is known to be able to calculate the impossibility of data in a particular class by identifying that the data is in another class. This is the focus of research related to the development of the CNB method in dealing with imbalance datasets for detection of Instagram spam comments. The study uses weighting TF-IDF with the Support Vector Machine (SVM) classification method and Naïve Bayes Classifier (NBC) as a comparison classification.

Based on the test results with training data of 2500 comments (25% spam and 75% non-spam) and test data for 100 comments in the case of imbalanced data, the results of accuracy were 92%, precision 86%, recall 100% and f-measure of 93% on the CNB method. The SVM method produces an accuracy of 87%, precision of 79%, recall of 100% and f-measure of 88%. NBC has an accuracy of 92%, precision of 90%, recall of 94% and f-measure of 93%. In conclusion the CNB method is more suitable for detecting spam comments with cases of imbalanced datasets.

Keywords: Instagram, Spam, Complementary Naïve Bayes, Naïve Bayes Classifier, Support Vector Machine.