

INTISARI

KLASIFIKASI KOMENTAR *SPAM* DAN ANALISIS *SPAMMER* PADA *INSTAGRAM*

Yuliawan Rizka Syafaat

12/334659/PA/14892

Perkembangan dunia media sosial memicu perkembangan *spam* yang tinggi. Permasalahan *spam* disebabkan oleh *spammer*, yaitu pengguna yang cenderung lebih banyak melakukan *spam* bila dibandingkan dengan seluruh kiriman yang pengguna tersebut miliki.

Penelitian ini menggunakan pendekatan *Supervised Learning* yaitu algoritma *Naive Bayes* dan *Complement Naive Bayes* untuk melakukan klasifikasi *spam* pada komentar di Instagram. Pada setiap algoritma di atas, kalimat komentar dipecah menggunakan tokenisasi untuk dijadikan pemodelan dan digunakan untuk melakukan klasifikasi. Kemudian algoritma *clustering* digunakan untuk mengetahui persebaran *spam* dan *spammer* berdasarkan hasil klasifikasi.

Penelitian ini menunjukkan hasil klasifikasi menggunakan *Complement Naive Bayes* dapat mencapai nilai *Precision* 98,86% pada kelas bukan *spam* dan 81,47% pada kelas *spam*, nilai *Recall* 81,23% pada kelas bukan *spam* dan 98,88% pada kelas *spam*, nilai *Accuracy* 89,26% dan nilai *F-Measure* 89,18%. Hasil parameter pengujian menunjukkan bahwa *Complement Naive Bayes* cocok digunakan untuk penelitian ini, yaitu dimana jumlah masing-masing data berbeda. *Clustering spammer* dilakukan menggunakan *k-Means* dengan 2 *cluster* yaitu *spammer* dan bukan *spammer*. *Clustering* menghasilkan nilai *silhouette coefficient* dapat mencapai nilai 0,930 yang berarti hasil *clustering* memiliki hasil yang memiliki kepadatan yang tinggi dan keterpisahan *cluster* yang cukup jauh dari *cluster* lainnya.

Kata kunci: klasifikasi, *clustering*, *spam*, *spammer*, *instagram*, *naive bayes*, *complement naive bayes*, *k-means*.

ABSTRACT

CLASSIFICATION OF SPAM COMMENTS AND SPAMMER ANALYSIS ON INSTAGRAM

Yuliawan Rizka Syafaat

12/334659/PA/14892

The development of social media lead to the development of spam. This spam problems caused by spammers, which is users who have tendency to spam when compared with the rest of the posts that users have.

This research used Supervised Learning approach which is Naive Bayes and Complement Naive Bayes algorithm to classify spam in Instagram comment. In the algorithm used, comment sentences is tokenized and used as models for classifications. Clustering approach then used to determine the distribution of spam and spammers in Instagram comment using classification results.

This study indicate the classification results using the Complement Naive Bayes can reach a value of Precision 98.86% in non spam class and 81.47% in spam class, Recall value of 81.23% in non spam class and 98.88% in spam class, Accuracy value is 89.26% and F-Measure value is 89.18%. This result shows that Complement Naive Bayes is suitable to used in this study, in which the respective number of data is different. Spammer clustering performed using k-Means using 2 clusters, spammer and non spammer. Clustering producing silhoutte coefficient value that can reach 0.930 which means this clustering have the densest cluster while at the same time every cluster is separated from each other cluster.

Keywords: classification, clustering, spam, spammers, instagram, naive bayes, complement naive bayes, k-means.