

INTISARI

KLASIFIKASI TWEET BERITA *CLICKBAIT* DAN *NON-CLICKBAIT* PADA TWITTER MENGGUNAKAN *TEXT MINING*

Oleh

Sastra Anugrah Bimantara
15/379624/PA/16682

Clickbait adalah judul atau cuplikan berita yang kata-katanya menunjukkan fakta yang dilebih-lebihkan atau fakta yang tidak lengkap. Fenomena ini merugikan baik bagi pembaca maupun bagi penyedia berita lain yang masih kompeten. Salah satu penyebaran *clickbait* adalah melalui media sosial Twitter. Pendeteksian *clickbait* pada Twitter perlu dilakukan untuk meminimalisasi kerugian yang ditimbulkan oleh *clickbait*.

Pada penelitian ini dilakukan klasifikasi *tweet* berita *clickbait* menggunakan data *tweet* dari beberapa akun Twitter portal berita menggunakan metode Support Vector Machine dan Multinomial Naïve Bayes. Data yang digunakan adalah sejumlah 3.400 dengan komposisi yang seimbang. Pengambilan data dilakukan bersamaan untuk lima dari enam akun yaitu pada tanggal 29 Desember 2019, sementara akun sisanya diambil lebih akhir yaitu pada tanggal 22 Maret 2019. Metode ekstraksi fitur yang digunakan adalah metode TF-IDF dan nilai performa yang dicari adalah akurasi, presisi, *recall*, dan *F1-Score*.

Hasil dari penelitian ini adalah performa berupa akurasi sebesar 85,6%, presisi sebesar 86%, *recall* sebesar 85,8% dan *f1-score* sebesar 85,6% untuk model Support Vector Machine, serta akurasi sebesar 76,4%, presisi sebesar 82,4%, *recall* sebesar 77,1% dan *f1-score* sebesar 75,6% untuk model Multinomial Naïve Bayes.

Kata Kunci: Clickbait, Support Vector Machine, Multinomial Naïve Bayes, TF-IDF, Klasifikasi Dokumen Teks

ABSTRACT

CLICKBAIT AND NON-CLICKBAIT NEWS TWEETS CLASSIFICATION ON TWITTER USING TEXT MINING

By

Sastra Anugrah Bimantara

15/379624/PA/16682

Clickbait is a title or snippet of news which words indicate exaggerated facts or incomplete facts. This phenomenon is detrimental to both the reader and other competent news providers. One of the spreads of clickbait is through social media Twitter. Clickbait detection on Twitter needs to be done to minimize the loss caused by clickbait.

In this study the classification of clickbait news tweets is done using tweet data from several news portal Twitter accounts using the Support Vector Machine and Multinomial Naïve Bayes method. The data used is a number of 3.400 with a balanced composition. Data retrieval is done simultaneously for five of the six accounts, which is on December 29, 2019, while the remaining account is taken later, which is on March 22, 2019. The feature extraction method used is the TF-IDF method and the performance values calculated are accuracy, precision, recall and f1-Score.

The results of this study are performances in the form of accuracy of 85.6%, precision of 86%, recall of 85.8% and f1-score of 85.6% for Support Vector Machine models, and accuracy of 76.4%, precision of 82.4%, recall of 77.1% and f1-score of 75.6% for the Multinomial Naïve Bayes model.

Keywords: Clickbait, Support Vector Machine, Multinomial Naïve Bayes, TF-IDF, Text Document Classification