

INTISARI

IDENTIFIKASI DAN KLASIFIKASI ARGUMEN PADA TWITTER

Oleh

ILHAM FATHONI

15/379620/PA/16678

Melalui media sosial seperti Twitter, semua pengguna dapat mengutarakan pendapatnya. Fenomena ini dapat memantik adu argumen antar pengguna. Pengklasifikasian argumen dapat dijadikan sebagai indikator diskusi yang konstruktif dan pencegahan penyebaran disinformasi.

Pada penelitian ini dilakukan klasifikasi untuk membedakan argumen dari non-argumen dan argumen faktual dari opini. Data yang digunakan berupa *tweet* berbahasa Indonesia. Klasifikasi dilakukan menggunakan metode Random Forest dan Logistic Regression. Ekstraksi fitur dilakukan dengan metode TF-IDF dengan fitur *unigram* dan *uni-bigram*.

Hasil terbaik klasifikasi *tweet* argumen/non-argumen didapat dengan menggunakan metode Random Forest dengan f_1 score sebesar 0,77. Sedangkan, hasil terbaik klasifikasi *tweet* argumen faktual/opini didapat dengan menggunakan metode Logistic Regression dengan f_1 score sebesar 0,83.

Kata kunci: Klasifikasi Teks, Media Sosial, Argumen, Opini, Random Forest, Logistic Regression

ABSTRACT

ARGUMENT IDENTIFICATION AND CLASSIFICATION ON TWITTER

By

ILHAM FATHONI

15/379620/PA/16678

Through social media such as Twitter, everyone can express their opinion. This phenomenon can trigger heated arguments between user. Arguments detection can be used as an indicator of constructive discussions and to prevent the spread of disinformation.

In this study, classifications were carried out to distinguish arguments from non-arguments and factual arguments from opinions. Indonesian tweets are used as data source. The classifications were done using Random Forest and Logistic Regression method. Feature extraction were done using TF-IDF with unigram and uni-bigram features.

The best result of arguments/non-arguments classification was obtained using Random Forest method with f_1 score of 0,77. On the other hand, the best result of factual arguments/opinions classification was obtained using Logistic Regression method with f_1 score of 0,83.

Keywords: Text Classification, Social Media, Argument, Opinion, Random Forest, Logistic Regression