

INTISARI

PERBANDINGAN KINERJA *STOPWORD REMOVAL* DALAM KLASIFIKASI TEKS BERITA EKONOMI BERBAHASA INDONESIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Oleh

Andi Ariska Yogatama

12/334720/PA/14952

Perkembangan teknologi yang semakin pesat membawa perubahan yang besar terhadap kehidupan masyarakat. Salah satunya adalah akses berita *online* yang semakin meningkat. Penelitian sebelumnya telah dilakukan mengenai klasifikasi teks berita ekonomi menggunakan metode Multinomial Naïve Bayes. Hasil dari penelitian tersebut diperoleh rata-rata 74,77%. Dari hasil tersebut maka dilakukan penelitian mengenai klasifikasi teks berita ekonomi namun dengan metode yang berbeda.

Pada penelitian kali ini dilakukan klasifikasi teks berita ekonomi menggunakan metode *Support Vector Machine*. Dataset yang digunakan dalam penelitian ini adalah berita ekonomi *online* sejumlah 1233 berita yang terbagi dalam 9 kategori sektor ekonomi yang mengacu pada pendekatan produksi dalam menentukan angka Produk Domestik Bruto. Dalam penelitian kali ini juga dilakukan perbandingan performa antara penggunaan *stopword removal* dan tanpa *stopword removal* dalam *preprocessing*. Fitur yang digunakan adalah 1000 kata yang paling sering muncul dalam dataset dengan pembobotan tf-idf.

Dari fitur tersebut dilakukan klasifikasi dengan metode Support Vector Machine. Hasil dari pengujian didapatkan hasil rata-rata akurasi 75,27% untuk data tanpa *stopword*. Sedangkan untuk data dengan *stopword* didapatkan hasil rata-rata 75,43%.

Kata kunci: *text mining*, klasifikasi teks, *Support Vector Machine*, *stopword removal*

ABSTRACT

PERFORMANCE COMPARISON OF USING STOPWORD REMOVAL FOR INDONESIAN ECONOMY NEWS CLASSIFICATION USING SUPPORT VECTOR MACHINE METHOD

By

Andi Ariska Yogatama

12/334720/PA/14952

The rapid development of technology has brought great changes to people's lives. One of them is increasing online news access. Previous research has been conducted on the classification of economic news texts using the Multinomial Naïve Bayes method. The results of the study obtained an average of 74.77%. From these results, a study of the classification of economic news texts was conducted but with different methods.

In this research the economic news text classification was conducted using the Support Vector Machine method. The dataset used in this research is online economic news totaling 1233 news items which are divided into 9 categories of economic sectors that refer to the production approach in determining Gross Domestic Product figures. In this study also made a comparison between the performance of using stopword removal and without stopword removal. The feature that is used is 1000 words that appear most frequently in a dataset with weighted tf-idf.

From this feature classification is supported by the Support Vector Machine method. The results of these tests obtained an average accuracy of 75.27% for the dataset without stopword. While for testing dataset with stopword, the average yield is 75.43%.

Keywords: text mining, text classification, Support Vector Machine, stopword removal