

INTISARI

PERBANDINGAN KINERJA PORTER STEMMER DAN NAZIEF-ADRIANI STEMMER PADA KLASIFIKASI TEKS BERITA BERBAHASA INDONESIA MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES

Oleh

Muhammad Renaldy
12/331461/PA/14714

Meningkatnya pengguna internet dalam mengakses berita online khususnya berita ekonomi, mendorong Trisniantari D. (2016) untuk melakukan penelitian mengenai klasifikasi teks berita ekonomi. Metode klasifikasi yang digunakan dalam penelitian tersebut adalah Multinomial Naïve Bayes dan metode untuk *stemming* adalah Nazief-Adriani Stemmer. Dan dalam penelitian tersebut diperoleh rata-rata *processing time* untuk melakukan *preprocessing* berita cukup lama yaitu 7.5182 detik untuk satu teks berita ekonomi. Maka dalam penelitian kali ini akan dilakukan perbandingan *processing time* tahap *preprocessing* dan akurasi klasifikasi dengan tiga metode *stemming*, Nazief-Adriani Stemmer, Porter Stemmer, dan Porter Stemmer dengan tambahan kamus, yang diterapkan pada tahap *preprocessing*. *Processing time* pada tahap *preprocessing* didapatkan dengan menghitung selisih waktu saat *preprocessing* dimulai dan waktu saat *preprocessing* selesai. Sedangkan akurasi klasifikasi teks berita ekonomi didapatkan dengan menggunakan metode 10-Fold Cross Validation.

Pada penelitian kali ini didapatkan hasil bahwa Porter Stemmer Tala (2013) tanpa kamus memiliki kecepatan *preprocessing* paling cepat dengan waktu rata-rata 0.0044 detik untuk setiap teks berita ekonomi pada data *training* dan waktu rata-rata 0.0044 detik untuk setiap teks berita ekonomi pada data *test* namun akurasi klasifikasi yang dihasilkan paling rendah yaitu 73.31%. Sedangkan Nazief-Adriani memiliki kecepatan *preprocessing* paling lambat dengan waktu rata-rata 2.1048 untuk setiap teks berita ekonomi pada data *training* dan waktu rata-rata 1,8853 detik untuk setiap teks berita ekonomi pada data *test* namun rata-rata akurasi klasifikasi yang dihasilkan paling tinggi yaitu 74,45%. Dan untuk Porter Stemmer Tala (2013) dengan kamus rata-rata akurasi klasifikasi yang didapatkan adalah 74.12% dengan waktu rata-rata *preprocessing* 1.6735 detik untuk setiap teks berita ekonomi pada data *training* dan waktu rata-rata 1.7 detik untuk setiap teks berita ekonomi pada data *test*.

Kata kunci : klasifikasi teks, *text mining*, multinomial Naïve Naves, *stemming* bahasa Indonesia, Nazief-Adriani Stemmer, Porter Stemmer untuk bahasa Indonesia

ABSTRACT

PERFORMANCE COMPARISON OF PORTER STEMMER AND NAZIEF- ADRIANI STEMMER FOR INDONESIAN ECONOMY NEWS CLASSIFICATION USING MULTINOMIAL NAÏVE BAYES METHOD

By

Muhammad Renaldy
12/331461/PA/14714

The growing of internet users in accessing online news especially economic news, encouraging Trisniantari D. (2016) to conduct research on the classification of economic news text. The classification method used in the research is Multinomial Naïve Bayes and the method for stemming is Nazief-Adriani Stemmer. And in the research, the average processing time for preprocessing the news is quite long, about 7,5182 seconds for one text of economic news. So in this research will be done comparative processing time for preprocessing stage and classification accuracy with three methods stemming, Nazief-Adriani Stemmer, Porter Stemmer, and Porter Stemmer with additional dictionary, which is applied at the preprocessing stage. Processing time at the preprocessing stage is obtained by calculating the time difference when preprocessing begins and the time when preprocessing is complete. While the accuracy of text classification of economic news is obtained by using the 10-Fold Cross Validation method.

In this study we found that the Porter Stemmer Tala (2013) without a dictionary has the fastest preprocessing speed with an average time of 0.0044 seconds for each text of economic news on training data and an average time of 0.0044 seconds for each text of economic news on test data but the accuracy of the resulting classification is the lowest at 73.31%. While Nazief-Adriani has the slowest preprocessing speed with an average time of 2.1048 for each text of economic news on training data and an average time of 1.8853 seconds for each text of economic news on the data test but the accuracy of the resulting classification is the highest at 74.45%. And for Porter Stemmer Tala (2013) with additional dictionary classification accuracy obtained is 74.12% with an average preprocessing time of 1.6735 seconds for each text of economic news on training data and an average time of 1.7 seconds for each text of economic news on the data test.

Keywords : text classification, text mining, multinomial Naïve Nayes, bahasa Indonesia steeming, Nazief-Adriani Stemmer, Porter Stemmer for bahasa Indonesia