

## INTISARI

### PERBANDINGAN KINERJA ALGORITMA *IMPROVED K-NEAREST NEIGHBORS* DAN *MULTINOMIAL NAÏVE BAYES CLASSIFIER* UNTUK *KLASIFIKASI BERITA ONLINE BERDASARKAN CABANG OLAHRAGA*

Oleh

Farhan Ramadhani  
14/363805/PA/15861

Peningkatan jumlah artikel berita yang ada di internet memberikan permasalahan baru terhadap tugas seorang editor dalam mengkategorikan berita. Metode *text mining* mampu melakukan proses pengkategorian suatu data teks secara otomatis menggunakan proses klasifikasi. Akan tetapi, permasalahan yang timbul dari proses klasifikasi teks adalah ukuran dimensi dari data yang bisa mempengaruhi performa dari proses klasifikasi itu sendiri. Sehingga ukuran dimensi dari data harus dikurangi dengan melakukan seleksi terhadap atribut-atribut yang ada. Metode ini disebut sebagai seleksi fitur.

Seleksi fitur akan mengurangi dimensionalitas dari suatu data dengan mengabaikan fitur-fitur yang dianggap tidak representatif terhadap label manapun dari suatu data. Beberapa studi menunjukkan proses seleksi fitur bisa meningkatkan efektivitas dan akurasi. Pada penelitian ini digunakan metode seleksi fitur *TF-IDF* dan *chi square*. Hasil fitur yang telah diseleksi dari kedua metode, selanjutnya akan digunakan untuk klasifikasi oleh algoritma *K-Nearest Neighbors Decision Rules*. Hasil evaluasi performa dari algoritma klasifikasi *K-Nearest Neighbors Decision Rules* akan dibandingkan dengan *Multinomial Naive Bayes* dengan mengamati parameter evaluasi akurasi, presisi, *recall* dan *f-measure*.

Hasil penelitian ini menunjukkan metode seleksi fitur *chisquare* & *TF-IDF* mampu meningkatkan akurasi, presisi, *recall* & *f-measure* dari algoritma *K-Nearest Neighbors Decision Rules*. Akurasi dari klasifikasi menggunakan *chisquare* meningkat 10.68% sedangkan menggunakan *TF-IDF* meningkat 11.32%. Parameter presisi meningkat 5.8% dan 5.6% menggunakan *TF-IDF* & *chisquare* secara berurutan. *Recall* meningkat 10.50% menggunakan *TF-IDF* dan meningkat 10.7% menggunakan *chisquare*. Parameter *f-measure* meningkat 9.9% menggunakan metode seleksi fitur *chisquare*, *TF-IDF* meningkatkan *f-measure* hingga 8.7%. Kedua metode seleksi fitur juga meningkatkan performa running time dari algoritma *K-Nearest Neighbors Decision Rules*. Namun peningkatan evaluasi parameter dari *K-Nearest Neighbors Decision Rules* belum mengungguli dari performa klasifikasi *Multinomial Naive Bayes* dalam mengklasifikasikan berita olahraga berbahasa Indonesia.

**Kata Kunci :** klasifikasi, text mining , *chisquare*, *TF-IDF* , *K-Nearest Neighbors Decision Rules*

## ABSTRACT

### *PERFORMANCE COMPARISON OF K-NEAREST NEIGHBORS DECISION RULES AND MULTINOMIAL NAÏVE BAYES CLASSIFIER AS A SPORT NEWS CLASSIFICATION*

by

Farhan Ramadhani  
14/363805/PA/15861

The increasing number of news articles that exist on the internet gives new challenges to the editor when categorizing the news. Text mining method is able to categorize a text data automatically using the classification process. However, the problem arising from the text classification process is the size of the dimension of the data that may affect the performance of the classification process itself. So the dimension of the data should be reduced by selecting some attributes that are considered representative. This method is called feature selection.

Feature selection will reduce the dimensionality of the data by ignoring features that are deemed unrepresentative to any label of a data. Several studies shows that feature selection method can improve effectiveness and performance. This research used TF-IDF feature selection method and chisquare. Both feature selection method is continued by the K-Nearest Neighbors Decision Rules algorithm to perform classification. The performance evaluation results of K-Nearest Neighbors Decision Rules algorithm will also be compared by Multinomial Naïve Bayes Classifier by observing each of its accuracy, precision, recall and f-measure.

The results show that the chisquare & TF-IDF feature selection method improves the accuracy, precision, recall & f-measure of K-Nearest Neighbors Decision Rules algorithm. The accuracy of the classification using chisquare increased by 10.68% while using TF-IDF increased 11.32%. Precision improves up to 5.8% and 5.6% using TF-IDF & chisquare respectively. Recall increased 10.50% using TF-IDF and increased 10.7% using chisquare. F-measure parameter increased by 9.9% using chisquare feature selection method, TF-IDF increased f-measure up to 8.7%. Both feature selection methods also improve the running time performance of the K-Nearest Neighbors Decision Rules algorithm. However, the increased evaluation of the parameters of K-Nearest Neighbors Decision Rules has not outperformed the classification performance of Multinomial Naïve Bayes in classifying Indonesian sports news.

**Keyword:** classification, text mining, chisquare, TF-IDF, K-Nearest Neighbors Decision Rules