

ABSTRAK

Penyaluran informasi atau berita sudah mulai menggunakan sistem berbasis komputer untuk menyampaikan berita secara *up to date*. Pada umumnya berita yang disampaikan dalam portal tersebut terdiri dari beberapa kategori seperti berita politik, olahraga, ekonomi dan lain sebagainya. Namun, dalam membagi berita ke dalam kategori-kategori tersebut untuk saat ini masih dilakukan secara manual. Hal ini sangat merepotkan apabila berita yang ingin dikategorikan berjumlah banyak. Oleh karena itu perlu adanya sistem yang bisa mengklasifikasikan berita secara otomatis. Pada penelitian ini, akan dilakukan *text mining* klasifikasi terhadap konten berita tersebut. Secara garis besar penelitian ini dilakukan dalam empat tahapan. Tahap awal adalah proses *preprocessing* data dengan cara *case folding*, *tokenization*, *stopword removal* dan *stemming*. Tahap selanjutnya adalah melakukan seleksi *term* dengan *bag-of-words* dan pembobotan *TF-IDF*. Kemudian dilanjutkan dengan klasifikasi konten berita dengan metode *Naïve Bayes* dan *K-Nearest Neighbor*. Setelah proses klasifikasi dilakukan evaluasi dan analisis performansi sistem dengan menghitung nilai *F1-Measure*. Berdasarkan hasil evaluasi, diketahui bahwa metode *Naïve Bayes* lebih baik dibandingkan dengan metode *K-Nearest Neighbor* dimana nilai *F1-Measure* tertinggi yang dihasilkan metode *Naïve Bayes* adalah 85.94% sedangkan nilai *F1-Measure* tertinggi yang didapat metode *K-Nearest Neighbor* adalah 65.52%.

Kata kunci: *Sistem Klasifikasi, Berita, Text Mining, Preprocessing, Naïve Bayes, K-Nearest Neighbor, F1-Measure*

ABSTRACT

Distribution of information or news have started using computer-based systems to deliver up to date news. In general, the news is delivered in the portal consists of several categories such as politics, sports, economy and so forth. However, the news divide into these categories for the moment is still done manually. This is particularly troublesome when the news you want categorized numerous. Hence the need for a system that can automatically classify news. In this research, text mining classification will be made to the content of the news. Broadly speaking, the study was conducted in four phases. The initial stage is the process of preprocessing of data by folding case, tokenization, stopword removal and stemming. The next stage involves the selection of terms with the bag-of-words and TF-IDF weighting. Then proceed with the classification of news content by the method of Naïve Bayes and K-Nearest Neighbor. After the classification process evaluation and performance analysis system to calculate the value of F1-Measure. Based on the evaluation, it is known that the Naïve Bayes method is better than the K-Nearest Neighbor method in which the highest value of F1-Measure generated Naïve Bayes method was 85.94%, while the highest value of F1-Measure acquired K-Nearest Neighbor method is 65.52%.

Keywords: Classification System, News, Text Mining, preprocessing, Naïve Bayes, K-Nearest Neighbor, F1-Measure