

ABSTRAK

Data mining adalah proses pengumpulan informasi penting dari pola sebuah data yang mana hasil informasi dari data *mining* dapat digunakan dalam berbagai hal yang berhubungan dengan data. Pada penelitian ini, peran data *mining* bertujuan untuk mengidentifikasi kategori dari suatu topik berita dari berbagai jenis kategori. Melalui pengkategorian berita ini, suatu stasiun tv dapat memberi serta menentukan kategori yang tepat pada berita. Tidak hanya pada stasiun tv, pemberitaan juga sering beredar pada koran dan bahkan media sosial, sebagai media alternatif untuk menyampaikan informasi.

Oleh karena itu, dilakukannya proses pengkategorisasian berita berdasarkan kategori yang telah ditetapkan. Dalam penelitian ini, data yang digunakan adalah data yang mempunyai informasi relevan terkait dengan berita, dengan menggunakan data berita tv di Amerika yang terdiri dari 23189 judul berita. Dalam penelitian ini akan melibatkan perbandingan dari dua algoritma yang digunakan dalam klasifikasi, yaitu *Multinomial Naïve Bayes* dan *K-Nearest Neighbor*. Perbandingan akan dilakukan dalam tiap iterasi variasi nilai $k - fold$ dari kedua algoritma ini, yang bertujuan untuk mendapatkan hasil terbaik dari perbandingannya. Pengujian dilakukan dan mendapatkan bahwa algoritma *Multinomial Naïve Bayes* unggul dengan memperoleh nilai akurasi sebesar 97.55% dengan iterasi nilai variasi $k - fold$ sebesar 9.

Kata Kunci : *Data Mining*, Kategorisasi Teks, *Multinomial Naïve Bayes*, *K-Nearest Neighbor*, $K - fold$, Berita.

ABSTRACT

Data mining is the process of collecting important information from the pattern of a data where the results of information from data mining can be used in various matters related to data. In this research, the role of data mining aims to identify the category of a news topic from various types of categories. Through this news categorization, a TV station can provide and determine the right category for the news. Not only on TV stations, news is also often circulated in newspapers and even social media, as an alternative media to convey information.

Therefore, the news categorization process is carried out based on predetermined categories. In this study, the data used is data that has relevant information related to news, using TV news data in America consisting of 23189 news titles. This research will involve a comparison of two algorithms used in classification, namely Multinomial Naïve Bayes and K-Nearest Neighbor. The comparison will be done in each iteration of the variation of the k - fold value of these two algorithms, which aims to get the best results from the comparison. Tests were conducted and found that the Multinomial Naïve Bayes algorithm excelled by obtaining an accuracy value of 97.55% with an iteration of the k-fold variation value of 9.

Keywords : *Data Mining, Text Categorization, Multinomial Naïve Bayes, K-Nearest Neighbor, K-fold, News.*