

## Abstract

During these recent years, World Wide Web has become an important source for information search. Therefore, in line with the growing number of web, the ability to obtain specific and relevant information is as important as the reliability of the web itself to display the information in accordance with their respective class. Basically, the web page is a hypertext. In addition to text and other multimedia components, web also comprises other content in the form of hyperlinks, HTML tags, and meta data. Web page classification becomes not so easy in term of the attributes in HTML document in the form of text that are more prevalent compared with other media components such as audio, video, or image.

Classification problems were solved by applying the multinomial naïve bayes classifier and by utilizing components of web pages that contain the meta tag description and meta tag keywords. Then, from its application, the result of classification accuracy and prediction, also the factors affecting the accuracy of the classification results will be analyzed.

Multinomial bayes classifier works based on the collection of evidence (*evidence*) and class (*class*). By doing the training (*training*) on some sample data, we can determine the probability of certainty (*likelihood probability*) of a proof that if given a certain class. Multinomial bayes classifier is also using the previous probability (*prior probability*) of a class, which calculations can be based from sample data. From analysis of these data samples, if given a new web page consisting of a collection of evidence, the probability of each class to the web page (*posterior probability*) can be determined.

**Keywords:** web page, meta tag, classification, multinomial naïve bayes classifier