

ABSTRACT

Classification is one of the data mining techniques and also text mining used in the process of searching the set of the model or functions explaining or distinguishing data classes and the concept enables to apply this model in carrying out the testing towards the testing data.

Classification is comprised of the data training which attributes is determined as the class for data lines. The result of study is a model called classifier. This model will eventually be used to predict and test the attribute of data testing class.

Today, the process of classification which is often carried out is English text document classification. Therefore, the writer will classify the Indonesian text document. Classification of the text is a process of classification of documents in one or more category defined or in the same document's class. Documents classified in the class defined before is the main aim of classification of the text. Respectively the document could be categorized in many categories, exactly in one category, or even not including in any category.

The classification of Indonesian text document, the writer used two methods or the approach received in information retrieval and machine learning. The two methods are Latent Semantic Indexing and Support Vector Machine. Using the two methods is hoped could facilitate the process of classification especially for the Indonesian text document.

key words: classification, text mining, data mining, machine learning, information retrieval, latent semantic indexing, support vector machine.