A classification paradigm is a data mining framework containing all the concepts extracted from the training dataset to differentiate one class from other classes existed in data. The primary goal of the classification frameworks is to provide a better result in terms of accuracy. However, in most of the cases we can not get better accuracy particularly for huge dataset and
dataset with several groups of data. When a classification framework considers whole dataset for training then the algorithm may become unusable because dataset consists of several group of data. The alternative way of making classification usable is to identify a similar group of data from the whole training data set and then training each group of similar data. In our paper, we first split the training data using k-means clustering and then train each group with Naive Bayes Classification algorithm. In addition, we saved each model to classify sample or unknown or test data. For unknown data, we classify with the best match group/model and attain higher accuracy rate than the conventional Naive Bayes classifier.

Reference


- Networks in the Classification of Training Web Pages, IJCSI International Journal of Computer Science Issues, Vol. 4, No. 1, 2009

Index Terms

Computer Science

Data Mining

Key words

Classification

Naive Bayes

Clustering

classification accuracy