Abstract

The developments of information technology and data electronic modelling quite increasing such as text xml, html, graph , chemical compound and others led to the number of applications to model data as a graph growing rapidly. One of the popular and widely developed is Graph. Graph is very poverful because it can modeled complex structure. The example of application graph are risk assessment, toxic prediction and regulatory decision. Studies on graph-based classification are still lacking on structure chemical compound and for its application is still rare used that more needs to be done using the technique of classification. Graph classification are the one that can be applied to chemical compound.

At this final project study will discussed about method Graph Classification is Graph Boosting with algorithm gSpan and boosting for the chemical molecular classification and counting accuracy are obtained. The goal is to determine and identify whether a chemical compound containing mutagen or not based on classification models are made. This classification model will make a prediction rule with a fewer iterations to get the patterns. This pattern is obtained b enumerating the frequent occurance of subgraph patterns which can be used as a feature in classification. the selection of feature extraction techniques appropriate for build model classification and the design of the right system in order to generate maximum system performance. Chemical molecules chosen for their excelent and uniqueness that has the characteristics of labeled verticle and edges are not directed so that the chemical molecules suitable if represent by a graph. Graph classification method will classify graphs which has the characteristics of structural information as well as using all subgraph are selected as feature set. Results from this study show the efficiency from algoritma gSpan and Boosting for maximum accuracy chemical compound is 78,18 %.

Keywords*: Graph Classification, Frequent Subgraph Mining, Classification, Cheminformatics*