

Abstract

Data mining is the process of semi-automatic to explore and analyze large amount of data to get a useful pattern. Data mining is a process of combination between the fields is mainly machine learning, statistical analysis and database. Data mining seeks to find rules and patterns from data.

One important task in data mining is classification. This classification can be described as follows: the input consists of data that is also known as the *training set* consists of a number of examples (*records*) who each have a number of attributes or features is also called. The purpose of this classification is to analyze the data input and to develop an accurate model for each class based on several variables predictor.

Soft Decision Tree - ID3 (SDT - ID3) algorithm is one of the analyze methods data. SDT - ID3 algorithm find a solution to the decision *tree* is formed based on fuzzy set theory, which has a range of possibilities for some or all classes in the data that will be *tested*.

This final project to analyze performance data for classification SDT - ID3 with breast cancer, and the wine glass data, analysis of the influence of the number of linguistic variables and membership function of the value and performance analysis of the excess, algorithm in the case of SDT - ID3 third classification *datasets* (breast cancer, wine and glass).

Test results obtained from the conclusion that SDT - ID3 algorithm more accurate than the J4.8 algorithm as the *Soft Decision Tree* - ID3 algorithm uses soft discretization form of fuzzy sets which can handle the uncertainty of doubt, and incomplete information.

Keywords : decision *tree*, fuzzy set, soft discretization, accuracy