Abstract

Decision tree is a data mining technique used for the classification and forecasting of the data. It is the supervised learning algorithm that follows the greedy approach and works in a top down manner. Decision tree uses white box model approach and classifies the data in a hierarchical structure. It makes data easy to represent and understand. It can handle a large database and works well with both numerical and categorical variables. A variety of decision tree algorithms are proposed in the literature like ID3 (Iterative Dichotomiser 3), C4.5 (successor of ID3), CART (Classification and Regression tree), CHAID (Chi-squared Automatic Interaction Detector). These algorithms have specific mechanisms based on certain criteria’s. The study of these criteria are important and requisite for analysis of DT algorithms. The aim of this paper is to identify and inspect these vital criteria’s or factors of DT algorithms. The major contribution of this review is to provide a path to select a specific factor for improvement of DT algorithm as per requirement or problem.

References


4. Lior Rokach and Oded Maimon, DECISION TREES, Chapter 9, online source: http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf

5. http://physics.about.com/od/glossary/g/entropy.htm


8. Lior Rokach and Oded Maimon, DECISION TREES, Chapter 9, Data Mining and Knowledge Discovery Handbook online source: http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf


11. S. Santhosh Kumar, Dr. E. Ramaraj, Modified C4.5 Algorithm with Improved Information Entropy and Gain Ratio International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 9, September – 2013

Index Terms

Computer Science Algorithms

Keywords

Data Mining, Decision Tree Technique, Decision Tree Methodology, Decision Tree Factors.