

Data Mining Approach: Classification Model Evaluation

Authors : Lubabatu Sada Sodangi

Abstract : The rapid growth in exchange and accessibility of information via the internet makes many organisations acquire data on their own operation. The aim of data mining is to analyse the different behaviour of a dataset using observation. Although, the subset of the dataset being analysed may not display all the behaviours and relationships of the entire data and, therefore, may not represent other parts that exist in the dataset. There is a range of techniques used in data mining to determine the hidden or unknown information in datasets. In this paper, the performance of two algorithms Chi-Square Automatic Interaction Detection (CHAID) and multilayer perceptron (MLP) would be matched using an Adult dataset to find out the percentage of an/the adults that earn > 50k and those that earn ≤ 50k per year. The two algorithms were studied and compared using IBM SPSS statistics software. The result for CHAID shows that the most important predictors are relationship and education. The algorithm shows that those are married (husband) and have qualification: Bachelor, Masters, Doctorate or Prof-school whose their age is > 41<57 earn > 50k. Also, multilayer perceptron displays marital status and capital gain as the most important predictors of the income. It also shows that individuals that their capital gain is less than 6,849 and are single, separated or widow, earn ≤ 50K, whereas individuals with their capital gain is > 6,849, work > 35 hrs/wk, and > 27yrs their income will be > 50k. By comparing the two algorithms, it is observed that both algorithms are reliable but there is strong reliability in CHAID which clearly shows that relation and education contribute to the prediction as displayed in the data visualisation.

Keywords : data mining, CHAID, multi-layer perceptron, SPSS, Adult dataset

Conference Title : ICITCS 2016 : International Conference on Information Technology and Computer Science

Conference Location : Istanbul, Türkiye

Conference Dates : July 21-22, 2016