Abstract

Data mining means to find out some useful information from a big warehouse of data and the process is aimed at unfolding old records and identifying novel patterns from the data. Data mining is used for classification and prediction. Many techniques and algorithms are available for mining the data. Out of many techniques, the decision tree is the simplest. This paper focuses on comparing the performance accuracy of ID3 and C4.5 techniques of the decision tree for predicting customer churn using WEKA. The data used for this research work has been collected by designing a survey form and getting it filled by around 150 mobile phone users belonging to a different gender, age groups and having different types of connection providers. For the data analysis in WEKA, the cross-validation method is used where a number of folds n (10 as standard as per the software) is used. From the results, it is observed that C4.5 algorithm exhibits better performance than ID3.

References


Index Terms

Computer Science Information Sciences

Keywords

Data mining, Decision tree, ID3, C4.5