5th International Conference of Multidisciplinary Approaches (iCMA), 2018
Faculty of Graduate Studies,
University of Sri Jayewardenepura,
Sri Lanka
ISSN: 2386 – 1509

A SYSTEMATIC LITERATURE REVIEW OF DATA MINING IN FIELD OF EDUCATION

Herath D.S.* and Mathotaarachchi. Y.T.
University of Colombo School of Computing, Sri Lanka
dherath10@gmail.com

ABSTRACT

This paper aims to understand the different data mining techniques upon the educational data. We investigated the literature on case studies carried out in the area throughout the past ten years (2009-2018). Search terms identified 67 pieces of research work, but insertion criteria restricted the key studies to 46. We analyzed the year of publication, type of the publication, the learning environment, database searched, methods, algorithms, tools, topic, the dataset used, keywords used, respective educational outcomes and findings of these case studies. In this investigation, we have identified methods and algorithms with some experimental data. The most popular method is classification, followed by clustering, association mining and the widely used algorithm is K-Means (Clustering) followed by C4.5 (J48) (Classification), ID3 (Classification), apriori (Association), Naive Bayes (Classification). The most of the researchers focused on analyzing learner behavior/learner behavior modeling (n=16) area. But there were significant contributions towards prediction learner's performance (n=15) and outcomes, identifying learner's risk and drop-out (n=3) areas. Those areas, the prediction was an important method. Therefore classification has become the most common method that researchers used. The most number of the recommended algorithm was K-Means. It is a clustering algorithm. It was a free-dominated one in this review. Because a less number of other clustering algorithms were used by the researchers such as Markov Clustering (MCL), Expectation Maximization (EM), and Fuzzy C-means. We have discovered another factor, that is, WEKA it is mostly used data mining tool. It is an open source tool which provides a group of many data mining and machine learning algorithms. Finally, we have concluded educational data mining helps to investigate students' learning behaviors and the settings which they learn in. We expect to deliver a better influence towards the development of a quality learning environment for both students and teachers and some insights are outlined for further research.

Keywords: Educational Data Mining, Data Mining, e-Learning, Learning behaviors, Clustering