

Global Colloquium in Recent Advancement and Effectual Researches in Engineering, Science and Technology (RAEREST 2016)

Student academic performance prediction model using decision tree and fuzzy genetic algorithm

Hashmia Hamsa*, Simi Indiradevi, Jubilant J. Kizhakkethottam

Department of Computer Science and Engineering, Musaliar College of Engineering and Technology, Pathanamthitta, 689653, India

Abstract

The research on the educational field that involves Data Mining techniques is rapidly increasing. Applying Data Mining techniques in an educational background are known as Educational Data Mining that aims to discover hidden knowledge and patterns about student's performance. This work aims to develop student's academic performance prediction model, for the Bachelor and Master degree students in Computer Science and Electronics and Communication streams using two selected classification methods; Decision Tree and Fuzzy Genetic Algorithm. Parameters like internal marks, sessional marks and admission score were selected to conduct this work. Internal marks are the combination of attendance marks, average marks obtained from two sessional exams and assignment marks. Admission score for degree students is the weighted score obtained from 10th and 12th examination marks and entrance marks. In the case of master's degree students, it includes degree examination marks and entrance marks. Resultant prediction model can be used to identify student's performance for each subject. Thereby, the lecturers can classify students and take an early action to improve their performance. Systematic approaches can be taken to improve the performance with time. Due to early prediction and solutions are done, better results can be expected in final exams. Students can view their academic information and updates. Reputed companies having a tie-up with the institution can search students according to their requirements.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the organizing committee of RAEREST 2016

Keywords: Educational Data Mining; Classification; Prediction; Decision Tree; Genetic Algorithm; Fuzzy Logic

* Corresponding author

E-mail address: hashmia.hamsa@gmail.com

1. Introduction

Education is a very important issue regarding the development of a country [1]. The main objective of higher education institutions is to present quality education to its students. One way to accomplish the higher level of quality in higher education scheme is by predicting student's academic performance and thereby taking early actions to improve student's performance and teaching quality. The relevant knowledge is hidden with the educational data set and it is extractable during data mining techniques. The present paper is planned to validate the capabilities of data mining techniques in the background of higher education by offering a data mining model. In this work, classification task is used to evaluate student's performance [2]. Constant evaluations of student's performance have to be done for each subject. So that the exact area, the student lose his/her marks can be identified. That helps the lecturer to take necessary actions like more attention to the student for that particular subject, teaching in a different way that he/she can grasp quickly, conducting exams, etc. Finally, that improves student's calibre and academic status. The application of Data Mining in the Educational framework is referred as Educational Data Mining (EDM).

The appliance of analytics in the educational background has increased in the last decade. Ferguson presents in [3] three drivers for this to arise: primarily, the volume of data that are composed of educational institutions have seriously improved; secondary driver is the use of e-learning: it has helped to collect data, still brought some learning issues such as possible lack of motivation and difficulties for the educators to collect direct feedback regarding the mood, level of interest or even the understanding of the students; after all, the political concerns: countries are getting a superior understanding of the significance of higher education in their development and government have an attention in improving it, to propose enhanced learning opportunities that direct to better academic results.

In this work, two data mining approaches are proposed to predict student's performance. Prediction is done using two algorithms: Decision Tree (DT) and Fuzzy Genetic Algorithm (FGA). Prediction is carried out with academic records along with initial academic information.

The rest of the text is ordered as follows: section II present background study and related work; section III present the data sets and preparation; the proposed data mining model is presented in section IV; section V describe the experimental setup; results and analysis are presented in section VI; finally, the conclusions.

2. Background Study and Related Work

The study in [4] identifies that Bayes classifier performance result was decreased by adding more academic records. This may be caused by the assumption of independence required by the algorithm. For getting accuracy in prediction, selected attributes should be relevant and noise free. By adding more relevant attributes the accuracy can be increased. Related studies are:

2.1. Classification Methods

In a classification task, the objective is to assign a predefined label or class to a record based on a set of known attributes. An important feature of a classification model is that it is built by part of the data, also known as the training set, which is used to train the model. All the attributes in the subset are known, even the class. After the model is built, it is used to assign the label to new records where the class attribute is unknown. To build the models, two techniques are used: DT and FGA.

A DT is a representation made of nodes and arcs where an internal node present a decision based on attribute values and the arcs stand for the option made in the node. It ends on a leaf node, which represents the label or the class to be assigned. To categorize a record with DT, it starts from the root node and goes one level down at a time that depends on the results of the condition tested on every node; when it ends on a leaf node, the record is classified according to the label on the leaf node. The knowledge represented by DT be extracted and presented in the form of IF-THEN rules [5].

Fuzzy Logic (FL) addresses applications that resemble human decision making with a skill to generate exact solutions from certain or approximate information. The use of FL based techniques is for either improving GA behaviour or modelling GA components; the results obtained have been called FGA. A fuzzy fitness finding (FFF) mechanism guides the GA through search space by combining various criteria/features that have been identified as

Download English Version:

<https://daneshyari.com/en/article/4962543>

Download Persian Version:

<https://daneshyari.com/article/4962543>

[Daneshyari.com](https://daneshyari.com)