A REVIEW ON EDUCATIONAL DATA MINING TECHNIQUES AND RECOMMENDATION MODEL IN ANALYZING STUDENT’S PERFORMANCE

R. Aswin Herbert Sathish, Mary Jacob, Riju K Sajee

1. Department of Computer Science, Kristu Jayanti College, Bangalore, India
2. Department of Computer Science, Kristu Jayanti College, Bangalore, India
3. BCA Student, Department of Computer Science, Kristu Jayanti College, Bangalore India

ABSTRACT:

The Mining of education data is an emerging trend in the learning analytics as it is time-consuming to analyze the data and to identify the hidden information automatically. In this paper, a detailed investigation of educational data mining technique is carried out. The application of the data mining includes storage and retrieval of student data in large repositories such as mark sheets, attendance sheets, student profiles, etc. The importance analysis is carried out on the retrieval of large data using machine learning algorithm in data mining. Along the retrieval of the data, these days, deep focus is made on prediction and recommendation models which provide more effectiveness to educational applications in terms of suggesting and extracting the correlation among students. However handling of large data from repositories leads to performance bottleneck, hence it is resolved by employing Map Reduce Paradigm from big data analytics. Through extensive study, classification and clustering provide more value for the management, hence Semantic and Opinion Mining is presented as the future research solution.

Keywords: Educational data mining, Clustering, Classification, Data Prediction, Recommendation, Map Reduce, Semantic and Opinion Mining

[1] INTRODUCTION
Educational Data Mining (EDM) is a field which uses data mining to extract data from large repositories using knowledge discovery model to generate information [1]. It is also used to determine patterns in data, organize the information embedded in hidden relationships, estimate the unknown labels, predict the value to classify the object and compose a new cluster to the homogenous objects [2]. The Major focus of the EDM is made on knowledge discovery, decision making, and recommendations. EDM emerges as learning paradigm algorithms for exploring data to find out patterns and make predictions.

The objective of the EDM is to develop models which would upgrade learning experience and institutional effectiveness. One of the preprocessing algorithms of EDM is known as Clustering. It is an unsupervised approach for analyzing data [3]. It refers to collecting similar objects together to form a group or cluster. Each cluster contains objects that are similar to each other but dissimilar to objects of other groups [4]. Data clustering enables academicians to predict student performance, correlate behaviors and learning styles of different types of students thereby collectively improving on an overall institutional performance. Though handling of large data from repositories enhances analysis, it leads to performance bottleneck which would require the need for big data solutions, which is resolved by employing Map Reduce Paradigm from big data analytics [5]. The predictive and recommendation algorithms work on grade and enrollment data to suggest rationally auspicious courses to a student based on their analytically predicted interest [6]. The rest of the paper is organized as follows - Section 2 describes the review of literature related to study, it then continues to provide the future solutions in Section 3 and lastly Section 4 provides the conclusion.

Fig I [EDM]
2.1 EDUCATIONAL DATA MINING AND ITS ROLE IN DETERMINING FACTORS AFFECTING STUDENTS ACADEMIC PERFORMANCE

In this literature, various supervised data mining techniques used in problem-solving and decision-making skills of individuals. The Supervised algorithms are to improve the overall quality and effectiveness of education. Clustering techniques have broadly been classified into two types, hierarchical and partitional. Techniques like Centroid-based clustering, Graph-based clustering, Grid-based clustering, Density-based clustering, Neural network-based clustering, etc can be used to predict the student’s characteristics on various aspects. Clustering algorithms are also applied to voluminous data sizes of the student data if it exceeds the particular limit. The notion of voluminous, enormous quantities of data is referred to as Big Data [7].

2.2 CLUSTERING METHODS IN EDUCATIONAL DATA MINING

This literature discusses various clustering techniques using hierarchical and non-hierarchical methods to differentiate the groups of students, according to their interaction and performance characteristics. There are different learning models to perceive the groups obtained, to determine a similarity between the results, confirming the acquired knowledge from the clustering and demonstrating that the choice of method on the knowledge obtained from interactions and students performance on the course is reliable [8].

2.3 A DATA PREDICTION MODEL TO PREDICT SCHOOL FAILURES AND DROPOUT

This literature applies data prediction algorithms to predict school failures and drop outs. Decision Tree Algorithms have been employed to predict if a student has the possibility to fail in Central Examinations and drop out due to an on performing attitude in their studies. High dimensional data is taken for processing; in this feature reduction algorithm is also employed to remove the unwanted attributes in the dataset, after which class label employed to that category of data. Data normalization is also employed on the feature which reduces attributes in the resultant dataset [9].

2.4 MUTUAL REINFORCEMENT OF PREDICTION AND RECOMMENDATION ON ACADEMIC PERFORMANCE

This literature identifies the importance of prediction and recommendation in academic performance. It is one of the most important tasks in educational data mining and has been widely used in intelligent tutoring systems. Academic performance could be affected by factors like personality, skills, social environment and the use of library books and so on (Fig II). The supervised content-aware matrix factorization for mutual reinforcement of academic performance prediction and library book recommendation is analyzed. This model not only addresses the
sparsity challenge by employing dimension reduction techniques but also promotes recommendation by frequently using an item mining technique named Apriori algorithm for students based on their performance levels and book meta information [9].

**Fig II [Factors Influencing Student’s Performance in Academics]**

3 FUTURE SOLUTIONS

Enabling the semantic and opinion mining on the current learning model improves the adaptive learning system. Education content representation as semantics provides better retrieval efficiency. The semantic can be presented in terms of conceptual metadata representation. Automated acquisition of hierarchical relationships between relevant domain terms from educational content, which constitutes a fundamental step in the semantic composition of an educational course. It not only improves the speed of the domain model creation but also reduces the overall difficulty of the task, and finally, it has significant improvement in the quality of resulting domain models.

4 CONCLUSION

This paper has presented the detailed review of the Educational data mining carried out in the areas of classification, clustering, prediction and recommendation models. In Summary, a key advantage of clustering provides the grouping of heterogeneous objects using the learning models. The Classification algorithms are used to categorize the objects of the cluster. Prediction and recommendation algorithms are used to suggest the students with future outcomes in order to improve overall performance and enhance the student retention effort. This paper also
highlights the future solutions of EDM in terms of employing semantic mining and opinion mining to improve the performance of present solutions.

REFERENCES
[9].Carlos Marquez-Vera et.al "Predicting School Failure and Dropout by Using Data Mining Techniques" in IEEE Latin American journal of learning technologies, vol: 8, issue: 1, Feb2013
[10]. Antonio Garrido et.al " E-Learning and Intelligent Planning: Improving Content Personalization" in IEEE Ibero