



A LITERATURE SURVEY ON ANALYSIS OF STUDENT ACADEMIC RESULTS USING ROUGH K-MEANS

Computer Science

Teena A. Dorkhande

Mtech Scholar, PIET, Nagpur, India

Leena Patil

Associate Prof. (CSE), PIET, Nagpur, India

ABSTRACT

Educational data mining is mainly care with developing new methods to discover knowledge from educational database for analysing and enhancing the educational organization. One of the biggest challenges is to improve the quality of the educational processes so as to enhance student's performance. In order to analyse student academic results towards education an attempt to improve record of student. In the existing research the researchers have used classification and clustering data mining technique.

This paper surveys an educational data mining in education system & also previous result predict and analysing base on attributes and condition using Rough K-Means algorithm and WEKA tool. As we know large amount of data is stored in educational database, so in order to get required data to Rough K-means clustering techniques are develop & use. In this research we use dataset as final year Student of Computer Sciences Engineering, PIET College, Nagpur. Using this college can help how many academic students are distinctions, fail and pass. Also can be predicting result of each student base on previous result of those students before get final exam to be failed or low marks.

KEYWORDS

Educational data mining, Clustering, Classification, Rough set Theory, K-means, Rough K- means, WEKA tool

INTRODUCTION

Data Mining (Knowledge Discovery in databases) is the process of extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) pattern or information from large databases using various data mining techniques such as classification, clustering, association rule etc. which helps in various decision making[1].

Large data is mining in educational environment. This process is called Educational Data Mining. The data mining process is to extract information from huge datasets and transform it into understandable structure for further use. Data mining techniques which extract information from huge amount of data have been becoming popular in education domains. Educational Data Mining is a learning analytics and quantitative observation method in order to understand how student respond to educational system and their responses impact their learning. Its objective is to analyze educational data in order to resolve educational research issues.

In recent years there is rapid growth in education sector which leads to growing of education data so mining of education data become important to understand student behavior during learning process and take appropriate action to improve the student performance and the quality of the courses delivery[2].

The techniques of grouping data of an object are similar, both to deal with certain conditions and in conditions of uncertainty, incompleteness of knowledge, especially with high complexity, speeds and stability will be a problem in itself. To deal with such uncertainty problems, researchers have proposed several methods such as fuzzy theory and rough set theory [3]. Obtained clusters and applied queries do not necessarily have boundaries.

Rough set theory is a powerful mathematical tool for dealing with vagueness, uncertainty, and incompleteness of knowledge in information systems. Rough set theory is not verifying all properties. Main idea of that property is separate the distinguishable items from the indistinguishable ones. Also another soft computing k-means algorithm based on rough set recently introduced. It has describes the core concept of rough sets, the lower and upper approximation of a set that can be used in clustering. This algorithm called as rough k-means algorithm. This paper presents analysis of student academic result based rough sets to incomplete information systems. These propose system can be based on rough K-means clustering technique.

LITERATURE SURVEY

In this section we review some recent research works regarding the use of classification technique of data mining for predicting student result and analysis in educational domain.

R. R. Kabra and R. S. Bichkar December 2011 in that student performance predicted. They used decision tree algorithm technique to predict student performance. They used the data of around 340 students of engineering field to predict their performance in their first year exam. They generated pattern using decision algorithm to create a training model for prediction. Accuracy of model generated was 60% .they used confusion matrix to check whether how many student will pass or how many will fail.

Parneet Kaura ,Manpreet Singhb and Gurpreet Singh Josanc in 2015 it used data as 164 high school student to predicted student result using data mining techniques as well as slow learning students present in that group of students. They used WEKA tool to predict result and to show comparison of various techniques to predict result.

L. Romdhae, N. Fadhel and B. Ayeb in 2016 showed that to predict student success rate used data mining techniques based on both regression and classification can be applied. The authors first showed prediction result using regression analysis and Bayes method. Later on he also showed the use of decision tree under j48 method of classification to predict result and they also studied about how clustering technique can be used to predict result.

S. Aher and L.M.R.J. Lobo in 2011 used classification and clustering technique of data mining to predict student result. The author used ZeroR algorithm of classification technique to predict result and also predicted result using DBSCAN algorithm of clustering of data mining.

How data mining algorithms can be used to predict student result. Review as various research papers from 2002 to 2016 and explained their basis which techniques are mostly used to predict result.

Table 1: Comparison of Classification Method
Classification Methods

Classification Methods	Accuracy Rate
DT-J48	92%
Neural Network	91%
Decision Tree	94%
Support Vector Machine	95%
K-Nearest Neighbour	85%
Naïve Bayes	75%

EXISTING SYSTEM

Educational data mining has vast amount of data that has to be organised in a consistent manner. To organise, analyse and classify students detail. The existing system model is designed by collecting Students academic result data from college. Thereby grouping the

student's result based on certain conditions. This system is designing using J48 algorithm provide a maximum accuracy in classifying the instances in an efficient way.

Disadvantages

1. In Classification techniques, model is built earlier for the analysis and then it applies the algorithm to the data in order to estimate the parameters for the model.
2. This technique is time consuming. Its take a lot of time in training the dataset and rest time take as testing.

PROPOSED SYSTEM

To improve the existing system is based on classification techniques we propose a new system as to analyse of student academic results. To replace J48 algorithm of classification to Rough K-Means algorithm of data mining clustering. By doing this improve maximum accuracy in clustering the instances in an efficient way.

Flowchart of the System

We consider the academic result of student of UG Computer sciences engineering of our institute for result analysis. There are consider six previous exam results (first semester, second semester, third semester, four semester, fifth semester, six semester). We consider attributes like Percentage (CGPA) Or Grads. Thereby grouping the student's performance based on certain conditions as:

- Distinction
- First class
- Second class
- ATKT

Base on this condition we analysed results as how many student are belong to particular type of condition. Using all these attributes and conditions a pattern will be arranged or information will be extracted using data mining on previous result by training that information under clustering technique of data mining and will be stored.

Student analysis system will be proposed for the predict the outcome of the student based on the architecture given below in Figure 1. This system provides an efficient analysis on academic results by data collection.

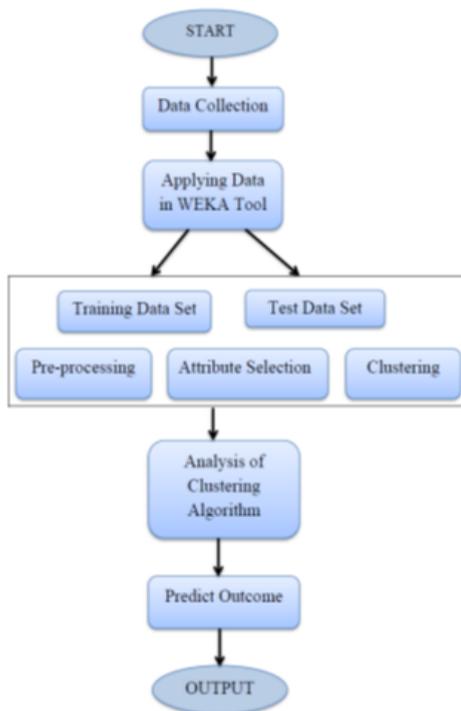


Fig.1:- Flowchart of Proposed System

Advantages

1. Improve the speed of predicting the result.

2. Reduce the time in training, it greatly improves speed. The step taken to reduce the training dataset may be by using clustering techniques.
3. To improve the quality of student's performance to apply the algorithm directly to the dataset to detect its result and later a model can be created on the basis of the obtained results using clustering techniques.
4. An enhanced work of proposed system the real time data from colleges or from universities can be used for the betterment of effective result.
5. This system can help those students before final exam whose result has been predicted fail or low marks.

METHODOLOGY

A. Data Selection

In this paper dataset has been collected from an Engineering College, PIET College Nagpur, India. The dataset contains different attributes and instances. Dataset can be analyses using a data mining technique namely rough k means. The programming environment use for application Java Programming language for building data mining model it was compatible with SQL, in which data was stored. Before processing of data we will be going to clean the data to remove noise and inconsistency. To remove missing values in the dataset, we will use the cleaning techniques. The experiments and observations will be carried out by using data mining tool i.e. WEKA.

B. Rough set theory

Rough set theory is new mathematical tool in data mining. This algorithm is applied on lager database. Rough set is used to deal with uncertainty, vague, insufficient information. When we have insufficient knowledge to precisely define clusters as sets, we use rough sets; here, a cluster is represented by a rough set based on a lower approximation and an upper approximation [11].

Some of the basic properties of rough sets are:

1. For a set A and object u, if $u \in$ lower approximation (Ai), then $u \in$ upper approximation (Ai).
2. An object u can be part of at most one lower approximation.
3. If an object u not part of any lower approximation then u belongs to two or more upper approximations.

C. Tool Selection –WEKA

Waikato Environment for Knowledge Analysis (WEKA) is data mining tool used for data mining. WEKA is a tool based on JAVA language. After applying data in WEKA tool. The clustering panel enables the user to apply clustering algorithms to the resulting dataset, to estimate the accuracy of the resulting predictive model, and to visualize erroneous predictions, or the model itself.

D. Rough K-Means Algorithm

The rough K-Means Algorithm is used in this paper. Traditional k-means clustering partitions a group of objects into a number of non-overlapping sets. It's find out k-means clustering results often have a vague, rough distinguish unclear boundary in data mining. After that combines k-means clustering algorithm with rough set theory, each cluster is seen as a rough set, each object is either determined to belong to the lower approximation set of one cluster, or belong to the upper approximation sets of multiple clusters.

Rough K-means algorithm has shown that it can provides a reasonable set of lower and upper bounds for a given dataset. With the conception of the lower and upper approximate sets, rough k-means clustering become valid algorithms in vague information clustering. Rough k-means clustering algorithm improves the accuracy of clustering boundary. And traditional rough k-means algorithm only considers the indiscernibility of boundary objects and unified weights of multiple objects, ignoring the differences of the objects in the cluster.

The Rough K-means algorithm provides a rough set hypothesis to the conventional K-means algorithm to deal with vague, uncertainty and insufficient knowledge involved in cluster analysis [12].

The Rough K-means algorithm can be stated as follows:

1. Select cluster in initially for u objects into n clusters.
2. Each u object assign to the Lower bound (L(x)) or upper bound (U(x)) of cluster/ clusters.
3. For each cluster Ai re-compute centre according to the equations

the weighted combination of the data points in its lower_bound and upper_bound.

CONCLUSION

Rough K-Means algorithm and WEKA tool has been used as attribute for the selection of a college student academic result. Data collected from the CSE department, PIET College. This algorithm is one of the clustering algorithms used in the large sets data. It will find out the result analysis and accuracy rate of the clustering. In this paper we are using Rough K-Means algorithm, which improves the speed of predicting analysis of student results. This system can help those students before final exam whose result has been predicted fail or low marks.

REFERENCES

- [1] Cristobal Romero, Member, IEEE, and Sebastian Ventura, Senior Member, IEEE, "Educational Data Mining: A Review of the State of the Art" NOVEMBER 2010.
- [2] N. Upadhyay, V. Katiyar, "A Survey On The Classification Techniques In Educational Data Mining", International Journal Of Database Management System (IJDMMS), 3(11), 2014.
- [3] Suhirman, Jasni Mohamad Zain, "Applying Rough Set Theory For Student Clustering On Assessment Datasets", 2013.
- [4] S. Aher, L.M.R.J. Lobo, "Data mining in educational system using weka", International Conference on Emerging Technology Trends, 2011.
- [5] Z. J. Kovacic, "Early prediction of student success: Mining student enrollment data", Proceedings of Informing Science & IT Education Conference
- [6] Parvesh Kumar and Shri Krishan Wasan "Comparative Study of K-Means, Pam and Rough K-Means Algorithms Using Cancer Datasets", International Symposium on Computing, Communication, and Control (ISCCC 2009).
- [7] P. Lingras and C. West, "Rough set clustering for Web mining", IEEE International Conference on Fuzzy Systems, vol. 2, 2002.
- [8] M. Mayilvaganan, D. Kalpanadevi, "Comparison of classification techniques for predicting the performance of students academic environment", In Communication and Network Technologies (ICCNT), International Conference on, IEEE, 2014, pp. 113–118.
- [9] Hidayah, A. E. Permasari, N. Ratwastuti, "Student classification for academic performance prediction using neuron fuzzy in a conventional classroom", Information Technology and Electrical Engineering (ICTEE), International Conference on, IEEE, 2013, pp. 221–225.
- [10] Parvesh Kumar and Siri Krishan Wasan "Comparative Study of K-Means, Pam and Rough K-Means Algorithms Using Cancer Datasets", 2009 International Symposium on Computing, Communication, and Control (ISCCC 2009).