ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

WEKA AS A DATA MINING TOOL TO ANALYZE STUDENTS' ACADEMIC PERFORMANCES USING NAÏVE BAYES CLASSIFIER- A SURVEY

Karan Manchandia*, Navdeep Khare, Mohit Agrawal

DOI: 10.5281/zenodo.438104

ABSTRACT

In Indian Education System, the student performance evaluation is done by faculty manually. This System of student performance evaluation is non-transparent and often leads to dissatisfaction of student. This project aims to solve this problem by designing a user interface which would work on learning using Naïve Bayes Classifier.In evaluating the marks of students by faculty, many times there is partiality done by faculty while giving marks to the students. Therefore to cease this problem the concept of data mining is introduced.

Data mining techniques are widely used in educational field to find new hidden patterns from student's data. The hidden patterns that are discovered can be used to understand the problem arise in the educational field. Data Mining (DM), or Knowledge Discovery in Databases (KDD), is an approach to discover useful information from large amount of data. DM techniques apply various methods in order to discover and extract patterns from stored data. The pattern found will be used to solve a number of problems occurred in many fields such as education, economic, business, statistics, medicine, and sport. The large volume of data stored in those areas demands for DM approach because the resulting analysis is much more precise and accurate.

KEYWORDS: WEKA, education, KDD, Neural Network, performance, evaluation.

INTRODUCTION

In Indian Education System, the student performance evaluation is done by faculty manually. This System of student performance evaluation is non-transparent and often leads to dissatisfaction of student. This project aims to solve this problem by designing a user interface which would work on learning using Naïve Bayes Classifier.In evaluating the marks of students by faculty, many times there is partiality done by faculty while giving marks to the students. Therefore to cease this problem the concept of data mining is introduced.

Data Mining techniques are widely used in educational field to find new hidden patterns from student's data. The hidden patterns that are discovered can be used to understand the problem arise in the educational field. Data Mining (DM), or Knowledge Discovery in Databases (KDD), is an approach to discover useful information from large amount of data. DM techniques apply various methods in order to discover and extract patterns from stored data. The pattern found will be used to solve a number of problems occurred in many fields such as education, economic, business, statistics, medicine, and sport. The large volume of data stored in those areas demands for DM approach because the resulting analysis is much more precise and accurate.

DATA MINING TOOLS

DM is a powerful technology invented with great ability to help organizations focus on the most important information in their data center. DM tools can predict future trends and behaviors and answer the questions that usually consume too much time to resolve. DM tools are among the variety of analytical tools that used for data analysis. It allows users to analyze the data from many different aspects, categorize it, and summarize the identified relationships. Currently, many DM tools are available for every one and different usage for research such as the WEKA, KNIME, Orange, SPSS Clemantine, MATLAB, and NeuroShell etc. These tools provide a set of DM methods and algorithms that help in better implementation of data and information available to users; NBC, Decision Trees, Neural Networks, Rule Learning, Association Rule etc. The available DM tools can be divided into 2 types which are open source/non-commercial software and commercial software. These types of tools have their own strengths and weaknesses in terms of data types and the application methods.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

WEKA:
Waikato Environment for Knowledge Analysis (WEKA) is developed on the Java platform that contain collection of machine learning and DM algorithms that widely used for data classification, clustering, association rule, and evaluation. The WEKA tool provides the interface that allows user to apply the DM methods directly to the dataset or user can embed their own programming Java code on WEKA to suit with their project. This tool also supports the variety file formats for mining include ARFF, CSV, LibSVM, and C4.5.

PROBLEM STATEMENT

In different educational institutions a huge amount of data is generated for the evaluation of student's performance. The data for each student is needed to be analyzed separately for his/her performance evaluation by the faculty. This type of evaluation takes much time.The process of student evaluation by the faculty is manual, also there is chances of partial evaluation by the faculty.
There are some research papers written on Student Performance Evaluation Using Data Mining. These are:

- Brijesh Kumar Baradwaj, Research Scholor, Singhaniya University, Rajasthan, India
  Mining Education Data to Analyse student's Performance, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011 [1]Present paper is designed to justify the capabilities of data mining techniques in context of higher education by offering a data mining model for higher education system in the university. In this research, the classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree method is used here.
- Firstman Noah, BaahBarida, Taylor Onate Egerton, Evaluation of Student Performance Using Data Mining Over a Given Data Space, International Journal of Recent Technology and Engineering (IJRTE)
  ISSN: 2277-3878, Volume-2, Issue-4, September 2013[2]
  In this paper, the performance evaluation of students, were presented using data mining technique and cluster checking. The system examined students who gained admission into the University of Port-Harcourt through the University Matriculation Examination (UME) and through Basic studies programme with the aim of finding out variations in their performance when they graduate from the university. The evaluation was done using data mining technique to find out the ratio that falls into grouping of the grading in the various classes using the cumulative grade point average (CGPA) and the students who failed out.
- AshwaniKharola, SwarnimaKunwar, Gopa B Choudhury, Institute of Technology Management, Defence Research & Development Organisation, Ministry of Defence, Government of India, Students Performance Evaluation: A fuzzy logic reasoning approach
  PM World Journal Vol. IV, Issue IX – September 2015 [3]
  This paper presents a new fuzzy logic reasoning based approach for performance evaluation of students in school or college. The attributes considered for evaluation cover academic as well as personality traits of the students. A Stage-wise fuzzy reasoning approach has been used to eliminate the issues of rule explosion. The comparison between fuzzy and traditional average technique shows the advantage of weightage allocation in fuzzy approach.
- Vaneet Kumar, Dr. Vinod Sharma, Student's Examination Result Mining: A Predictive Approach, International Journal of Scientific & Engineering Research, Volume 3, Issue 11, November2012. [4]
  This paper takes into consideration the various factors and their influences on student performance in education and predicts their final examination result whether pass or reappear. Various factors such as previous year results, attendance, financial status of family, parental education qualification, internet use for study material, parents visit to school, extra class etc. play an important role in education. This paper discusses the most common measurable factors among students. The result of this paper present an idea about these factors and their influence on student performance. MATLAB is used to predict outcome which helps to predict final examination result.

The following figure shows the process flow of existing system:

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

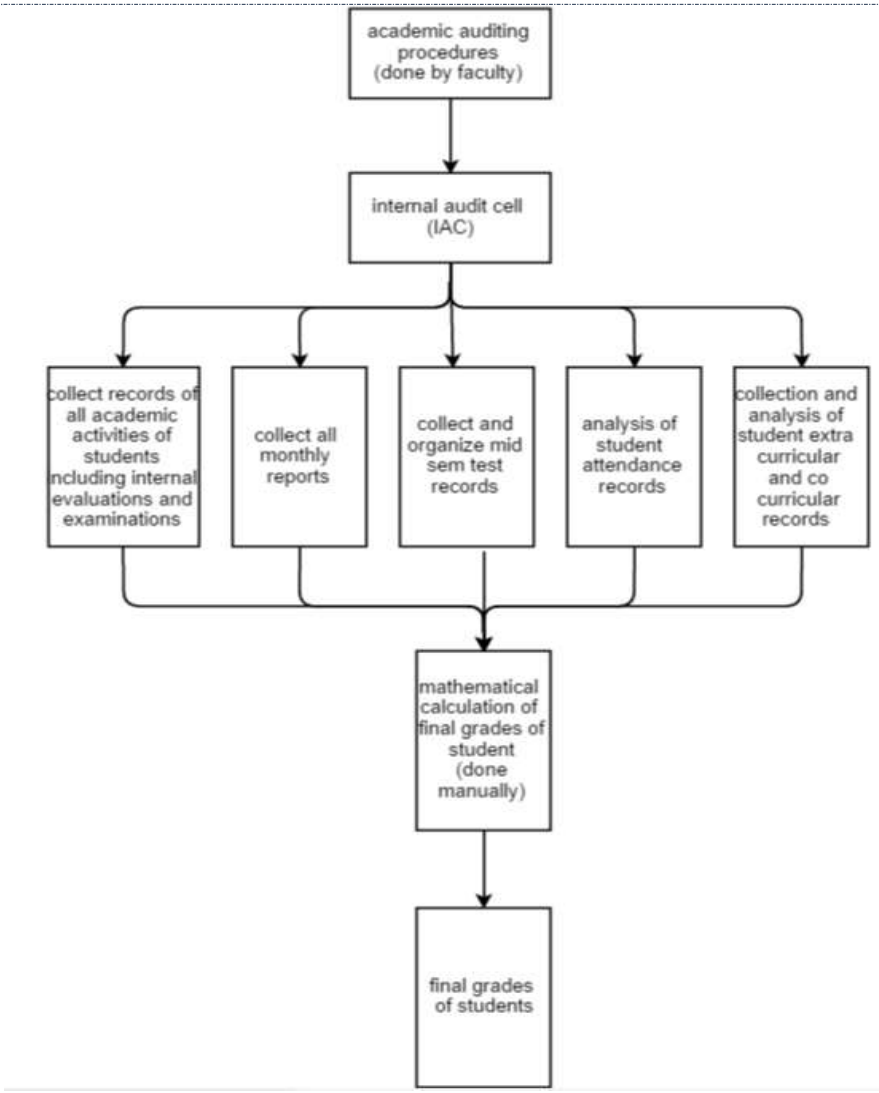

Figure.1:Process flow diagram of the existing system

The diagram shown in Fig.1 shows the process flow to find the final grades of the student as per existing system.

PROPOSED SYSTEM (SOLUTION)

Project features for the project Implementing WEKA as a Data Mining Tool to Analyze Students' Academic Performances using Naïve Bayes Classifierare:

- Understanding the Process of collecting Large Data Sets.
- Learning and Using Data as a Data Mining Tool.
- Performing Transformation of large data sets into normalized form using My SQL, PHP statements.
- Understanding and using the concept of Naïve Bayes Classifier.
- Pattern extraction ofStudent Academic Data.
- The faculty as well as the admin can view final mining result.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Process Flow:

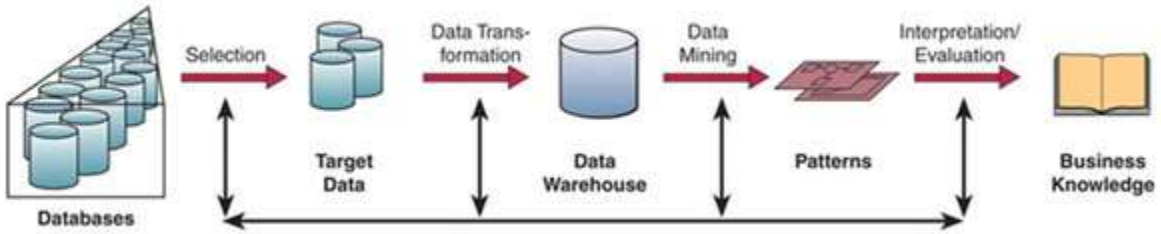

Figure 2: Process flow diagram for the proposed system

CONCLUSION & FUTURE ENHANCEMENT

The proposed system will help in Analyzing Students Academic Performance and it will be very useful for the user (faculty,student) to analyze students in a better way.Thus the proposed system would be more transparent and efficient for student grades evaluation.
Various limitations of the project are:

- Data Mining using Naïve Bayes Classifier does not give 100% efficiency.
- Student data/records used for may not be accurate as that data was entered manually.

You can add more different features to this projectlike:

- Modification in proposed algorithm to increase efficiency of results.
- Using different Data Mining tool for analyzing Students Academic Performance, and to study the effect of using a different tool on Efficiency.

REFERENCES

[1] http://www.cs.waikato.ac.nz/ml/weka/
[2] mlr.cs.umass.edu/ml/datasets.html
[3] Singh, S., & Kumar, V. (2013). Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques, 3(2), 31–37.
[4] Deshpande, S. P., &Thakare, V. M. (2010). DATA MINING SYSTEM AND APPLICATIONS: A REVIEW, 1(1), 32–44.
[5] Samuel. K., "Suitability analysis of data mining tools and methods", bachelor's thesis, 2012
[6] Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques Department of Natural Sciences Turks & Caicos Islands Community College Visualization and Articulation, 0–5. doi:10.1109/CERMA.2007.78 16.
[7] Brijesh Kumar Baradwaj, Research Scholor, Singhaniya University, Rajasthan, India, Mining Education Data to Analyse student's Performance, (IJACSA) International Journal of Advanced Computer Science and Applications Vol. 2, No. 6, 2011 http://www.ijrte.org/attachments/File/v2i4/D0796092413.pdf
[8] Firstman Noah, BaahBarida, Taylor Onate Egerton, Evaluation of Student Performance Using Data Mining Over a Given Data Space, International Journal of Recent Technology and Engineering (IJRTE)ISSN: 2277-3878, Volume-2, Issue-4, September 2013 http://arxiv.org/ftp/arxiv/papers/1201/1201.3417.pdf
[9] Implementing WEKA as a Data Mining Tool to Analyze Students' Academic Performances, NurHafieza Ismail, Fadhilah Ahmad, Azwa Abdul Aziz, University Sultan Zainal Abidin.
[10] Han and Camber, Data Mining: Concepts and Techniques, 2nd edition, Morgan Kaufmann Publishers.