ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)
Available Online at www.ijetcr.org
Volume 3; Issue 6; November-December-2015; Page No. 105-108

Analyzing an Efficient Approach for Educational Data Mining Using Web Based Application Development

Surjeet Kumar*

Assistant Professor, MCA Department, VBS Purvanchal University, Jaunpur UP India 222003

surjeet_k_yadav@yahoo.co.in

Abstract

Educational Data Mining (EDM) is an interdisciplinary research field which has the techniques and the methods to explore data arising in a scholastic fields. EDM has its Computational approaches that are used to examine the data to study educational questions. This paper provides a comprehensive summarization of the important aspects of EDM in today's education system. The major objective of this paper is to deliver a robust educational data mining web application, which allows users to do data extraction, analysis and prediction. The application can be used to mine the educational institute student data and greatly help the department ameliorate its admission strategies and facilitate the admission process. This paper also gives the background information and related studies of EDM. It also identifies the objectives and scope of web application and describes the approach and methodology used in application.

1. Introduction

The research work in the field of Data Mining is always concerned with the finding of patterns in the large warehouse of datasets. Data Mining is a widely used field in the business area but it has scarce applications to Education community. The main focus of Educational Data Mining is to apply tools and techniques of data mining to educationally related dataset. Educational data mining (EDM) is a growing research area that aims at discovering useful information and studying educational problems by exploring data originating from educational context [1]. The process of extracting knowledge from the student data, to enhance teaching and learning standards of educational institutes as well as of teachers and students performance is relatively recent but there are already a number of studies trying to do so and researchers are starting to merge their ideas. It is generally acknowledged that substantial valuable information, which could be vital to many educational decisions, is embedded in the educational data.

International Educational Data Mining Society has been formed with an aim to support collaboration and scientific development in this area. To realize its objectives EDM society organizes a series of conferences, bringing out a journal, development of community resources for sharing of data and techniques. Unfortunately, EDM is still not playing a part in the educational decision-making process in educational institutes. Each year, educational institute attract hundreds of applications from students all over the world for doing study. From these applications, numerous data have been collected, including student profiles, academic performance, research interests, scholarships, teachers' comments and admission decisions.

The usefulness of mining such data is promising but still needs to be proven and stereotypical analysis to be streamlined. However, these data have not been well use of since there lacks an EDM tool in the current admission system to mine these data and probe for useful underlying information. If an EDM application can be built, the administrators and teachers of educational institute can visualize and analyze the candidates' data, and thus discover significant patterns or trends that can provide insights into the admission strategies as well as admission decision-making. Therefore, the objective of this paper is to develop an educational data mining web application, which can use students' data to conduct data extraction, visualization and analysis [1][2].

Moreover, based on the analysis, the application can rank applicants by their performance, filter their

Corresponding author: Surjeet Kumar

applications and predict admission results. With these functionalities, the EDM application can greatly facilitate the admission process of educational institute and help teachers to make admission decisions. This paper proceeds as follows. It first considers the existing researches in the field of EDM and explicates how their achievements can be incorporated into the application. Then it identifies the objectives and the scope and describes the approach and methodology used. It also lists the potential challenges and proposes corresponding countermeasures.

2. Related Work

When there is a talk about the work done in the field of data mining then the most important aspect to focus is that data mining has been applied in a variety of industries, government, military, retail, and banking but it has not received much attention in educational context. It didn't get the attention of most researchers as well. Educational data mining is a field to solve educationally-related problems. Applying data mining this way can help researchers and practitioners discover new ways to uncover patterns and trends within large amounts of educational data

Mashat et al. designed a model to do association rule mining on the university admission data and extract relations between different attributes like GPA and courses already taken by the students. [2] This model provides insights into how to compare the importance of different assessment criteria during the admission process. The use of association rule mining is adaptable in our project.

Feng et al. established a university admissions decision-making model by utilizing the Self Organizing Map (SOM) neural network, cluster analysis, association rule and Fayyad data mining model. [3] Their model astutely takes geographical data into consideration and develops a new reference for admission scheme and propaganda. [3] Their experiment on applying SOM to analyze geographical data draws lessons on how to make use of the origins of the applicants in our project.

Sharma and Bvaghela presented a new algorithm, which makes use of Global Rule Binary Search Tree and distributed data mining techniques to predict the result of students' admission to college. [4] This algorithm can be of enlightening significance to the development of the prediction function of our application.

3. Scope

The main objective of EDM viewed by different Researchers as [5] [6]:

- Student Modeling,
- Domain Modeling,
- Learning System,
- Building the computational models,
- Study the effects of resources.

This paper aims to develop an advanced data manipulation application, which enables the administrators or professors of educational institute to interact with the student data, do data mining and analysis.

The application is vital to the student admission. It can significantly facilitate the admission process and help assessors make better admission decisions by providing a list of functionalities, including data visualization, pattern analysis, correlation analysis, performance analysis, student comparison, performance ranking, smart filtering, intelligent matching and outcome prediction.

The application is scalable and can be later adapted to be used by the Universities for Recruitment Purpose as well. In this paper, an educational data mining web application can be developed. Broadly speaking, the application has three layers, data extraction, data analysis and prediction.

In the data extraction layer, student data will be extracted from the educational institute database and presented to users in the form of statistical graphics, plots or information graphics. Moreover, some open-source data like university ranking will also be extracted and visualized. Users can interact with these visual representations of data and reason about the underlying information in these data.

In the data analysis layer, different analytic tools, such as pattern analysis, correlation analysis and performance analysis, will be developed. Along with these tools, a set of data mining algorithms and statistical models will be implemented. Furthermore, an elegant user interface will be developed so that users can easily and efficiently conduct these analyses.

In the prediction layer, the following five functionalities will be provided:

- Student comparison lets users to compare students' performance based on the specific criteria listed by the user.
- Performance ranking provides the ranking of all the applicants by considering their performance in each criterion and the corresponding proportion assigned to each criterion by the user.
- Smart filtering helps filter students' applications by different requirements and constraints.
- Intelligent matching provides an efficient tool to match applicants with professors based on both parties' preferences.
- Outcome prediction predicts the admission result of each applicant and helps professors make admission decisions.

4. Approach and Methodology

For this application development, the Agile Software Development approach can be applied in order to ensure the data mining application can satisfy the needs of educational institute in terms of the student admission. During the whole application development, developers will closely work with the administrators and professors of educational institute to understand their requirements and seek feedbacks for our designs and implementations. For the data extraction part, developers will first examine existing open-source data visualization tools like Charted and D3 and make use of the appropriate ones to represent and visualize the educational institute student data.

For the data analysis part, developers will research existing data mining techniques and algorithms by reading relative publications and papers and also take relative online courses to better understand different techniques. Then, prototypes to inspect the performance of different data mining algorithms and identify those that meet the requirements will be developed. In the end, based on the selected algorithms, the final version of the data analysis tool with a nice interface can be developed.

For the prediction part, developers will implement the five functionalities in the sequence of Smart Filtering, Student Comparison, Performance Ranking, Outcome Prediction, Intelligent Matching. In this way, later work can be build upon the previous one.

The implementation of this web application is divided into front end development and back end development. JavaScript, CSS and HTML will be used as the languages for front end development, which is responsible for visualizing data and interacting with users. Meanwhile, different open-source JavaScript libraries like d3.js will be used. Python will be used for back end development, which is responsible for interacting with database, implementing data mining algorithms and analytical logic. In particular, ScikitLearn, as a Python library that provides a number of functionalities of data mining and data analysis, [5] will be used in this application.

5. Algorithms and Tools

In data mining, there are different tools and algorithms that are diverse in their methods and aims [8]. There tools for data exploration and visualization to present results in a convenient way to users. Here some of the algorithms and tools are presented that can be used for the data mining purpose.

- Tools: A range of tools like Excel and Access can be used to perform simple SQL queries and visualization. Then there is a tool Clementine[9] for clustering and a data mining platform for teachers, Tada-Ed [10], for clustering, classification and association rule (Clementine is very versatile and powerful but Tada-Ed has preprocessing facilities and visualization of results more tailored to needs). SODAS [11] can be used to perform symbolic data analysis.
- Data exploration and visualization: Raw data and algorithm results can be visualized through tables and graphics such as graphs and histograms as well as through more specific techniques such as symbolic data analysis. The aim is to display data along certain attributes and make extreme points, trends and clusters obvious to human eye.
- Clustering algorithms aim at finding homogeneous groups in data. k-means clustering and its combination with hierarchic clustering [8] can be used for efficient clustering of data. Both methods rest on a distance concept between individuals.
- Classification is used to predict values for some variable. For example, given all the work done by a student, one may want to predict whether the student will perform well in the final exam. C4.5 decision tree can be used from TADA-Ed which relies on the concept of entropy. The tree can be represented by a set of rules such as: if x=v1 and y> v2 then t= v3. Thus, depending on the values an individual takes for, say the variables x and y, one can predict its value for t. The tree is built taking a representative population and is used to predict values for new individuals.

- Association rules find relations between items. Rules have the following form: X -> Y, support 40%, confidence 66%, which could mean 'if students get X incorrectly, then they get also Y incorrectly', with a support of 40% and a confidence of 66%. Support is the frequency in the population of individuals that contains both X and Y. Confidence is the percentage of the instances that contains Y amongst those which contain X. We implemented a variant of the standard Apriori algorithm [8][9] in TADA-Ed that takes temporality into account. Taking temporality into account produces a rule X->Y only if exercise X occurred before Y.

6. Evaluation

The web application delivered can be evaluated according to the performance of each of the three layers.

For the data visualization layer, administrators and professors of educational institute can determine whether the application makes the data more understandable and usable.

For the data analysis layer, developers will test whether the application can identify inconspicuous patterns, trends and correlations and whether these discoveries can help educational institute develop admission strategies.

For the prediction layer, developers will compare the rankings, predictions and matching results provided by the application with the actual rankings and admission results for the past three years and test its accuracy. If it achieves an 80% accuracy, this layer will be considered as completed.

7. Conclusion

The purpose of this paper is to synthesize and share our various experiences of using Data Mining for Education, especially to support reflection on teaching and learning, and to contribute to the emergence of stereotypical directions. This paper will develop a robust educational data mining web application for the administrators and teachers of educational institute to analyze the student data in order to more efficiently admission decisions and answer questions that can be crucial to the admission strategies of educational institute. It can be believed that this application can have a real impact on the admission process in the future and can be later adapted to help universities and companies to recruit new employees as well.

REFERENCES

1. Romero C, Ventura S., Educational Data Mining: A Review of the State of the Art. IEEE Trans. Syst., Man, Cybern. C. 2010; 40: 601-618.
2. Mashat A F, M. Fouad M,Yu P S, Gharib T F: Discovery of Association Rules from University Admission System Data. IJMECS International Journal of Modern Education and Computer Science. 2013; 5:1-7.
3. Feng S, Zhou S, Liu Y. Research on Data Mining in University Admissions Decision-making & quot. International Journal of Advancements in Computing Technology IJACT. 2011; 3: 176–186.
4. B Vaghela D, Sharma P: Students' Admission Prediction using GRBST with Distributed Data Mining. Communications on Applied Electronics CAE. 2015; 2: 15–19.
5. Agathe MERCERON, and Kalina YACEF, "Educational Data Mining: a Case Study", Proceedings of the 12th Iinternational Conference on Artificial Intelligence in Education AIED 2005.
6. O.S. Akinola, B.O. Akinkunmi, T.S. Alo, " A Data Mining Model for Predicting Computer Programming Proficiency of Computer Science Undergraduate Students", African Journal of Computing & ICT January, 2012, Vol 5. No. 1 - ISSN 2006-1781
7. Nelli F. Machine Learning with scikit-learn. Python Data Analytics. 2015; 237–264.
8. Han, J. & M. Kamber, Data Mining: Concepts and Techniques, San Francisco: Morgan Kaufman (2001).
9. SPSS, Clementine, www.spss.com/clementine/ (accessed 2005)
10. Benchaffai, M., G. Debord, A. Merceron, & K. Yacef. "TADA-Ed, a tool to visualize and mine students' online work" in Proceedings of International Conference on Computers in Education, (ICCE04), B. Collis (Eds), pp 1891-1897, Melbourne, Australia: RMIT (2004).
11. SODAS, http://www.ceremade.dauphine.fr/~touati/sodas-pagegarde.htm (accessed 2003)