ATTRIBUTE IDENTIFICATION OF CAMPUS BASED COLLEGE ATTENDEE USING DATA MINING

[1] Mrs. B. Himabindu,[2] K. Hanshu,[3] K. Sainath Reddy,[4] K. Ram Teja,[5] J. Hanumu
[1] Assistant Professor,[2345] B.Tech Students
Department Of Computer Science & Engineering
Sri Indu College Of Engineering & Technology,Sheriguda, Ibrahimpatnam

ABSTRACT

The rise and promotion of big data methods enables teachers to understand the behavior patterns of students in a timely and accurate manner, especially to find out the groups of students that need to be focused on in time, and to help promote the student affairs management from empirical qualitative knowledge to scientific quantitative analysis. This paper applies the clustering method of data mining to analyze the campus network behavior of 3,245 students in a certain grade of B university, obtains a total of 23.843 million Internet access data in 4 years. The result shows 4 groups of students with different characteristics of Internet access, finds 350 students with large network usage. Achievements and other aspects of performance of these students are affected. This study carried out data mining of student campus network behavior, which can be used as a practical operation case for student affairs management data mining, providing effective data support for the accurate and scientific development of student affairs management.

I. INTRODUCTION

A major problem for student affairs management is the contradiction between the limited energy of student counselors and the diversity of student behaviors, which results in many potential problem students losing the opportunity for early intervention. Since the beginning of the 21st century, the rapid development of information technology in education and the construction of digital campuses has made it possible for student counselors to conduct quantitative analysis of student school behaviors, especially to provide early warning to students who may have problems, so that the contradiction could be alleviated by applying the analysis and early warning methods. As contemporary college students who grew up in the Internet era, their daily life, learning and thinking are deeply influenced by the Internet. This provides us with the possibility to understand their campus network behavioral characteristics through big data. How to mine useful information for student counsellors from massive data in the explosive growth of data categories and data scales, is a challenge for current student counsellors, also an important opportunity to conduct work by new means. This study starting from the actual work problems and was conducted based on the network behavior data of B college students, combining big data thinking and big data mining methods, researching the characteristics of college students' network behavior rules, and detecting the students who need pay close attention because the large amount of campus network usage. This study could also carry out as a practical case of student work data mining for reference.

With the advancement of the modern management theory and decision-making science, as well as their application in university and college management, universities and colleges will shift from experience-based management to scientific or information management based on the modern management theories and methodologies. In light of the current situation, the information management method has been consistently supported in universities and colleges, and teaching information management system (IMS) has been established one after the other. However, as the number of students managed and the time spent on using the teaching management information system increases, a large amount of

management data based on teaching is collected. Education has become more flexible and broader as college enrollment continues to expand.

Most universities and colleges are dealing with a conflict between rising student numbers and tightening teaching resources, posing unprecedented difficulties to education management. As a result of such a novel occurrence, the IMS is growing in popularity and recognition among teachers and students because it allows them to work together effectively.

In the previous teaching management, school education management (SEM) only focused on the unique characteristics of the educational field, exaggerating the unique characteristics of school education, emphasizing the management mode based on experience, and to some extent ignoring the similarities between education and general management. Further, SEM only emphasizes the particularity of education field and ignores the commonness between general and education management to some extent. It places too much emphasis on the precision of school curriculum and not enough on the experience-based management approach.

II. LITERATURE SURVEY

TITLE: Variable selection in regression models including functional data predictors.
AUTHORS: Kesheng Liu, Siyang Wang. V
ABSTRACT: Modern research data, where a large number of functional predictors is collected on few subjects are becoming increasingly common. In this paper we propose a variable selection technique, when the predictors are functional and the response is scalar. Our approach is based on adopting a generalized functional linear model framework and using a penalized likelihood method that simultaneously controls the sparsity of the model and the smoothness of the corresponding coefficient functions by adequate penalization. The methodology is characterized by high predictive accuracy, and yields interpretable models, while retaining computational efficiency. The proposed method is investigated numerically in finite samples, and applied to a diffusion tensor imaging tractography data set and a chemometric data set.

TITLE: Penalized spline models for functional principal component analysis
AUTHORS: Yao F. Lee T.
ABSTRACT: We propose an iterative estimation procedure for performing functional principal component analysis. The procedure aims at functional or longitudinal data where the repeated measurements from the same subject are correlated. An increasingly popular smoothing utilizes two types of datasets from 505 university students, i.e., online learning records for a project-based course, and network logs of university campus network. A deep learning framework: Sequential Prediction based on Deep Network (SPDN) is proposed to predict students' performance in the course. SPDN models students' online behavioral sequences by utilizing multi-source fusion CNN technique, and incorporates static information based on bidirectional LSTM. Experiments demonstrate that the proposed SPDN model outperforms the baselines and has a significant improvement on early-warning. Furthermore, it can be learned that Internet access patterns even have a greater impact on students' academic performance than online learning activities.

TITLE: Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue
AUTHORS: Pooja Thakar,Anil Mehta and Manisha
ABSTRACT: In this era of computerization, education has also revamped itself and is not limited to old lecture method. The regular quest is on to find out new ways to make it more effective and efficient for students. Nowadays, lots of data is collected in educational databases, but it remains unutilized. In order to get required benefits from such a big data, powerful tools are required. Data mining is an emerging powerful tool for analysis and prediction. It is successfully applied in the area of fraud detection, advertising, marketing, loan assessment and prediction. But it is in nascent stage in the field of education. Considerable amount of work is done in this direction, but still there are many untouched areas. Moreover, there is no unified approach among these researches. This paper presents a comprehensive survey, a travelogue (2002- 2014) towards educational data mining and its scope in future.

TITLE: Analysis and Research of the Campus Network User's Behavior Based on k-Means Clustering Algorithm
AUTHORS: Quan Shi,Lu Xu,Zhenquan Shi and Yijun Chen
ABSTRACT: This thesis introduces the status and methods of data mining, aiming at the Nantong University campus network users access data preprocessing analysis, using the K-means clustering algorithm combined with SQL Server 2008 and Visual Studio 2008 business intelligence project function for data mining analysis, and the mining experimental results are analyzed and studied. The research indicates that the campus network users of Internet time has a positive relevance with the rate of

student's failing grades and a negative correlation with getting scholarship and CET4(College English Test 4) achievements. What's more, it not only has a positive effect on school leaders fully understand the behavioral characteristics of students and campus network users of campus network usage, timely feedback and guiding the students to form a good habit of learning, but also plays an important role in improving the campus network bandwidth, performance and application efficiency

TITLE: attribute identification of campus-based college attendee using data mining

AUTHORS: Hanan Abdullah Mengash

ABSTRACT: An admissions system based on valid and reliable admissions criteria is very important to select candidates likely to perform well academically at institutions of higher education. This study focuses on ways to support universities in admissions decision making using data mining techniques to predict applicants' academic performance at university. A data set of 2,039 students enrolled in a Computer Science and Information College of a Saudi public university from 2016 to 2019 was used to validate the proposed methodology. The results demonstrate that applicants' early university performance can be predicted before admission based on certain pre-admission criteria (high school grade average, Scholastic Achievement Admission Test score, and General Aptitude Test score). The results also show that Scholastic Achievement Admission Test score is the pre-admission criterion that most accurately predicts future student performance. Therefore, this score should be assigned more weight in admissions systems. We also found that the Artificial Neural Network technique has an accuracy rate above 79%, making it superior to other classification techniques considered (Decision Trees, Support Vector Machines, and Naïve Bay.

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM

A major problem for student affairs management is the contradiction between the limited energy of student counselors and the diversity of student behaviors, which results in many potential problem students losing the opportunity for early intervention. Since the beginning of the 21st century, the rapid development of information technology in education and the construction of digital campuses has made it possible for student counselors to conduct quantitative analysis of student school behaviors, especially to provide early warning to students who may have problems, so that the contradiction could be alleviated by applying the analysis and early warning methods.

DISADVANTAGES:

1. Privacy Concerns: Data mining involves analyzing large amounts of data, which can raise privacy concerns. Personal information of college attendees, such as grades, attendance records, and demographics, may be at risk of unauthorized access or misuse.
2. Bias and Discrimination: Data mining algorithms may unintentionally introduce bias or discrimination based on attributes such as gender, race, or socio-economic background. This can lead to unfair treatment or decisions regarding college attendees.
3. Data Quality Issues: Data mining heavily relies on the quality of the input data. Inaccurate or incomplete data can result in misleading conclusions or incorrect attribute identifications of college attendees.

PROPOSED SYSTEM

Data mining is the process of knowledge discovery which based on a large, incomplete, noisy, fuzzy, random, and original data set, revealing hidden information, previously unknown, but potentially valuable and ultimately understandable information [4]. Conventional data mining deals with traditional data, and mainly treats data as discrete data points. For data mining of functional data, there have been researches that extend traditional methods to functional data processing. These studies have laid a theoretical foundation for the development of this study [5- 6]. This study carried out the clustering analysis based on coefficient vectors obtained from principal component analysis of functional data. Cluster analysis is to classify samples according to their individual characteristics. Through continuous iteration, samples with similar characteristics and rules are in a class, and there are relatively obvious differences between classes. In this study, we use cluster analysis to group student groups with different campus network usage patterns. By analyzing the characteristics of different categories, we can better understand the students' campus network usage patterns and help to discover the group of students who need to pay more attention to the large degree of network usage. It can provide data support for

improving the scientific and accurate of students' affairs management. Currently, the most widely used cluster analysis methods are K- means clustering and hierarchical clustering. K-means clustering algorithm is simple in principle and convenient for processing large amounts of data, but K- values need to be determined through cross-validation and other methods. The calculation speed is slower, and the efficiency is lower when the data volume is large. Therefore, in this study, K-means method was finally selected for clustering.

ADVANTAGES:

We use cluster analysis to group student groups with different campus network usage patterns.

Help to discover the group of students who need to pay more attention to the large degree of network usage.

IV. IMPLEMENTATION

MODULES

- Admin

MODULE DESCRIPTION

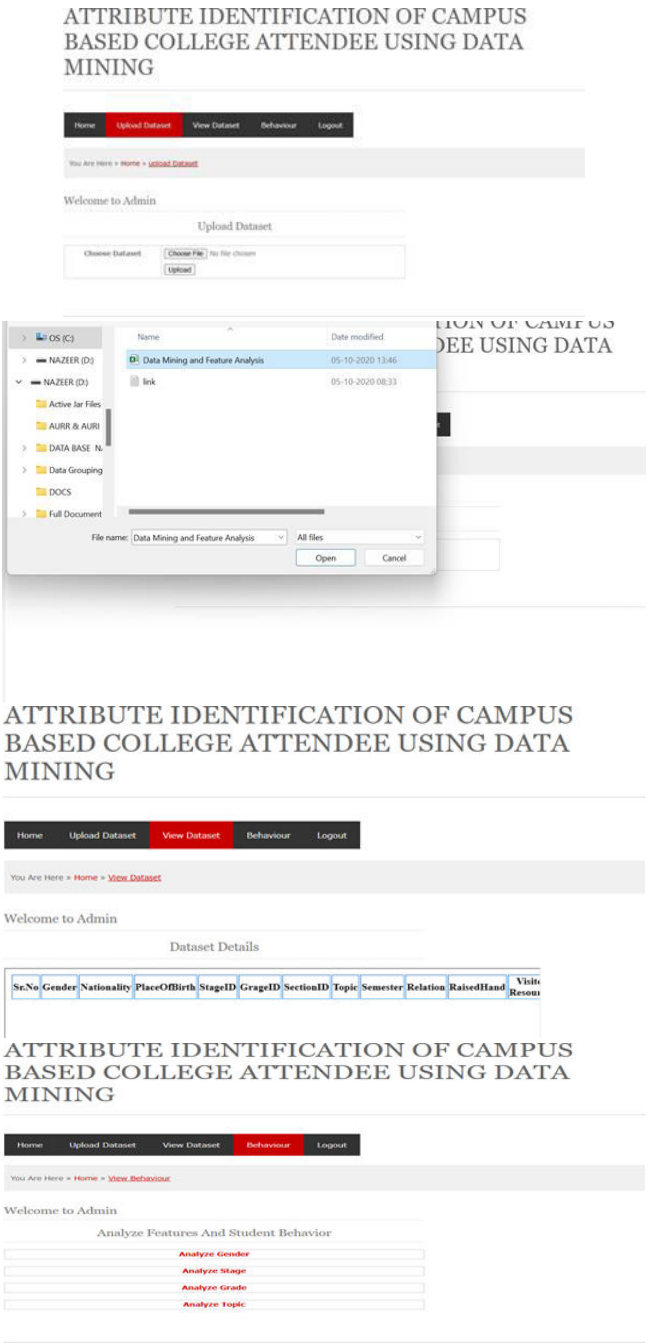
Admin

The Admin module is responsible for managing and controlling the overall system. It typically includes functionalities related to user authentication, authorization, and system configuration. Admins have the authority to add or remove users, assign roles and permissions, and oversee the system's general health. This module is crucial for maintaining the security and integrity of the entire application, ensuring that only authorized users can access and perform specific actions within the system.

In the context of campus-based college attendees, the admin module might focus on collecting and organizing student information, such as attendance records, grades, extracurricular activities, and demographic details. This data can then be analyzed using data mining techniques to identify patterns, trends, and insights that can help improve student performance and campus management.

The admin module could also be responsible for setting up data mining algorithms, defining data mining tasks, and overseeing the entire data mining process to ensure accurate and meaningful results are obtained. Additionally, the admin module might involve ensuring data privacy and security measures are in place to protect the sensitive information of college attendees.

V. SCREENSHOTS:



VI. CONCLUSION

CONCLUSION

This study involved the application of data mining techniques to analyze and extract valuable insights from the behavior of students on the campus network. The primary objective was to utilize these mined data as a practical case for enhancing student affairs management through the lens of data-driven decision-making. By scrutinizing the patterns and trends within student network behavior, the study aimed to contribute to the development of an accurate and scientifically informed approach to student affairs management. The data mining process likely involved

collecting and examining various aspects of student activities on the campus network. This could include tracking internet usage, access to specific resources, time spent online, and patterns of engagement with digital platforms. By employing data mining algorithms and techniques, the study sought to uncover hidden patterns, correlations, and insights within this vast dataset. The mined data could offer valuable information for understanding student behavior, preferences, and needs. For instance, it might reveal patterns indicating peak times of network usage, commonly accessed resources, or particular online activities that are prevalent among students.

The term "effective data support" implies that the findings from the data mining process can be practically utilized in the day-to-day operations of student affairs management.The goal is to enable administrators and educators to make informed decisions that are aligned with the specific needs and behaviors of the student population.

In summary, the study leveraged data mining techniques to analyze student campus network behavior, aiming to offer practical insights for enhancing the precision and scientific basis of student affairs management. The outcomes of this study could potentially inform strategies and policies that contribute to a more effective and tailored approach to student support and engagement.

FUTURE SCOPE

Personalized Student Support:

The identification of student groups with distinct internet behavior patterns opens the door to personalized student support strategies. Institutions can tailor interventions, resources, and guidance based on the specific needs and characteristics of each group, fostering a more individualized and effective approach to student affairs management.

Early Intervention and Support:

The ability to identify, in a timely manner, students who exhibit patterns of significant internet usage or other behaviors that may impact their academic performance allows for proactive interventions. Early identification can lead to timely support mechanisms, such as counseling, academic assistance, or targeted programs, to address challenges and enhance student success.

Refinement of Student Affairs Policies:

The insights gained from data mining can contribute to the refinement and development of student affairs policies. Institutions can make evidence-based decisions regarding internet usage policies, resource allocation, and the design of campus support services. This move from empirical qualitative knowledge to scientific quantitative analysis enhances the precision and effectiveness of student affairs management strategies.

Continuous Improvement in Educational Practices:

Continuous data mining and analysis can contribute to an ongoing cycle of improvement in educational practices. Institutions can adapt and refine their approaches based on changing student behavior patterns, technological advancements, and evolving educational needs. This dynamic feedback loop enables educational institutions to stay responsive and relevant.

REFERENCES

1. Ramsay J. O., Silverman B. W. Functional data analysis[M]. New York: Springer, 1997.
2. Ramsay J. O., Silverman B. W. Applied functional data analysis: methods and casestudies[M]. Vol. 77. New York: Springer, 2002.
3. Kesheng Liu, Siyang Wang. Variable selection in regression models including functional data predictors. Journal of Beijing University of aeronautics and astranautics, 2019, 45(10): 1990-1994.
4. Romero C., Ventura S., Data mining in education[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2013, 3(1):12-27.
5. Locantore N., Marron J., Simpson D., et al. Robust principal component analysis forfunctional data[J]. Test, 1999, 8(1):1–73.
6. Yao F., Lee T. Penalized spline models for functional principal component analysis[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2006,68(1):3–25.
7. iiMedia. w 2018 Chinese College Students' Online Leisure and Entertainment Behavior Monitoring Analysis Reportx[EB/OL].2018-11-].