"Examining Successful Attributes for Academic Performance using Machine Learning"

AMRUTHA N S[1], Mrs. PUSHPALATHA S[2]

[1]PG Scholar, [2] Assistant Professor

Department of Electronics and Communication Engineering,

Visvesvaraya Technological University Centre for Post Graduate Studies,

Mysuru, INDIA

Abstract

Identification of different factors which affects a student's learning behavior and performance during academic career. Analyzing student mental issues and low academic performances is a complex task in the current education sector. In current system it is difficult to track the students behavior and characteristics. There is no automation or tool which predicts or shows how to improvise the students academic performances. Finding the correlation between factors that affects student's performance and academic results pays a vital role in the current education sector. System uses data science technique called as "*Association Learning*" to find the patterns. We use either "*apriori algorithm"* or "*apriori TID algorithm*" or "*Eclat algorithm*" to find patterns. Proposed system build as real time application useful to colleges and lecturers to know the students behaviour patterns. System also aims at predicting the individual performance. System uses machine learning algorithm called as "*naive bayes*" algorithm for student's performance prediction individually.

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 counselor from massive data in the explosive growth of data categories and data scales, is a challenge for current student counselor, 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 science 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 science for reference.

Related Works

An Intelligent Student Advising System Using Collaborative Filtering

We propose a web based intelligent student advising system using collaborative filtering, a technique commonly used in recommendation systems assuming that users with similar characteristics and behaviors will have similar preferences. With our advising system, students are sorted into groups and given advice based on their similarities to the groups. If a student is determined to be similar to a group student, a course preferred by that group might be recommended to the student.

Disadvantages

1. System used to predict suitable course for students and data-set not compatible to predict student results.
2. Not all student behaviors connected to course advising.
3. Students are grouped and then system predicts the suitable course for the students. Grouping lacks over data for prediction.

Mining Students' Data for Performance Prediction

The ability to predict a student's performance is very important in educational environments. Students' academic performance is based upon diverse factors like personal, social, Psychological and other environmental variables. A very promising tool to attain this objective is the use of Data Mining. Data mining techniques are used to discover hidden information patterns and relationships of large amount of data, which is very much helpful in decision making. A single data contains a lot of information. The

type of information is produced by the data and it decides the processing method of data. A lot of data that can produce valuable information, in education sector contains this valuable information. Which helps the education sector to capture and compile low cost information for this information and communication technology is used. Now-a-days educational database is increased rapidly because of the large amount of data stored in it. The loyal students motivate the higher education systems, to know them well; the best way is by using valid management and processing of the students' database. Data mining approach provides valid information from existing student to manage relationships with upcoming students.

Disadvantages

1. System predicts student performance based on the student behaviors. Data-set not compatible for class results prediction.

2. Used more irrelevant parameters for student performance prediction such as father income, mother income, qualification etc..

Gap Analysis

- To build a real world application this is useful for colleges and students to improvise performance.

- To develop a system which predicts the relationship between students behavior and performance using the educational data-set.

- Proposed project is a students behavior analysis and prediction and management system which is meant for educational institute.

- System aim is to reduce the number of students with poor performances.

- System makes use of "*Association Learning*" to discover the educational patterns using data science algorithms.

Proposed Work

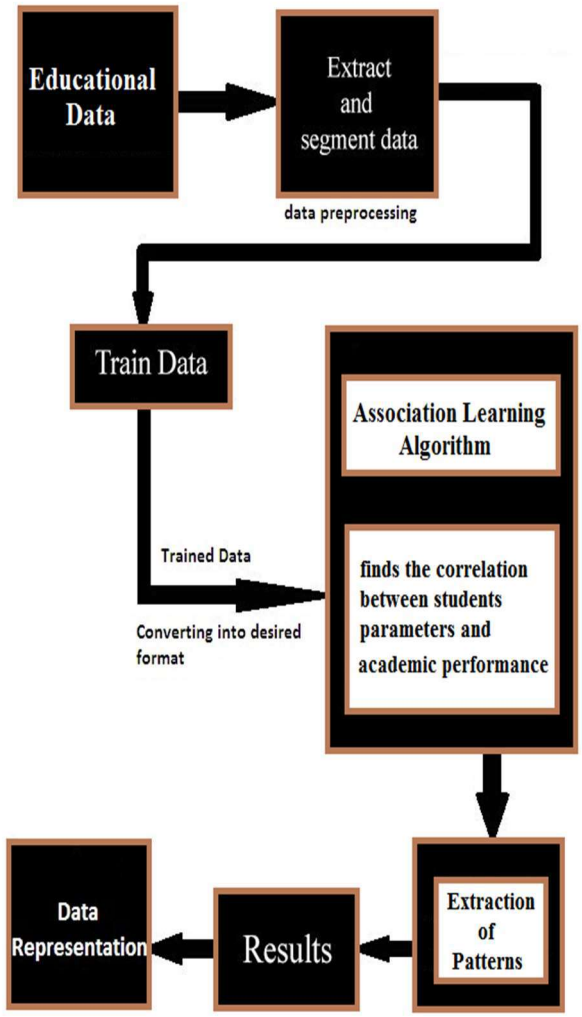
The major objective is to find behaviour patterns of students in a timely and accurate manner. Main aim is specifically to identify the categories of students who require extra attention. Without campus behaviour analysis, These students' academics and several other performance dimensions are impacted. System uses parameters such as attendance status, extra circular activities, grade, technical skills, previous semester

results, grasping capability, Aptitude grade, interaction with lecturers etc.. System helps lecturers to identify the most influential factors affecting the students performance.

System uses data science technique called as "*Association Learning*" to find the patterns. We use "*Eclat algorithm*" to find patterns and classification algorithm i.e "Naïve bayes Algorithm" to make individual predicition of student's performance. Proposed system to build as web application useful to colleges and lecturers to know the students behaviour patterns.

Objectives of the Proposed System

- Proposed project is a students behavior analysis and prediction and management system which is meant for educational institute.
- Proposed project makes use of ML technique or Data Science approach for the students behavior analysis.
- To provide valid information from existing students to manage relationships with upcoming students.
- Identification of different factors which affects a student's learning behavior and performance during academic career.
- Construction of a prediction model using ML technique - Unsupervised Learning on the bases of identified predictive variables.



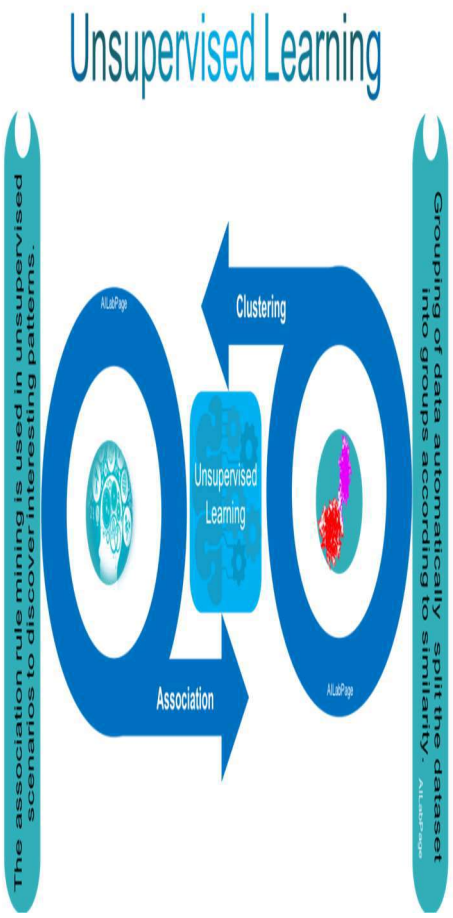
Methodology

Association (or relation) is probably the better known and most familiar and straightforward data science technique. Here, we make a simple correlation between two or

more items, often of the same type to identify patterns.

For example, Market-basket analysis, where we track people's buying habits, we might identify that a customer always buys cream when they buy strawberries, and therefore suggest that the next time that they buy strawberries they might also want to buy cream.

In our project Association Learning Algorithm "*Apriori Algorithm/ Eclat Algorithm*" is used to predict the relationship between students behavior and performance using the educational data-set. Eclat algorithm is one of the efficient algorithm and takes less time for data processing. This algorithm works fine for small data-sets as well as large data-sets.



Pattern Prediction Process

Step 1: Data Collection

We are working on real time application, we build a new application which contains data servers (used to store data). Data collection means collecting data from different sources. Data includes student parameters and student results.

Step 2: Data Preparation

Here data from servers extracted and analyzed. Complete data extracted and

analyzed where we remove irrelevant data and retain data required for processing. According to the project only student parameters and student results are required to generate outputs.

Step 3: Specify Constraints

SUPPORT COUNT

The relationship between the total number of transaction containing that item (A) with the total number of transaction in data set.

CONFIDENCE

Confidence of item set defined as total number of transaction containing the item set to the total number of transaction containing LHS.

Step 4: Association Rules Mining (Eclat Algorithm)

Association (or relation) is probably the better known and most familiar and straightforward data mining technique. Here, we make a simple correlation between two or more items, often of the same type to identify patterns.

For example, Market-basket analysis, where we track people's buying habits, we might identify that a customer always buys cream when they buy strawberries, and therefore suggest that the next time that they buy strawberries they might also want to buy cream.

We use eclat algorithm to process e commerce data and to find the patterns. Here we generate patterns related to education.

Eclat algorithm is selected because of the following reasons.

1. Quicker Results (takes less time for Prediction)

2. Works fine for small data set as well as Huge data set.

3. One scan of Database is Enough.

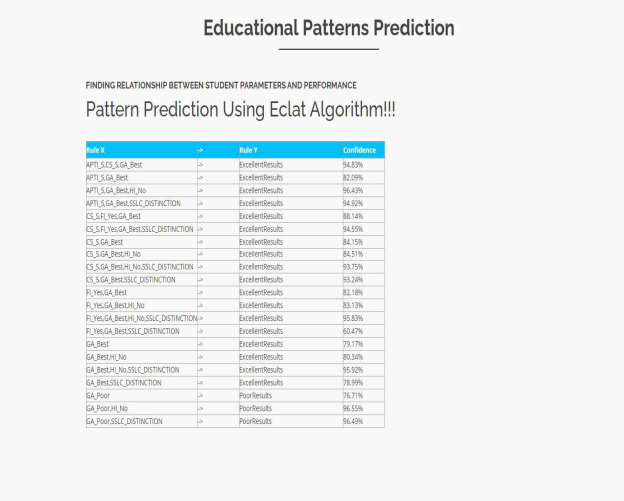
4.Works fine for multiple constraints.

Step 5: Patterns Prediction

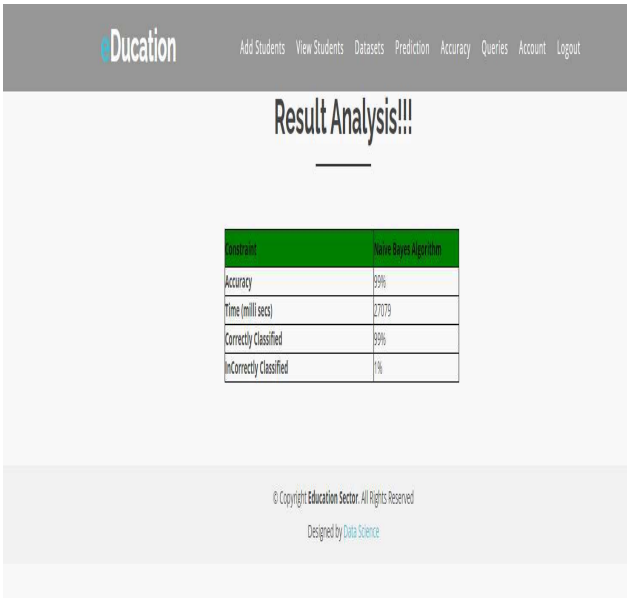
Here system predicts the relationship between students behavior and performance using the educational data-set.

Data Visualization

Educational Patterns Prediction



Individual Prediction Results



Conclusion

Identification of different factors which affects a student's learning behavior and performance during academic career. Analyzing student mental issues and low academic performances is a complex task in the current education sector. System uses data science technique called as "*Association Learning*" to find the patterns. We use either "*apriori algorithm"* or "*apriori TID algorithm*" or "*Eclat algorithm*" to find patterns. Proposed system build as real time application useful to colleges and lecturers to know the students behaviour patterns. System helps lecturers to identify the most influential factors affecting the students performance.

Future Enhancements

More training datasets can be used to find more related patterns. More algorithms can be used to find the students behaviour patterns and algorithms can be compared to identify the algorithm with better results.

References

[1] R. L. Ahadi, H. Haapala, and A. Vihavainen, "Exploring machine learning methods to automatically identify students in need of assistance," in Proc. 11th Annu. Int. Conf. Int. Comput. Educ. Res., 2015, pp. 121–130.

[2] K. Quille and S. Bergin, "Programming: Further factors that influence success," in Psychology of Programming Interest Group (PPIG). Cambridge, U.K.: Univ. Cambridge, 2016.

[3] C. Y. Ko and F. Y. Leu, "Analyzing attributes of successful learners by using machine learning in an undergraduate computer course," in Proc. 32nd IEEE Int. Conf. Adv. Inf. Netw. Appl. (AINA-2018), Krakow, Poland, 2018, pp. 801–806.

[4] S. Kotsiantis and D. Kanellopoulos, "Association rules mining: A recent overview," Int. Trans. Comput. Sci. Eng., vol. 32, no. 1, pp. 71–82, 2006.

[5] J.-L. Hung and K. Zhang, "Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching," J. Online Learn. Teach., vol. 4, no. 4, pp. 426–436, 2008.

[6] A. Ezen-Can, K. E. Boyer, S. Kellogg, and S. Booth, "Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach," in Proc. 5th Int. Conf. Learn. Anal. Knowl., 2015, pp. 146–150.

[7] J.-L. Hung, M. C. Wang, S. Wang, M. Abdelrasoul, Y. Li, and W. He, "Identifying at-risk students for early interventions—A time-series clustering approach," IEEE Trans. Emerg. Topics Comput., vol. 5, no. 1, pp. 45–55, Jan.–Mar. 2017.

[8] C. Romero, M.-I. López, J.-M. Luna, S. Ventura, "Predicting students' final performance from participation in on-line discussion forums," Comput. Educ. vol. 68, pp. 458–472, Oct. 2013.

[9] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," Comput. Educ., vol. 113, pp. 177–194, Oct. 2017.

[10] S. Amershi and C. Conati, "Unsupervised and supervised machine learning in user modeling for intelligent learning environments," in Proc. 12th Int. Conf. Intell. User Interfaces, 2007, pp. 72–81.

[11] B. J. Zimmerman, "Attaining of self-regulation: A social cognitive perspective," in Handbook of Self-Regulation, Research, and Applications, M. Boekaerts, P. Pintrich, and M. Zeidner, Eds. Orlando, FL, USA: Academic, 2000, pp. 13–39.

[12] B. J. Zimmerman, "Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects," Amer. Educ. Res. J., vol. 45, no. 1, pp. 166–183, 2008.

[13] A. Bandura, Social Learning Theory. Oxford, U.K.: Prentice-Hall, 1977.

[14] A. Bandura, Self-Efficacy: The Exercise of Control. New York, NY, USA: Freeman, 1997.

[15] R. Lynch and M. Dembo, "The relationship between self-regulation and online learning in a blended learning context," Int. Rev. Res. Open Distance Learn., vol. 5, no. 2, pp. 1–16, 2004.

[16] M. V. J. Veenman, B. H. A. M. Van Hout-Wolters, and P. Afflerbach, "Metacognition and learning: Conceptual and methodological considerations," Metacogn. Learn., vol. 1, pp. 3–14, Mar. 2006.

[17] P. R. Pintrich, "The role of goal orientation in self-regulated learning," in Handbook of Self-Regulation, M. Boekaerts, P. R. Pintrich, and M. Zeidner, Eds. San Diego, CA, USA: Academic, 2000, pp. 451–502.

[18] D. H. Schunk, "Self-regulated learning: The educational legacy of Paul R. Pintrich," Educ. Psychol., vol. 40, no. 2, pp. 85–94, 2005.

[19] A. Kitsantas, A. Winsler, and F. Huie, "Self-regulation and ability predictors of academic success during college: A predictive validity study," J. Adv. Acad., vol. 20, no. 1, pp. 42–68, 2008.

[20] D. Compeau and C. Higgins, "Computer self-efficacy: Development of a measure and initial test," MIS Quart., vol. 19, no. 2, pp. 189–211, 1995.

[21] R. J. Roiger and M. W. Gaetz, Data Mining: A: Tutorial-Based Primer. Boston, MA, USA: Addison-Wesley, 2003.

[22] I. H. Wittlen, E. Frank, and A. H. Mark, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.

[23] S. B. Kotsiantis, "Decision trees: A recent overview," Artif. Intell. Rev., vol. 39, pp. 261–283, Apr. 2013.

[24] J. Ross Quinlan. C4.5: Programs for Machine Learning. San Mateo, CA, USA: Morgan Kaufman. 1993.

[25] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," Informatica, vol. 31, no. 3, pp. 249–268, 2007.

[26] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Machine learning: A review of classification and combining techniques," Artif. Intell. Rev. vol. 26, no. 3, pp. 159–190, 2006.

[27] W. S. Noble, “What is a support vector machine?” Nat. Biotechnol., vol. 24, no. 12, pp. 1565–1567, 2006.

[28] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, “Efficient kNN classifi- cation with different numbers of nearest neighbors,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 5, pp. 1774–1785, May 2018.

[29] P. R. Pintrich, D. A. F. Smith, T. Garcia, and W. J. McKeachie, “Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ),” Educ. Psychol. Meas., vol. 53, no. 3, pp. 801–813, 1993.

[30] S. Cassidy and P. Eachus, “Developing the computer user self-efficacy (CUSE) scale: Investigating the relationship between computer selfefficacy, gender and experience with computers,” J. Comput. Res., vol. 26, no. 2, pp. 133–153, 2002.

[31] P. Marriott and A. Lau, “The use of on-line summative assessment in an undergraduate financial accounting course,” J. Account. Educ., vol. 26, no. 2, pp. 73–90, 2008.

[32] D. J. Nicol and D. Macfarlane-Dick, “Formative assessment and selfregulated learning: A model and seven principles of good feedback practice,” Stud. Higher Educ., vol. 31, no. 2, pp. 199–218, 2006.