ISSN 2348 – 9928
IJAICT Volume 4, Issue 11, March 2018 Doi:01.0401/ijaict.2018.03.19 Published on 05 (04) 2018

PROGNOSTICATION OF STUDENT DROPOUTS IN COLLEGE

Ms. J. Shanthalakshmi Revathy
Assistant Professor

Ms. M. Gayathri Rakshana
UG Scholar

Ms. P. R. Ranjani
UG Scholar

Ms. P. Kavinisha
UG Scholar

Velammal College of Engineering and Technology,
Madurai, Tamilnadu, India.

Abstract *– Nowadays predicting student dropouts is a major problem in the educational system. A Student's dropout have future implications not only for themselves but also for the society as a whole. Dropouts often face unemployment, involvement in criminal activities and depression. This paper examined the prediction of dropouts using Data mining approach-Naïve Bayes and collaborative filtering methodology. Naïve Bayes is a classification technique based on Bayes theorem of probability to predict the class of unknown dataset which consists of attributes. Attributes is nothing but real data of students that is collected from college. Using* **this** *method, we try to boost the correctness for computing the students may not pass or dropout first with all accessible characteristics. The technique is applied on data that is readily available in the institution's database. The collected data includes student study history, lack of parent's engagement, economic needs of parents and many more. The analysis and the information about predication is more helpful for college management and teachers to improve education in a better way and also make changes if necessary.*

Keywords *– Datamining, Prediction, Dropout Rate of Students.*

I. INTRODUCTION

A facet of students achievements often overshadowed by standardized test scores is the students dropout rate. Costs associated with students dropout are not only felt by the dropouts themselves. A summary table on the National Dropout Prevention Center's (2004)website showed a 31% decrease in the average hourly wage (adjusted for inflation) of dropouts between 1973 and 1997. McDill, Natrielo, and Pallas (1986) estimated the lifetime earnings losses for 516,000 dropouts in 1980 at approximately $55 billion, after adjusting the estimated lifetime earnings,downward by half to be adjusted for biases related to differences in ability. Catterall (1987) estimated the total loss in the lifetime tax revenue associated with dropouts at around $70 billion,for one of the 8th grade students in the U.S. LeCompte and Dworkin(1991) claimed that between 1986 and 1991, New York City spent close to $40 million annually on dropout prevention programs. Rumberger (1995) cited the increase in the amount of money spent on the dropout prevention, job training, and welfare programs.In addition,increasing dropout rates not only bode well for the advancement of a highly-educated workforce but also reduces the future labor productivity potential.

II. LITERATURE SURVEY

Educational dropouts is one of the major problem in india.There are lots of factors which affects the student's education,but financial and domestic responsibilities are the major reason for educational dropouts.A national survey office shows that in India 13 out of 100 between the age group of 5-25 years is educational dropouts because they did not consider education important.In another survey conducted by the independent agency,that one out of four students is educational dropout due to the similar reason.Recently, many researchers in the area of machine learning and data mining tries to address the student retention phenomenon in college and university.
In this section, we briefly discuss the works which is similar techniques as our approach but serve for different purposes.

Loretta Auvi, Anthony Don, Ben Shneiderman, Elena Zheleva, Catherine Plaisan, Machon Gregory, Tanya Clement, and Sureyya Tarkan in proposed about the Feature Lens, visualize a text or data compilation at numerous stages of granularity and facilitate the consumers to discover interesting text or data patterns in the data warehouse. The current accomplishment focuses on everyday entry sets of n-grams, as they incarcerate the replication of accurate or comparable terminology in the compilation. Users can locate meaningful co-occurrences of data patterns or text by envisaging them within and transversely documents in the text collection in the databases. This also consents the users to recognize the sequential progression of tradition such as goes up and down or sudden appearance of text prototypes. The boundary could be worn to discover other copy features as fine.

Ah-Hwee Tan proposed data or text mining, It is also known as text data mining or knowledge discovery. From textual

© 2018 IJAICT (www.ijaict.com)

databases refers to the procedure of removing interest and significant model or knowledge from copy documents. There is a fast growth in the computer and network technologies in recent years. In this technology, numeric data's also made available in the current time and it show the fast growth in this field. This critique challenges to shack some lights to the query text mining structure involves two components: unstructured text documents transform into intermediate form by using text refining and knowledge sanitization that deduces patterns or knowledge from the intermediate form. In conclusion, we emphasize the upcoming challenges of text mining and the opportunities it offers. M. Rajman, and R.Besancon, proposed the common framework of knowledge discovery, This type of technologies is simple to gather and provisions in a huge quantity of unstructured or semi-structured text or data and are present in form of WebPages, HTML/XML archives, emails, and text files. And these copy information can be an idea with the great level text types of databases, it becomes significant to expand discipline to determine exciting knowledge or news from such data warehouses. In Collaborative data publishing more providers. desire to calculate an identify view of their data Without disclosing any responsive and personal information.

III. RELATED WORK

Research in Educational Data Mining has gained momentum over the past few years. Various aspects of learners and learning styles have been studied. Classification, Clustering, Association rules have been used widely. Students learning situations, attitudes, tendencies, and behaviours have also been studied.It has been observed that different parameters are studied and number of predictions have been obtained by implementing techniques on them. Some of them include their scores in individual subjects, economic status, academic progress, psychological profile ,demographic data, their past scores in school, personal and family information, mothers' and fathers' occupations. Considering the tendency of effect of environment on psychology of the student, the family details such as family income, working status of parents and the residing place of the student throughout his education is considered. Different techniques were implemented for various datasets. Norlida Binyamin et. Al applied the most commonly used classifiers techniques in Educational Data Mining and mentioned an outline of Neuro- Fuzzy classification Techniques of how to obtain knowledge from databases such as large arrays of student data from academic Institution databases . Neuro –Fuzzy works with incomplete data but it does not support mixed variable.

B.A classification model is developed by Bo Guo et. Al to predict student using Deep Learning which automatically learns multiple levels of representation. They pre-trained hidden layers of features layer wisely using an unsupervised learning algorithm sparse auto-encoder from unlabeled data, and then use supervised training for fine tuning of the parameters. Estimation of Student Performance by Considering Consecutive Lessons proposes a new method of data mining to predict student performance. The process deploys Latent Dirichlet Allocation (LDA) and Support Vector Machine (SVM) to tell about student grades in each lesson to be obtained in future. Anjana Pradeep, et. Al performed data mining to identify the weak students who are likely to perform poor in their academics. Various classification techniques such as induction rules and decision tree are implemented. V.Vivekananda and Devipriya used a properly designed Decision Support System to help decision makers compile useful information from a combination of raw data, documents, and personal knowledge, or models in business to detect and solve problems and take decisions.

IV. PROPOSED SYSTEM

The proposed system is quite different since it uses .NET programming instead of using datamining tools.Hence it is platform independent.The project can be run on any machine.The datamining technique which is used in this system is Naïve Bayes and collaborative filtering methodology which is an added advantage to the project.

V. DEMOGRAPHIC DATA

Attributes	Percentage of students	Percentage of dropout students
Not interested in studying	74.17	62.32
Attendance	82.50	65.22
Academic scores	27.54	12.50
Family Income	25.83	37.68
Understanding	37.50	47.48
Emotional	88.33	91.30
Fight with friends	5.00	7.25
Bad Habit	49.32	50.21
Police Complaint	78.34	83.23
Health Issues	67.32	63.27
Family size	11.45	8.23
Extracurricular activities	12.72	11.01

© 2018 IJAICT (www.ijaict.com)

5.1 Data Gathering

The process of data gathering is that involves in collecting all available information about students .the set of factor should be identified that can affect student's performance and collected from different available data sources .the collected characteristics or risk factors that can influence to students failure or dropped out. Risk factors contain the information about student's cultural, social, educational background, socioeconomic status, psychological profile and academic progress .In which most of the students are aged between 15 and 16 and this is the years with the highest rate of failure. Finally the survey is to obtain personal and family information to identify important risk factors of all students and school services provides the score obtained by the students in all subjects of course. All those information are integrated into single dataset.

5.2 Pre-Processing

In this stage dataset is prepared for applying data mining technique. Before applying data mining technique, pre-processing methods like cleaning, variable transformation and data partitioning and other technique attribute selection is must be applied. Here new attribute of age is created using date of birth of each students. The continues variables are transformed into discreet variable that is scores obtained by each student is changed into categorical values (i.e) Excellent score between 9.5 and 10,Very good the score between 8.5 and 9.4.all information's are integrated in single dataset that is stored in .arff format of Weka tool. Finally entire dataset is divided randomly into 10 pairs of training and test data files. After pre-processing we have attributes or variables for each student. Each test file will contain best attributes and rebalanced.

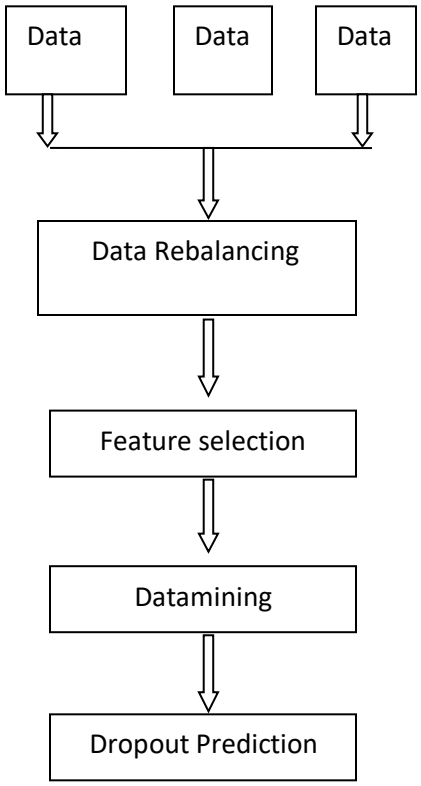
5.3 Data Mining

In this stage Data mining technique is going to be applied. Here the data mining technique is mainly used for classification. The classification is based on best attribute selection from data set. In which the naive bays algorithm is implemented for classification of data. Traditionally the Weka Software tool is used for data mining. It contains verity of data mining algorithms. Weka implements decision tree, it is a set of condition organized in hierarchical structure. Here the classification algorithms were executed using cross- validation and all available information. Finally the result with the test file of classification is shown.

5.4 Interpretation

In which, the obtained results are analyzed to predict student failure or dropped out. To achieve this previous test results are taken for comparison. At this stage classification rules are applied for predict relevant factors and relationships that lead to student pass or fail. There are attribute that indicate that student who failed are older than 15 year and some of the attribute are shows marks of poor, not presented and regular students. Finally the risk factors are analyzed from previous results of classification algorithms.

5.4 System Architecture



5.5 Working Method

The working of this project is that the dataset is gathered from the instituition's database and then the parameters are passed as arguments.The parameters are the behavioural attributes which are used in the project.The Naïve Bayes classifier is used .The collaborative filtering is used in order to get a valid output after processing both the structured and unstructured data. Finally the criterias are driven as output.Based on the criterias the student is found and considered to be a dropout or not.

© 2018 IJAICT (www.ijaict.com)

VI. CONCLUSION

This project will help in determining the dropout accuracy in terms of percentage value. There will be an immediate action taken, since the student's dropout situation if found will be intimated to the parents, thereby will be helpful in taking further steps or actions to improve the student's academic performance.

References

[1]. Patricia A. Aloise-Young and Ernest L. Chavez, "Not All School Dropouts are the same: Ethnic Differences in the Relation between Reason for Leaving School and Adolescent Substance Use", Psychology in the Schools, Vol. 39, No. 5, pp. 539-547, 2002.

[2]. Carlos Márquez-Vera, Cristóbal Romero Morales and Sebastián Ventura Soto, "Predicting School Failure and Dropout by using Data Mining Techniques", IEEE Journal of Latin-American Learning Technologies, Vol. 8, No. 1, pp. 7-14, 2013.

[3]. R. Arulmurugan and H. Anandakumar, "Early Detection of Lung Cancer Using Wavelet Feature Descriptor and Feed Forward Back Propagation Neural Networks Classifier," Lecture Notes in Computational Vision and Biomechanics, pp. 103–110, 2018. doi:10.1007/978-3-319-71767-8_9

[4]. Haldorai, A. Ramu, and S. Murugan, "Social Aware Cognitive Radio Networks," Social Network Analytics for Contemporary Business Organizations, pp. 188–202. doi:10.4018/978-1-5225-5097-6.ch010

[5]. Haldorai and A. Ramu, "The Impact of Big Data Analytics and Challenges to Cyber Security," Advances in Information Security, Privacy, and Ethics, pp. 300–314. doi:10.4018/978-1-5225-4100-4.ch016

[6]. H. Anandakumar and K. Umamaheswari, "A bio-inspired swarm intelligence technique for social aware cognitive radio handovers," Computers & Electrical Engineering, Sep. 2017. doi:10.1016/j.compeleceng.2017.09.016

[7]. R. Arulmurugan, K. R. Sabarmathi, and H. Anandakumar, "Classification of sentence level sentiment analysis using cloud machine learning techniques," Cluster Computing, Sep. 2017. doi:10.1007/s10586-017-1200-1

[8]. H. Anandakumar and K. Umamaheswari, "An Efficient Optimized Handover in Cognitive Radio Networks using Cooperative Spectrum Sensing," Intelligent Automation & Soft Computing, pp. 1–8, Sep. 2017. doi:10.1080/10798587.2017.1364931

[9]. Francisco Araque, Concepción Roldán and Alberto Salguero, "Factors Influencing University Dropout Rates", Computers and Education, Vol. 53, No. 3, pp. 563-574, 2009.

[10]. C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art", IEEE Transactions on Systems, Man, and Cybernetics, Part C, Applications and Reviews, Vol. 40, No. 6, pp. 601-618, 2010.

[11]. Rajni Jindal and Malaya Dutta Borah, "A Survey on Educational Data Mining and Research Trends", International Journal of Database Management Systems, Vol. 5, No. 3, pp. 53-73, 2013.

[12]. Bharat Inder Fozdar, Lalita S. Kumar and S. Kannan, "A Survey of a Study on the Reasons Responsible for Student Dropout from the Bachelor of Science Programme at Indira Gandhi National Open University", International Review of Research in Open and Distance Learning, Vol. 7, No. 3. pp. 1-15, 2006.

[13]. Mohammed M. Abu Tair and Alaa M. El-Halees, "Mining Educational Data to Improve Student's Performance", International Journal of Information and Communication Technology Research, Vol. 2, No. 2, pp. 140-146, 2012.

[14]. Sotiris B. Kotsiantis, "Educational Data Mining: A Case Study for Predicting Dropout-Prone Students", International Journal of Knowledge Engineering Soft Data Paradigms, Vol. 1, No. 2, pp. 101-111, 2009.

[15]. L. Fourtin, D. Marcotte, P. Potvin, E. Roger and J. Joly, "Typology of Students at Risk of Dropping Out of School: Description by Personal, Family and School Factors", European Journal of Psychology of Education, Vol. XXI, No. 4, pp. 363-383, 2006.

[16]. L.G. Moseley and D.M. Mead, "Predicting Who Will Drop Out of Nursing Courses: A Machine Learning Exercise", Nurse Education Today, Vol. 28, No. 4, pp. 469-475, 2008.

[17]. M.A. Hall and G. Holmes, "Benchmarking Attribute Selection Techniques for Data Mining," Working paper No. 00/10, Department of Computer Science, University of Waikato, 2002.

[18]. L. Breiman, J.H. Friedman, R.A. Olshen and C.J. Stone, "Classification and Regression Trees (Wadsworth Statistics/Probability)", New York: Chapman & Hall, 1984.

[19]. Yoav Freund and Llew Mason, "The Alternating Decision Tree Algorithm," Proceedings of the Sixteenth International Conference on Machine Learning, pp. 124- 133, 1999.

[20]. S. Nandni, R. Subashree, T. Tamilselvan, E. Vinodhini, and H. Anandakumar, "A study on cognitive social data fusion," 2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT), Mar. 2017. doi:10.1109/igeht.2017.8094075

[21]. H. Anandakumar and K. Umamaheswari, "Supervised machine learning techniques in cognitive radio networks during cooperative spectrum handovers," Cluster Computing, vol. 20, no. 2, pp. 1505–1515, Mar. 2017. doi:10.1007/s10586-017-0798-3

[22]. M. Suganya and H. Anandakumar, "Handover based spectrum allocation in cognitive radio networks," 2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE), Dec. 2013. doi:10.1109/icgce.2013.6823431

[23]. Roshini and H. Anandakumar, "Hierarchical cost effective leach for heterogeneous wireless sensor networks," 2015 International Conference on Advanced Computing and Communication Systems, Jan. 2015. doi:10.1109/icaccs.2015.7324082

[24]. S. Divya, H. A. Kumar, and A. Vishalakshi, "An improved spectral efficiency of WiMAX using 802.16G based technology," 2015 International Conference on Advanced Computing and Communication Systems, Jan. 2015. doi:10.1109/icaccs.2015.7324098

[25]. K. Mythili and H. Anandakumar, "Trust management approach for secure and privacy data access in cloud computing," 2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE), Dec. 2013. doi:10.1109/icgce.2013.6823567

© 2018 IJAICT (www.ijaict.com)