International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 2
February 2018

A Survey on Academic Progression of Students in Tertiary Education using Classification Algorithms

Sujith Jayaprakash
Research Scholar, Dr. N.G.P Arts & Science College, Department of PG & Research
Dr. Jaiganesh V
Assistant Professor, Dr. N.G.P Arts & Science College, Department of PG & Research

Abstract— Education Data Mining has taken a big leap in the area of research. Several researcher scholars have taken Education Data Mining to the next level through their research findings. Academic progression of students in Higher Education system is one of the key areasunder EDM where majority of the research works are developed which not only helps the institutions to take a preventive measure to better the progression of their students but also minimizes the attrition rate. This paper surveys the various research works performed in this area and tries to identify an optimal solution which can be enhanced for the future research works. It also discusses about various classification algorithms, attributes and tools used by the researchers to predict the academic progression.

Index Terms—Academic progression, Education Data Mining, Classification Algorithms, Student Attrition

I. INTRODUCTION

Producing high quality graduates is a key objective of every higher education institution. But very few institutions adhere to or strive to achieve it. Higher education institutions in the developed countries like US, UK, Canada, etc., invest a lot on applications and resources on analyzing the student's academic performance and progression whilst developing countries still lack in touching up those grey areas. Education Data Mining (EDM) has taken a multitude dimension in the area of research and a myriad of data mining techniques have been applied to a variety of educational contexts. Several research works have been carried out to address the student attrition and prediction of academic performance but a solid recommender system is yet to be habituated among institutions. Major challenge for every tertiary institution is to identify the students at risk in the early stage and address them to improve their performance. EDM helps the stakeholders of institution to identify and overcome these problems by applying various data mining techniques. Academic performance of a student is largely measured based on the following parameters.

- Student demography,
- Educational history,
- Psychological data,
- Personal detail and
- Other environmental variable.

Several corporate organizations have come up with the high-end applications for behavioral modeling and predictive analysis. Applications like NLP Logix, IBM Watson, SAS onDemand for Academics are used in various universities to strengthen the academics by analyzing the behavior of students and predicting the performance. Nearly 70% of the institutions around the world use their academic analytics primarily for transactional reporting. Very few uses it for predictive modeling or what-if analysis [1]. Cost of high end applications slogs the institutions from using it and hence it's always advisable to develop a CRM with a recommender system which entails the analysis of the student demographic details, academic data, feedbacks about the instructors and institution, involvement in extracurricular activities and attendance to make a behavioral study and predict the performance.

In this paper a detailed systematic survey has been conducted to identify the various parameters which influence the performance of a student and different mining algorithms used.

II. METHODOLOGY

The reason behind this systematic review is to find out the influential parameters in predicting the student performance used by various researchers. A

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 2
February 2018

qualitative research is conducted by using SPIDER tool.

Developing a question is a critical step to effectively searching for research evidence. While the PICO (Population, Intervention, Comparison, and Outcome) tool has been a fundamental tool for evidence-based practice and systematic reviews, searching qualitative research is more problematic. The SPIDER tool, designed using the PICO tool as a starting point, has been created to develop effective search strategies of qualitative and mixed-methods research [2].

A questionnaire has been made based on the SPIDER Tool helps to develop effective search strategies of qualitative and mixed-methods research.

1. What is the objective of the research paper?
2. What are the dependent attributes?
3. Which are the selected attributes to perform the behavioral analysis and prediction?
4. What is the accuracy of various algorithms used in the prediction?
5. What are the different tools used for mining?

The above questionnaires are applied in the research papers which used Classification mining algorithms. During the survey it has been found that 90% of the prediction based research papers used classification mining algorithm.

III. Related Research Works

Amirah, WahidahNur'aini (2015) made a comprehensive systematic survey for identifying the most important attributes in a student's data. In the proposed systematic review, the study has been made to identify the gaps in existing prediction methods, variables used in analyzing the student performance and methods for predicting student's performance [3]. Satya and Sankar (2017) conducted a survey on different education mining prediction algorithms like Classification, decision tree algorithm, C4.5, Feature with Graph structure, Bayesian, RIPPER, SVM and compared the best performance. Satya and Sankar concluded that Naïve Bayes, C4.5, Ripper are the best suitable algorithm for finding the knowledge from datasets with utmost accuracy [4]. Pooja, Anil and Manisha (2015) made a comprehensive literature review of relevant researches done in last decade from year 2002 to 2014. The research was mainly focused on various areas in education field like predicting academic performance with pre/post enrollment factors, Comparison of data mining techniques used and correlation among pre/post enrollment factors and employability [5]. Romero and Ventura (2006) carried out a survey on the application of data mining to traditional educational systems, particular web-based courses, well known learning content management systems, and adaptive and intelligent web-based educational systems. Through this work it has been found that the recommendation agents proposed for e-learning systems are promising and can be widely used in the area of education mining [6].

IV. Attributes Influencing Academic Progression of Students

In this section we intensively discuss about various factors influencing the progression of students like:

a. Attributes used.

b. Accuracy of algorithms.

c. Tools used.

In total, 21 research papers have been reviewed. Reviewed research works have mainly used GPA as a dependent variable and all these papers have used Classification techniques for mining the data.Several attributes are used and the attributes are classified based on below parameter,

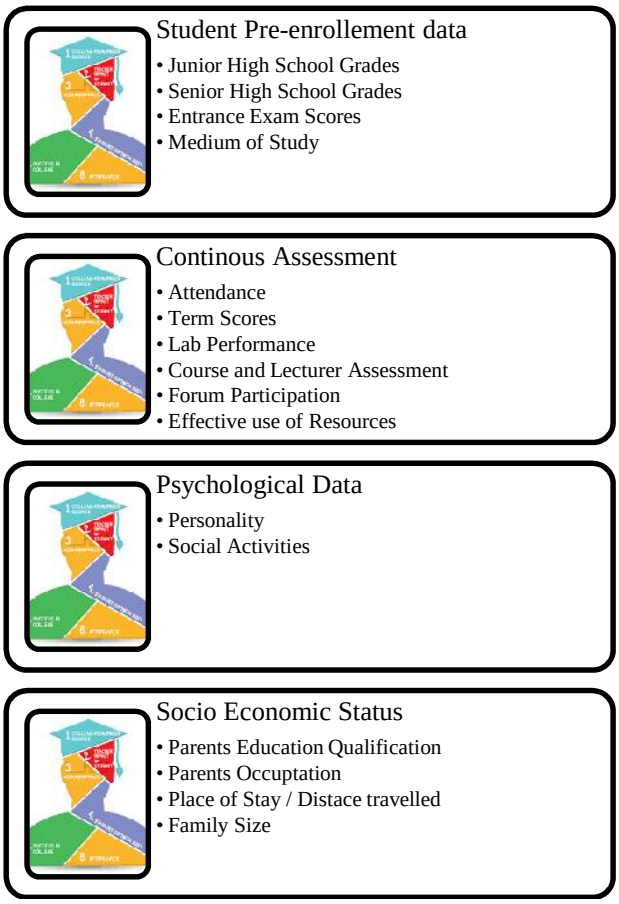
a. Student Pre-enrollment data:

This data plays a vital role in understanding the students past academic performance. Junior and Higher Secondary school results help us to understand the consistency in their performance and also provide an insight on their interest on various subjects. This data also includes any entrance exam scores taken by the candidate to apply for the program.

Twelve (12) out of Twenty One (21) research papers have used the student pre-enrollment attributes to predict their performance [7,10,11,16,17,18,20,22,24,25,26,27].

Fadhilah et al., [7] used Pre-enrollment data intensively to predict the academic performance of the students. University entry mode and School grade in selected subjects are the highly influential parameters of this research work.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 2
February 2018



b. Continuous Assessment:

Continuous assessment data is vital for the prediction of a student's performance. Class Tests, Assignments and Lab Practical Tests become an aid to prepare for the final exams. A continuous assessment not only helps the students to know their understandability in the subjects but also helps the institution to monitor and take necessary steps to improvise his/her grades for the semester. Grades are also calculated on the basis of attendance in the particular course. If a student is a regular absentee of the course, it adversely affects his grades.

Apart from the regular class tests and assignment another key factor to judge the involvement of a student is to analyze his participation in open forums and discussions. The magnificence of western education system is that it provides a liberty to the student to express his ideas and throw his views. The active participation in these events helps the student to get an exemplary knowledge on the course.

A series of continuous assessment is a preparatory phase for the final exams. Parameters to be considered during this phase are:

- Attendance
- Term Scores
- Lab Performance
- Course and Lecturer Assessment
- Forum Participation
- Effective use of Resources

Majority of the researchers have used continuous assessment as the key tool for the prediction. 18 out of 21 research papers have taken continuous assessment as an important tool. Few have considered almost all the parameters in the aforementioned whereas majority has taken parameters like Attendance, Term Scores and Course and Lecturer Assessment for the analysis.

Bo et al., [10] has taken Term Exam Scores, Final Term Exam, Average, Course Score and Attention as the parameters for his prediction analysis. Similarly, Brijesh K. &Saurabh Pal has used Test Grade, Seminar Performance, Assignment, General Proficiency, Attendance and Lab Work as their key factors in considering the performance.

c. Psychological Data:

Apart from the regular classroom activities, it is important to assess the student's involvement in extracurricular and social activities. Student's interpersonal communication should be monitored and taken into consideration while assessing his progression.

A student with a sluggish attitude might not be able to perform better and improve his grades, in such cases the student has to be counseled to better his performance.

Ashwin S. &Mariusz N [26] used multiple classifiers to improve the quality of student data which comprises of Personal, Socio-Economic and other environment attributes.

Similarly, six of the research works surveyed in this work has used Psychological data as the key parameter.[10, 11, 12, 16, 19, 24]

d. Socioeconomic Status:

Several researches indicate that students from low socioeconomic status develop academic skills slower than those from high socioeconomic status.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 2
February 2018

Parent's occupation and their educational background may also be a deciding factor for the performance of a student. It also involves the income status of the family and the resources available for the student to enhance his skills.

Location of residence is another key factor to judge the performance of a student. A student who travels from far distance to the college tend to focus less on his studies due to the tiresome journey he has to undertake every day.

Apparently, all the aforementioned points play a key role in assessment of a student's progression and among the papers surveyed, Twelve out of Twenty one research works have taken socio economic data as an important attribute [7,10,11,12,13,14,15,20,22,24,25,27]

V. Prediction Methods Used For Student Performance

Predictive modeling is a process that uses data mining and probability to forecast outcomes. In Education Data mining, Predictive modeling is used to predict the performance of students. A number of classification algorithms are used in predictive analysis and this survey paper has analyzed various algorithms used and its accuracy.

From the Fig 1, it's clear that among the other classification algorithms used in the prediction analysis, Naïve Bayes algorithm has given more accuracy.

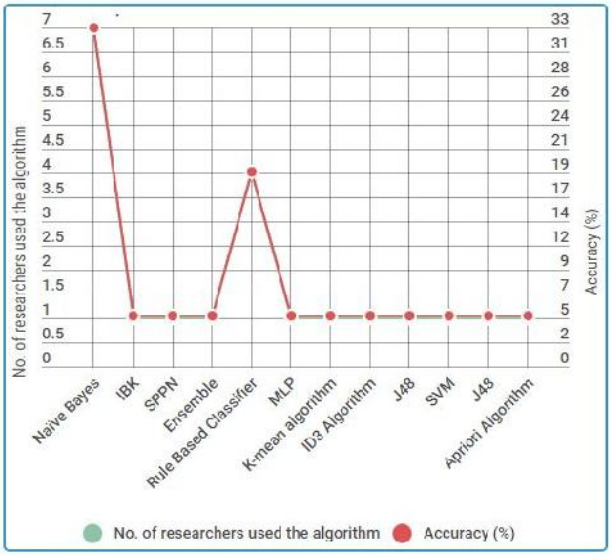


Fig 1. Comparison of Algorithms used in various research works and its accuracy

a. Naïve Bayes Algorithm:

This algorithm is based on Bayes Theorem which describes the probability of an event, based on prior knowledge of conditions that might be related to an event. A Naïve Bayes classifier considers each of the attribute to contribute independently to the probability. It represents a supervised learning method as well as a statistical method for classification. Bayesian Classification provides a useful perspective for understanding and evaluating many learning algorithms. It calculates explicit probabilities for hypothesis, and it is robust to noise in input data.

Predominantly known for the prediction based on related attributes, Naïve Bayes algorithm seems to be one of the robust classification algorithms to be considered or used for predicting academic progression.

Naïve Bayes algorithm is widely used in Recommender systems, Spam filtering, emotional intelligence and news classifiers.

b. Rule Based Classifier

Rule Based classifier makes use of a set of IF-Then rules for classification. A set of rules are created to derive the decision from the dataset. A Rule is expressed as follows,

If [Condition] Then [Conclusion]

If the rule is satisfied then it leads to another rule which in turn covers the subsequent tuples of a dataset. Few of the research papers in this survey paper has used the Rule based algorithm and received high accuracy rate comparing to C4.5, Random Forest and MLP.

c. SPPN (Students Performance Prediction System)

It is a proposed algorithm by Bo et al., [10] to predict student performance using emerging trend Deep Learning approach which is demonstrated to be a very effective method to predict outcomes with a high level of accuracy, especially when large data sets are available. SPPN involves millions of parameters to train, which require massive

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 2
February 2018

computation power. The learning can be made more efficient by using a layer-by-layer pre-training phase that initializes the weights sensibly. The pre-training also allows the variation inference to be initialized sensibly with a single bottom-up pass.

SPPN uses a six layer neural networks to implement deep learning algorithm. Bo et al., proposed model on a 120,000 students dataset with two Tesla K40 12GB GPUs, and the experimental results show the effectiveness and efficiency of the proposed method which can be applied into the academic pre-warning mechanism.

d. Ensemble Learning

Ensemble learning is a process by which multiple classification algorithms are used for better predictive performance. Zahyab et al.,[12] used a combination of classification models to predict the students who secured Good Honours degree and Not Good Honors degree. Bayesian Network, CHAID Decision tree and Logistic regression are combined using ensemble with confident-weightage voting. Comparing to other individual models ensemble shown the highest accuracy rate.

VI. Tools Used For Applying Prediction Algorithm

Quite a lot of open source tools are available for mining. These are machine learning tools which helps the researcher to analyze the dataset using various algorithms. These tools are widely used for predictive analysis, visualization and statistical modeling. Based on the survey conducted on various researches, it has been observed that WEKA is the widely used tool for predictive modeling. Fig 2. depicts the usage of tool by various researchers in the survey.

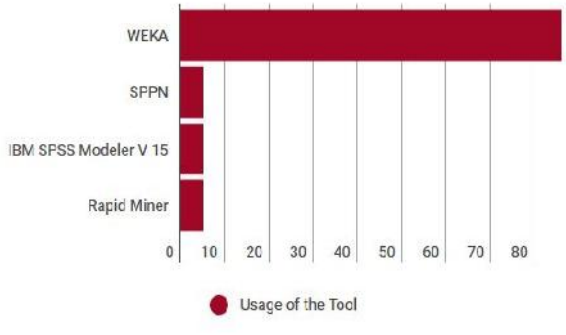


Fig: Usage of mining tool in various research works.

a. WEKA

WEKA is a collection of machine learning algorithms purposefully used for data mining tasks. WEKA has pre-built tools for data preprocessing, classification, regression, visualization and association rules. It's easy to use and it also allows the user to develop new machine learning algorithms.

From this survey paper, it's obvious that majority of the researchers are using WEKA for applying prediction algorithms. WEKA is an easy to use software and it allows us to implement various algorithms but it does not support big datasets for which programming is required.

b. IBM SPSS Modeler

IBM SPSS Modeler is an application used for data mining and text analytics. It is a product of IBM and widely used to build predictive modelers. Though IBM SPSS is a powerful tool comparing to WEKA, the latter being an open source it is widely used. IBM SPSS is an extensive predictive analytics platform which is designed to bring predictive intelligence.

c. Rapid Miner

Like IBM SPSS Modeler and WEKA, Rapid mine also provides an integrated environment for pre-processing, machine learning, predictive analysis and text mining. Comparing to WEKA, Rapid miner can handle large datasets and can produce effective results.

VII. Conclusion And Further Recommendation

Predicting student performance is an ongoing research work wherein several contributions are made from various researchers. This topic is still in the nascent stage in developing countries. Prediction of academic progression is highly evaluated from the student's high school performance, continuous assessment and demography. Universities in developed countries have started showing a significant improvement in analyzing a student data. Apart from analyzing the performance in high school and university, student's active involvement, behavior, social activities and financial status are to be analyzed. These attributes or data provides 360 degree information about a student and helps us to build a

recommender system which can predict the performance of the student more accurate.

Many universities in developing countries lack ERP solutions. Hence student information is stored in hard copy or in software which only stores the basic information. Universities in USA, UK and other developed countries keep track of student's accessibility to resources like library and online systems. These universities also monitor the student's relationship among his peers. Hence, developing countries should implement ERP solutions to gather the complete information of a student. Also, a recommender system has to be implemented to analyze the students' performance, and necessary measures have to be taken to improve their performance. Recommender system not only helps the institution to better the student performance but also help the institution in reducing the attrition rate.

Student pre-enrollment data, continuous assessment reports, psychological data and socio-economic data are all major attributes in predicting the performance of a student. Therefore, institutions should devise strategies to gather thisinformation and analyze it. From this paper, it is also understood that algorithms like Naïve Bayes, Rule Based or ensemble learning can be used for the predictive analysis with the help of mining tools like WEKA, IBM SPSS or Rapid Miner.

VIII. REFERENCES

[1] J. K. Jothi and K. Venkatalakshmi, "Intellectual Performance Analysis of Students by Using Data Mining Techniques," *International Journal of Innovative Research in Science, Engineering and Technolog*, Vol. 3, no. 3, pp. 1922-1929, Mar., 2014

[2] N. Shelke and S. Gadage, "A Survey of Data Mining Approaches in Performance Analysis," *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 5, no. 4, pp. 456-459, Apr., 2015

[3] M. S. Mythili and A. R. Mohamed Shanavas, "An Analysis of students' performance using classification algorithms," *IOSR Journal of Computer Engineering*, Vol. 16, no. 1, pp. 63-69, Jan., 2014

[4] A. D. Kumar and Dr. V. Radhika, "A Survey on Predicting Student Performance," *(IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 5, no. 5, pp. 6147-6149, Oct., 2014

[5] Bayer J., Bydzovska H., Geryk J., Obsivac T and Popelinsky L, "Predicting Drop-Out from Social Behaviour of Students," in *International Educational Data Mining Society, Paper presented at the 5thInternational Conference on Educational Data Mining (EDM)*, Chania, Greece, 2012, pp. 19-21

[6] C. Romero and S. Venutra, "Educational data mining: A survey from 1995 to 2005",*Science Direct – Expert Systems with Applications*, Vol. 33, no. 1, pp. 135-146, Oct., 2006

[7] F. Ahmad, N. H. Ismail and A. A. Aziz, "The Prediction of Students' Academic Performance Using Classification Data Mining Techniques", *Applied Mathematical Sciences*, Vol. 9, no. 129, pp. 6415-6426, Oct., 2015

[8] A. Mueen, B. Zafar and U. Manzoor, "Modeling and Predicting Students Academic Performance using Data Mining Techniques", *I.J. Modern Education and Computer Science*, Vol. 11, no. 1, pp. 36-42, Nov., 2016

[9] M. Ilic and M. Veinovic, "Students' success prediction using Weka tool", *INFOTEH-JAHORINA*, Vol. 15, no. 1, pp. 684-688, Mar., 2016

[10] B. Guo, R. Zhang, G. Xu, C. Shi and L. Yang, "Predicting Students Performance in Educational Data Mining," *in International Symposium on Educational Technology (ISET)*, Wuhan, China, 2016, pp.125-128

[11] T. Devasia, Vishnushree, V. Hedge, "Prediction of Students Performance using Educational Data Mining," *in International Conference on Data Mining and Advanced Computing (SAPIENCE)*, Ernakulam, India, 2016, pp.125-128

[12] Z. Alharbi, J. Cornford, L. Dolder and B. Iglesia, "Using Data Mining Techniques to Predict Students at Risk of Poor Performance," *in SAI Computing Conference, London, UK, 2016*, pp. 523-531

[13] M. S. Mythili and Dr. A. R. Shanavas, "An Analysis of students' performance using classification algorithms" *IOSR Journal of Computer Engineering (IOSR-JCE)*,Vol. 16, no. 1, pp. 63-69, Jan., 2014

[14] L. P. Khobragade, P. Mahadik, "Students' Academic Failure Prediction Using Data Mining," *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 4, no. 11, pp. 290-298, Nov., 2015

[15] S. F. Salim, Y. C. Kulkarni, "Precognition of Students Academic Failure Using Data Mining Techniques," *International Journal of Engineering Research and General Science*, Vol. 3, no. 3, pp. 507-513, Jun., 2015

[16] K. V. Kishore,Venkatramaphanikumar S, S. Alekhya, "Prediction of student academic progression: A case study on Vignan University,"

published inInternational Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2014, pp. 507-513, Jun., 2015

[17] M. M. Abu Tahir and A. M. El-Halees, "Mining Educational Data to Improve Students' Performance: A Case Study," *International Journal of Information and Communication Technology Research*, vol. 2, no. 2, pp. 140-146, Feb., 2012

[18] B. K. Bharadwaj and S. Paul, "Mining Educational Data to Analyze Students' Performance," *(IJACSA) International Journal of Advanced Computer Science and Applications*,,vol. 2, no. 6, pp. 63-69, Feb., 2011

[19] J. Bayer, H. Bydzovs, J. Eryk and L. Popelinsk, "Predicting drop-out from social behaviour of students," presented at *5thInternational Conference on Educational Data Mining (EDM), Chania, Greece*, 2012

[20] B. K. Bharadwaj and S. Paul, "Data Mining: A prediction for performance improvement using classification," *(IJCSIS) International Journal of Computer Science and Information Security*,vol. 9, no. 4, April., 2011

[21] P. Kavirpriya, "A Review on Predicting Students' Academic Performance Earlier, Using Data Mining Techniques," *International Journal of Advanced Research in Computer Science and Software Engineerin*,,vol. 6, no. 12, pp. 101-105, Dec., 2016

[22] W. Singh and P. Kaur, "Comparative Analysis of Classification Techniques for Predicting Computer Engineering Students Academic Performance," *International Journal of Advanced Research in Computer Science*, vol. 7, no. 6, pp. 31-36, Dec., 2016

[23] B. Guo, R. Zhang, G. Xu, C. Shi and L. Yang, "Predicting Students Performance in Educational Data Mining," *presented at International Symposium on Educational Technology (ISET)*, Jul.,2015

[24] K.P. Shaleena and S. Paul, "Data mining techniques for predicting student performance,"*International Journal of Computer Science and Mobile Computin*,vol. 2, no. 7, pp. 273-279, Jul., 2013

[25] E. Osmanbegovic and M. Suljic, "Data mining approach for predicting student performance," Economic Review – Journal of Economics and Business, vol. 10, no. 1, pp. 3-8, May., 2012

[26] A. Satyanarayana and M. Nuckowski, "Data Mining using Ensemble Classifiers forImproved Prediction of Student AcademicPerformance," presented at spring 2016 Mid-Atlantic ASEE Conference,April. 2016

[27] A. A. Saa, "Educational Data Mining & Students' Performance Prediction," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, pp. 212-220,, Jun. 2016