International Journal for Modern Trends in Science and Technology
Volume: 02, Issue No: 11, November 2016
http://www.ijmtst.com
ISSN: 2455-3778

Predictive Analytics in Education Context

Neha Kawchale[1] | Prof. Rachana Satao[2]

[1,2]Department of Computer Engineering, Smt.Kashibai Navale College of Engineering, Pune, Maharashtra, India.

To Cite this Article
Neha Kawchale, Rachana Satao, "Predictive Analytics in Education Context", International Journal for Modern Trends in Science and Technology, Vol. 02, Issue 11, 2016, pp. 77-82.

ABSTRACT

Education plays a vital role in nation's overall development process. To be effective, analysis must be timely and cope with data scales. The scale of the data and the rates at which they arrive make manual inspection infeasible. Predictive analytics can help and improve the quality of education by analyzing the historical data of the student and allow the decision makers address factors such as increased drop-out rate, fees structure in the upcoming years, unemployment, Recommender Systems for Professional Development and curriculum Development. This paper presents an analytical study of student progress report and help to plan accordingly to achieve success.

KEYWORDS: *Descriptive analytics, Predictive analytics, Academic analytics and Learning analytics.*

I. INTRODUCTION

Data mining is the process in which huge amount of generated data from any source in the form of online, offline; structured, unstructured, etc. data is analyzed from different perspectives and finally summarized to generate useful information. In order to generate such useful information a process called Analytics is used which generate meaningful patterns by analyzing the data. To make predictions about unknown future events, branch of the advanced analytics which is called as Predictive Analytics [8, 9] is used. The goals of predictive analytics are to produce relevant information, actionable insight, better outcomes, and smarter decisions, and to predict future events by analyzing the volume, veracity, velocity, variety, value of large amounts of data and interactive exploration.

Predictive analytics uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions about future. It uses a number of data mining, predictive modeling and analytical techniques to bring together the management, information technology, and modeling business process to make predictions about future.

The patterns found in historical data of student progress report can be used to identify the risks involved and the opportunities for future. Predictive analytics models capture the relationships among many factors to assess risk with particular set of conditions to assign a score, or weight age.

Predictive analytics allows institutions to become proactive, forward looking, anticipating outcomes and predict behaviors based upon the historical data and not on assumptions. It goes further and suggests actions to the people involved in the process of performance of the student benefit from the prediction and also provide decision options to benefit from the predictions and its implications.

A) Definition

1. A predictive model is simply a mathematical function that is able to learn the mapping between a set of input data variables, usually bundled into a

record, and a response or target variable. (http://www.ibm.com/developerworks/library/ba-predictive-analytics1/)

2. It is a process of extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive business decisions and actions. (www.accenture.com/sitecollectiondocuments/pdf/accenture-business-intelligence-and-predictiveanalytics.pdf).

3. Predictive analytics is the use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data. The goal is to go beyond knowing what has happened to providing a best assessment of what will happen in the future (http://www.sas.com/en_us/insights/analytics/predictive-analytics.html)

Finally, we can break predictive analytics into the following categories: data mining, predictive Modeling, pattern recognition and alerts, forecasting, and root cause analysis.

B) Terms used in Analytics

In order to understand Analytics in education context, the following variety of terminologies is used.

a) Descriptive analytics (what happened): It gains insight from historical data by reporting, scorecards, clustering, etc.

b) Predictive analytics (what will happen): It is the use of data, statistical algorithms and machine learning techniques to identify the likely hood of future outcomes based on the historical data collected.

c) Prescriptive analytics (what needs to done): It is the area of business analytics (BA) dedicated to finding the best course of action for a given situation. Prescriptive analytics is related to both descriptive and predictive analytics.

d) Academic analytics: An early academic analytics initiative which can be used for seeking student's academic difficulty across academics level. It can also help faculty members of institution and advisors to provide necessary and useful input to student by suggesting specific learning paths.

e) Learning analytics (academia) [3]: The use of data and models to predict student progress and performance and the ability to act on that information. It can also used to generate appropriate intervention by understanding the student behaviors in academics.

f) Action Analytics: It is an encompassing academic and administrative processes, and recognizing the need for transformative reinvention and reimagining focused more broadly on academic and administrative productivity and performance.

The goals of predictive analytics are to produce relevant information, actionable insight, better outcomes, and smarter decisions, and to predict future events by analyzing the volume, veracity, velocity, variety, and value of large amounts of data. Formally, we can define predictive analytics as follows:

C) Utilizing Predictive Analytics in Education

For any nation in the world, Education plays a vital role in a nation's overall development process which can be done by delivery appropriate syllabus to student. For many countries, "Quality Education" remains the basis for a sustainable and prosperous future. Several factors contribute to a quality education, including actuality and goal-based information;

Before utilizing predictive analytics in education, the following questions should be addressed:

• Which students are likely to drop out?
• What materials need to be referred clear competitive exams?
•What will be the cutoff for the particular college with course-wise particular years?
• Which students will be eligible and enroll in particular course programs?
• Which students will need assistance to graduate?
• What types of courses will attract more students as per industry needs?
•What will be the fees structure in the upcoming years?

Moreover, quality education will be value-oriented and will provide an understanding of industry needs. Acquiring all these factors in the same place to effectively develop successful and goal-oriented education systems is a difficult task the following factors are considered.

1 .Early Alert and Targeted Support Systems to Improve Student Persistence.

A common use of predictive analytics in education [10] is in the implementation of early alert systems that rely on data about student behavior to identify students at risk of dropping out and to target interventions that assist them in completing their course and/or program of study. For example, Purdue University's Course Signals

system predicts which students are at risk of performing poorly using data such as student demographics and academic history, engagement with online resources, and current performance in a given course. Based on the prediction [6], students receive an indicator - red, yellow, or green light -that delivers instructor feedback to the students about their performance as well as steps to take next steps.

As the report notes, by studying historical data, colleges can build profiles of students who are most at risk of not persisting and develop steps to intervene in a timely manner. However, when historical data are combined with data accessible after a student enrolls, even more powerful predictions can be made about a student's likelihood of persisting. The following risk factors are most prevalent among at-risk students:

- number of logins to the college's learning management system;
- level of self-confidence (assessed with a diagnostic tool);
- level of social integration to campus life (meeting with faculty outside of class,
- studying with a mentor or other students;
- club or student government membership, learning community participation);
- study skills (time management, prioritization).

Identifying students at-risk of dropping out is the first step in the process, but perhaps more important is the use of predictive analytics to identify and communicate the biggest contributors to that risk; in other words, to tell faculty and administrators why a student may be struggling in a class or program. The use of interactive, data visualization tools that identify the reasons why students may be at risk helps colleges develop targeted interventions to minimize that risk.

2. Predicting Student Outcomes

Predictive analytics [7] can be utilized to project enrollment and to identify at-risk students of dropping out or failing a class, it can also be used to predict student outcomes based on historical rates of student progress and success; the ethnic, socioeconomic, and academic composition of the student body; the availability of transfer slots at nearby four-year institutions; the growth or decline of industries in which students are employed and the relative strength of the economy.

Predicting student outcomes can help administrators set realistic, data-driven targets for improved graduation and transfer rates and educate legislators and the public at large about the factors that may affect progress toward those goals.

3. Recommender Systems for Professional Development

An effective recommender system for professional development would guide employees to skill building opportunities most suited to their needs and interests.

4. Text Mining

Where operational data are supported by information that is not easily quantified, such as information related to manual processes or processes that require human judgment, big data text analytics can be used to optimize operations. Actionable knowledge can be identified by deploying text analytics tools that mine information that may be accessible online, within an organization's local intranet, and embedded within organizational communications, potentially revealing complex patterns of interactions .Supporting the analysis of unstructured, not easily quantified data is a growing trend and allows organizations to process information at a greater rate than if done so manually. When used in conjunction with structured data, institutions can develop more accurate predictive models.

5. Student Placement Prediction.

Predictive analytics can inform classroom curriculum, direct resources more appropriately, and help measure teaching strategies in order to gauge which benefit students most. One of the more interesting potential applications of these findings would be the creation of an advising application that would allow high school students, parents, and/or counselors to enter data points for an actual (or theoretical) student and receive a predicted placement and grade, given the data on the student's past performance and state of readiness.

6. Recruitment and Enrollment Management.

Predictive analytics can help community colleges and other institutions of education to statistically identify likely students based on a variety of factors, including geographic location [5], anticipated program of study, ethnicity, socio-economic status, high school grade point average (GPA), and source of first contact. Predictive analytics can also help institutions

identify and target marketing materials to specific high schools that yield high proportions of students most likely to enroll (or that the college would most like to target for enrollment).

7. Other Opportunities for Predictive Analytics. Analytics can also be used to assist in the scheduling of courses and classrooms. With the advent of long-term student education plans, a new data source has been created that could be harnessed for scenario planning of course schedules for future terms.

D) Issues faced in Education system.

Following are several crucial issues faced in education systems.

1. **Drop-out rate –** Students now-a-days are unable to meet the scheduled plan for the academics
which results in increased dropping out from the current academic years for example unable to clear the exam?
2. **Resource utilization within an institute** - This requires gathering real-life data through a communications channel and having the facility to store this data in various forms and types. Identifying which data belong to a particular category and maintaining it in the relevant repository or database is another challenging aspect.
3. **Decision prediction** - University applicants have various options when it comes to the branch of study, courses, and programs in which they can enroll; this access to options creates confusion. Perfect prediction thus helps society by producing well-trained and skilled professionals to better serve societal needs.
4. **Missing Required Skill set for Recruitment** - Companies and recruiters still find it difficult to find students equipped with the skills these organizations require. So, there is a need for a quality education system that can help planners design a curriculum focused on the demands of the future- workforce.
5. **Need of standardization of education system** - It's difficult to obtain best practices in online courses hence, to promote quality in online education, standardization is unavoidable.

By analyzing current and historical facts to make predictions about the future, decision makers can take action and make decisions today to attain tomorrow's goals. Predictive analysis is one way to effectively address these issues.

Predictive analytics in educational systems comprises of following steps, as shown in Figure 1.

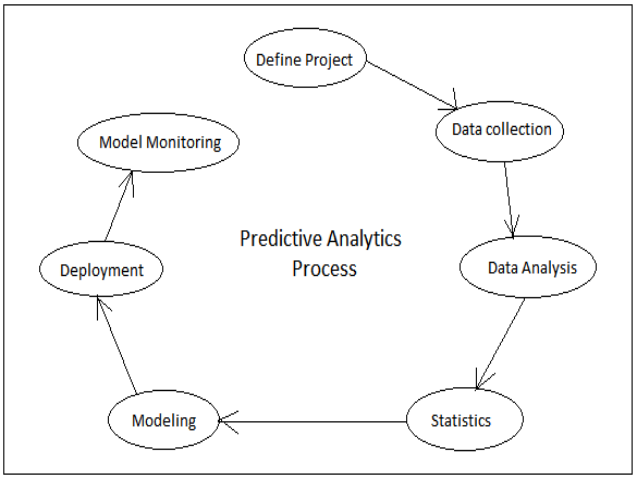

Figure 1: The main steps for predictive analytics in educational systems.

II. Literature Survey

Predictive analytics as a valuable tool [2] with which to engineer positive change throughout the student life cycle. As the cost to recruit a student rises, it becomes ever more important to retain students until they graduate, which will:

1. Improve student learning outcomes.
2. Improve retention and graduation rates.
3. Improve the institutional return on investment (ROI) on recruitment costs.
4. Increase operational efficiency.
5. Help the institution demonstrate success in a key area of focus for accrediting agencies and the Federal government.
6. Demonstrate positive efforts to other important entities (e.g., state legislatures that allocate funding to public schools, colleges and universities).

In the era of big data,[1,7] the challenges of predictive analytics include the quality of the data, because the prediction model's quality depends on it, the quantity of the data, because limited data provided during the training phase can make the analysis incapable of generalizing the derived knowledge when fed the new data; and the ability to satisfy analytical performance criteria—that is, results must be accurate and make statistical sense, and outcomes must be actionable—so that the analytics can identify the actual necessity for predicting an educational goal[4].

III. Proposed Work

In order to apply predictive analytics in education system, we examine the following questions.

We wanted to answer the following questions:

•What will be the cutoff for the particular college with course-wise?

• Which students will enroll in particular course programs?

• What is current demand, which programs are trending, [4] which are becoming obsolete?

•Which students are likely to drop out?

•What will be the fees structure in the upcoming years?

• What is the level of satisfaction of students in the current education system?

• Which students will need assistance to graduate (future grade prediction)?

To answer these questions, we applied predictive analytics. We based this work primarily on the first two categories: data mining and predictive modeling. We will execute predictive analytics according to the seven steps outlined in the below section.

Predictive Analytics Process.

Step 1: Define Project:

It is must and required step where we need to define the project outcomes, deliverables, scoping of the effort, business objectives, identify the data sets which are going to be used.

Step 2: Data Collection:

Data mining for predictive analytics prepare data from multiple sources for analytics. It is the first and foremost step in the predictive analytics process. To plan for quality education, it is necessary that any analysis collect huge amount of educational data and focus on gathering knowledge that can improve future prospects of student. Educational data combine offline data, online interaction data which can be structured and unstructured data. Offline data includes learner/educator information, students' attendance records, and student's progress report. Online and interaction data would be distance and Web-based education, computer-supported collaborative learning, social networking sites and online group forums, email, chat transcripts, and so on. Collected data must be preprocessed—that is, cleaned, transformed, and integrated.

Step 3: Data Analysis:

Data Analysis is the process of inspecting, cleaning, transforming and modeling data with the objective of discovering useful information arriving at conclusions.

Step 4: Statistics:

Statistical Analysis enables to validate the assumptions, hypotheses and test them with using standard statistical models.

Step 5: Modeling:

Predictive Modeling provides the ability to automatically create accurate predictive models about future. There are also options to choose the best solution with multi model evaluation. The following models are most widely used for analysis of data.

• decision tree—predicts categories of item class (supervised learning);

• clustering—discovers data clusters (unsupervised learning);

Step 6: Deployment:

Predictive Model Deployment provides the option to deploy the analytical results in to the every day decision making process to get results, reports, and output by automating the decisions based on the modeling.

Step 7: Model Monitoring:

Models are managed and monitored to review the model performance to ensure that it is providing the results expected.

IV. Conclusion

We have discussed steps that need to be considered for enhancing the quality of education to sustain a nation. We have studied the process to help student to achieve success in academics by understanding the huddles such as critical issues and challenges of education—specifically, technical education and the role that predictive analytics which can play in order to address these issues. Factors such as miniaturization of various sensors, improved logging and tracking of systems and improvements in the quality and capacity of both disk storage and networks should be considered in future.

Acknowledgment

This survey would not have been possible without the kind support and help of many individuals. I would like to extent my sincere thanks to all of them. I am highly indebted to **Prof. Rachana Satao**, Department of Computer Engineering of Smt. Kashibai Navale College of Engineering affiliated to Savitribai Phule Pune University for her guidance and constant

supervision as well as all the other staff members for providing important information regarding the survey.

REFERENCES

[1] Jindal Rajni and Dutta Borah Malaya, "Predictive Analytics in a Higher Education Context "IT Professional ,2015 IEEE Journals & Magazines Volume: 17, Issue: 4,Pages: 24 - 33, DOI: 10.1109/MITP.2015.68

[2] Shankar M. Patil "Predictive Analytics in Higher Education", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 12, December 2015

[3] A.V. Barneveld et al., "Analytics in Higher Education: Establishing a Common Language," Educause, Jan.2012; www.educause.edu/library/resources/analytics-higher-education-stablishing-common-language.

[4] KarthikeyanNatesan Ramamurthy, Moninder Singh, Michael Davis, J. Alex Kevern,Uri Klein, and Michael Peran ," Identifying Employees for Re-Skilling using an Analytics-Based Approach ",2015 IEEE 15th International Conference on Data Mining Workshops, pp. 345-354, DOI 10.1109/ICDMW.2015.206

[5] Matthew Malensek; SangmiPallickara; ShrideepPallickara, "Analytic Queries over Geospatial Time-Series Data Using Distributed Hash Tables ",2016 IEEE Transactions on Knowledge and Data Engineering,DOI: 10.1109/TKDE.2016.2520475

[6] JosepBerral;Nicolas Poggi; David Carrera; Aaron Call; Rob Reinauer; Daron Green ,"ALOJA: A Framework for Benchmarking and Predictive Analytics in Hadoop Deployments" ,Computing Year: 2015, Volume: PP, Issue: 99Pages: 1 - 1, DOI: 10.1109/TETC.2015.2496504

[7] P. Raj and G.C. Deka, "Big Data Predictive and Prescriptive Analytics," A Handbook of Research on Cloud Infrastructure for Big Data Analytics, IGI Global, 2014,pp. 370–391; doi:10.4018/978-1-4666-5864-6.ch015.

[8] Predicting the Future of Predictive Analytics, SAP report, Dec. 2013; http://tinyurl.com/ooygmtq.

[9] E. Sigel, "Seven Reasons You Need Predictive Analytics, "Prediction Impact, 2010; http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=YTW03080USEN.

[10] "Possibilities for Improving Student Success Using Predictive Analytics" by the RP group from Findings from an Environmental Scan of Predictive Analytics.