A Review Study on Predictive Analytical Tools and Techniques in Education

V. Belsini Gladshiya [1]
[1] Research Scholar,
Dept. of Computer Science, VISTAS,
Assistant Professor,
Department of Computer Science,
Agurchand Manmull Jain College, Chennai, India.

Dr. K. Sharmila [2]
[2] M.Sc.,M.Phil.,Ph.D
Associate Professor,
School of Computing Sciences,VISTAS.
Pallavaram, Chennai-6000117

India

Abstract:- Education plays an important role in each individual. The scope and value of higher education is challengeable. Using Predictive analytics, decisions can be made for future enhancement of students to improve their academic performances. Predictive Analytics is the analytics that answer the question what will happen in the future. The two words Analytics and Predict means analyzing the data and predicting the future.Using data and evidence transformation can be made by implementing predictive algorithms to predict the facts and risk factors in higher education. This paper surveys the problems and challenges and finds the solution using the technologies of Predictive Analytics to penetrate the facts and information in higher education.

Key words: Predictive Analytics, Higher Education, Predictive Algorithms.

I. INTRODUCTION

The process of collecting, organizing, analyzing large sets of data with useful information that helps the organization to take right decision is termed as Big Data Analytics. The advanced type of big data analytics is Predictive Analytics. This analytics process analyses the historic data and based on that it produces the future occurrences and events. It consists of different analytical techniques which are used to develop different algorithms that are used to predict the future occurrences.

A Predictive Analytical Model can be created using Predictive Analytics Modeling Techniques. Depending upon the data different Predictive Models can be used. Linear Regression, Logistic Regression, Classification, Decision Trees, Random Forest, Support Vector Machine, Graphical Models are some of the techniques used in Predictive Analytics. It deals with continuous and discontinuous data and its changes.

Predictive Analytics plays an important role in education to predict the student's activities to improve the success rate of students identifying the risk the students facing in their earlier stage of academics, environment, family and society.The approach and usage of data and evidence makes the analysis leads to deeper understanding of students of weaker sections to make decisions which transforms their success in many fields.

II. PREDICTIVE ANALYTICAL MODELING:

The Predictive Analytic Process can be done by three phases.

- Analysis
- Monitoring
- Prediction

Analysis:
This phase deals with collecting the preprocessed raw data and transforms the data by analyzing to develop a new model that can be easily handled.

Monitoring:
In this monitoring phase the data can be observed and the learning methods can be created using the transformed data.

Prediction:
The Prediction Phase gives the prediction reports to the users using the learning models created with previous data.

Fig:1 Phases of Predictive Analytics

Steps to build a Predictive Analytical Model:
To build a Predictive Analytical Model seven steps are included. They are

1. Understanding the Objective
2. Project Definition.
3. Data Collection.
4. Data Analysis.
5. Data Modeling.
6. Data Evaluation
7. Deployment

1. Understanding the Business Objective:
The first step in modeling a Predictive Analytics is to understand the objective of the business. (i.e.) asking the right question to define the goals clearly based on the objective. Some of the Business Objectives are

- Target Marketing
- Risk Management
- Strategy Implementation

- Operational Efficiency
- Increase Customer Experience
- Social Media Influences.

   Based on the business objectives the decisions should be made.

2. Defining the Project:

The Second step is defining the project that means defining the goal of the model. The objective of the model should be changed or transformed in to analytics to attain its goal. Some of the analytical terms are

- Profile analysis
- Segmentations
- Response Modeling
- Risk Modeling
- Net Present Value(NPV)
- Customer Life Time Value(CLTV)

3. Data Collection:

In this step the data for modeling should be collected. The data selected for modeling should have thorough understanding about the market, business and the objective. The data may be from internal sources like customer data, student data, transaction data and also the data from external sources like survey data, Research data, on line data etc.

4. Data Analytics:

The data collected should be prepared in a right format for analysis in this step. The analytical tool should be selected for analysis. The variable should be defined and data dictionary is created. The multiple data sets can be created and joined and validation process should be done for checking the correctness of the data and a basic summary of report should be produced according to it.

5. Data Modeling:

In this data modeling step the analyzed data is transformed into a model for prediction. Once the data is modeled it should undergo an univariate analyses to check the distribution of each variables and features and multivariate analyses to check the relationship with other variables and dependent variables. Based on the model type the variables will be transformed using the approach such as

- Binnig approach
- Transformation
- Polynomial
- Square Root
- Outlier Treatments
- Missing Value Treatment
- Dimension Reduction.

6. Data Evaluation:

The Modeled data is used for evaluation during this step. Using algorithms the data is validated and checked for multicollinearity and redundancies. The performance of the model is also checked. The robustness and stability of the model is also checked in this step.

7. Deployment:

The process of bringing the resources into effective action is called deployment. After evaluating the data it can be implemented. The model can be implemented within a system or it can be a real business data. After implementation an user acceptance test (UAT) can be performed to ensure the implemented model is correct and the outcomes are synchronized with desired outcome for prediction.
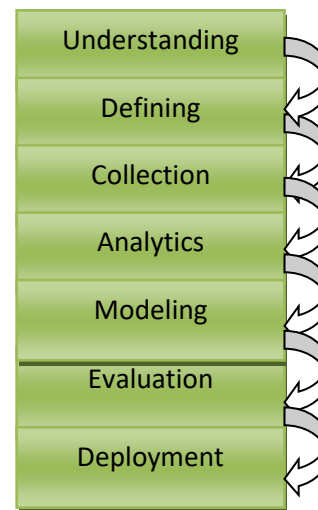

Fig 2: Predictive Analytical Model

III. ANALYTICS FOR EDUCATION SECTOR:

It is now becoming essential to analyze data for the development of both learning and academic activities while taking consideration in education sector. There are two broader analytics categories in education sector namely Learning Analytics (LA) and Academic analytics.

1. Learning Analytics (LA):

Learning analytics hold the promise of improving learning efficiency and effectiveness in primary, secondary, and post-secondary education.LA are directed toward providing educators, learners, and decision makers with actionable insight to classroom and course level activities It reduces attrition through detecting the risk of students earlier and generating an alert for learners.[1].

General Morphological Analytics (GMA):

General Morphological Analysis (GMA) is a method for structuring and investigating the total set of relationships contained in multi-dimensional, non-quantifiable, problem complexes. It was originally developed by Professor Fritz Zwicky, the Swiss astrophysicist and aerospace scientist based at the California Institute of Technology. [2]

Knowledge Discovery and Data Mining (KDD):

Knowledge discovery (KDD) is an interdisciplinary area focusing on methodologies for identifying and extracting useful and meaningful patterns from large data sets. KDD draws upon research in statistics, databases, pattern recognition, machine learning, data visualization, optimization and high performance computing. [3]

Factor analysis (FA):

Factor analysis is also a common approach in education, its goal is to find variables that can be naturally

grouped together, or splitting the set of variables (as opposed to the data points) into a set of latent (not directly observable) factors. [4]

2. Academic Analytics:

Academic analytics combines select institutional data, statistical analysis, and predictive modeling to create intelligence upon which students, instructors, or administrators can change academic behavior. [5]

Scholarship of Teaching and Learning Analytics:

The scholarship of teaching and learning (SoTL) is an action research to influence change within the classroom and among learners. [6]

Social Network Analysis:

It analyses the student-to-student and student-to-teacher relationships and interactions to identify disconnected students. **[7]**

Visual data analysis:

Visual data analysis blends highly advanced computational methods with sophisticated graphics engines to tap the ability of humans to see patterns and structure in complex visual presentations (Johnson et al. 2010). Visual data analysis is designed to help expose patterns, trends, and exceptions in very large heterogeneous and dynamic datasets collected from complex systems. [8]

Institutional analytics refers to variety of operational data that can be analyses to help with effective decisions about making improvements at the institutional level. Institutional analytics include assessment policy analytics, instructional analytics, and structural analytics. They make use of reports, data warehouses and data dashboards that provide an institution with the capability to make timely data driven decisions across departments and divisions. [8]

Information Technology Analytics:

Information technology (IT) analytics covers usage and performance data which helps with monitoring, required for developing or deploying technology for institutional use, developing data standards, tools, processes organizational synergies and policies. Information technology analytics largely aims at integrating data from a variety of systems such as student information, learning management, and alumni systems, as well as systems managing learning experiences outside the classroom. [9]

IV. PREDICTIVE ANALYTICS IN EDUCATION – SURVEY

Predictive analytics is the roof of advanced analytics which is to predict the future
Events. Predictive analytics is capsule with the data collection and modeling, statistics and
Deployment. Kavya.V, Arumugam.S (2016).

Predictive Analytics provide institutions with better and actionable insights based on data. It is used to help the students who are in risk behavior early in the semester .Ben Daniel (2014).

Shikha Anirban (2014) .It is essential to analyze data for the development of both learning and academic activities.

Tulasi (2013) stated Academic Analytics will be essential component for future. Adaptation to analytics will lead the institution to be more intentional, intelligent with data and evidence.

Predictive analytics is used to determine the probable future outcome of an event or the likelihood of a situation occurring. Nichol Mishra, Dr.Sanjay Silakari (2012).

The focus on academic analytics is the actions that can be taken with real time data reporting and with predictive modeling helps to suggest the out comes from familiar patterns of behavior. Paul Barpler, Cynthia James Murdoch (2010).

JAMES OGUNLEYE (2004) People, tools and algorithms are at the heart of predictive analytics and key to activate it. As a concept, predictive analytics combines human skills and expertise with technology, machine learning of patterns in current and historical data and the application of algorithms not only to see or identify patterns in the data but also to forecast future probabilities of the outcome of those patterns.

One of the major goals of Predictive Analytics in education is making the prediction of a student in his early stage in the semester before the student has fallen into a course material. Ryan S. Baker, David Lindrum, Mary Jane Lindrum, David Perkowski.

Learning Analytics in Higher education case study in Nottingham Trent University states that "In one of the most prominent learning analytics initiatives in the UK, Nottingham Trent University (NTU) has implemented an institution-wide dashboard to enhance the academic experience of its 28,000 students by facilitating dialogue between students and staff. The specific goals were: to enhance retention, to increase a sense of belonging to a community, and to improve attainment."

The case study from Marist College indicates the transfer of predictive models to other institutions. The Open Academic Analytics Initiative (OAAI) led by Marist College, a liberal arts institution in New York State, developed an open source early alert solution for higher education. The predictive model was successfully transferred to different institutions, and intervention strategies to help at-risk students were evaluated.

The guiding practice(3) of Predictive analytics in Higher education shows models (showing how different data points are related) and algorithms need data to build predictive tools that will support enrollment efforts or help students make academic progress. The guiding practice(5) states Adding predictive analytics to the student success toolbox may spark a culture change as interventions informed by data become central to your institution. To get the campus to embrace this change, it is important to communicate how faculty, staff, and students will benefit from using interventions that are informed by predictive analytics, and allow them to guide the change as well. (Manuela Ekowo, Iris Palmer)

V. ALGORITHMS USED FOR PREDICTION IN EDUCATION:

In data mining many methodologies are been used according to the related problems in which one should have the in-depth and complete knowledge of the problem according to that one could decide the technique and algorithm to be applied on the dataset. The most important methodologies used are Classification, Prediction, and Clustering.
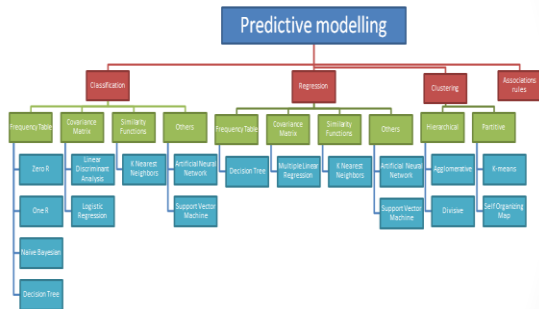

Fig 3: Families of algorithms used in predictive modeling

1. Weka tool of data mining, k-fold and association rule mining technique -Sumam Sebastian in June 2015

Sumam used 300 real data of the students to predict the result using these algorithms on the basis of artificial neural network under data mining.

2. ZeroR algorithm of classification, DBSCAN algorithm of clustering -S. Aher and L.M.R.J. Lobo in 2011.

3. **Regression analysis and Bayes method, decision tree under j48 method** -L. Romdhae, N. Fadhel and B. Ayeb in 2010**.**

The author showed that, to predict student success rate used data mining Regression method is used and Decision tree is used to predict the result.

This algorithm is also used to predict the result of the students.

4. Decision tree algorithm -R. R. Kabra and R. S. Bichkar.

They used this algorithm to predict the performance of engineering students. They used decision tree algorithm technique to predict student performance. of around 340 students during their first year exam. The accuracy of model generated was 60%.

5. IBK algorithm of classification.- Milos Ilic and Petar Spalevic

They used two different training sets as basis and generated two different types of models and used two different test sets and compared the two algorithms and showed the accuracy of prediction.

VI. OPEN CHALLENGES:

From the survey it is understood that a predictive techniques, models algorithms had been proposed for predicting the future probabilities in education such as predicting the student result, learning analytics, institutional analytics, relationship between students and teachers etc. But there are some other challenges focused on education to do with predictive analytics.

- It is essential to identify the correct patterns for suitable data will be an open challenge in Predictive Analytics.
- Finding suitable analytical techniques or tool and using of correct data are the challenge in Predictive analytics.
- A suitable data set should be used and examine all the assumptions and possibilities should be considered and according to it the analysis should be retraced.
- To identify students at risk of failing a course or program of study.
- Finding methods of making operative learning process, classifying learner's strength and weakness, outlining academic activities helps to predict the student's future performance.
- Identifying the risk factors and analyzing them with correct predictive models, appropriate data and suitable algorithm to give suggestions to get the students success in their academics. This is also considered as an open challenge in predictive analytics in the field of education.

VII. CONCLUSION

Analytics in education will lead to two different innovations as sustainable and disruptive. Sustainable means it improve the existing system and Process to a well fined proposed system. Disruptive means creating new ideas and activities which bring changes in behavior or process.

Hence as research is considered it is important to create new ideas that makes creative models, techniques and algorithms with real data set to predict the future probabilities of the students by finding their risk factors in their education and makes them to get success in their academics which makes their future life turns into success.

REFERENCES:

[1] George Siemens & Dragan Gasevic,Athabasca, , Canada Caroline Haythornthwaite & Shane Dawson, Simon Buckingham Shum & Rebecca Ferguson, "Open Learning Analytics: an integrated & modularized platform Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques." July 28, 2011

[2] T. Ritchey (Ed.), "General Morphological Analysis—A general method for non-quantified modeling."*Wicked Problems— Social Messes*. Retrieved from http://www.swemorph.com/pdf/gma.pdf

[3] Ben.K.Daniel. "Big Data Analytics in higher Education: Opportunities and Challenges" Article in British Journal of Educational Technology Sep 2015.

[4] . Kline, P. "An Easy Guide to Factor Analysis." (1993).

[5] Paul Barpler *University of Minnesota*, baepl001@umn.edu , Cynthia James Murdoch *University of Minnesota* "Academic Analytics and Data Mining in Higher Education", , macal001@umn.eduInternational Journal for the Scholarship of Teaching and Learning Volume 4 | Number 2 Article 17 7-2010

[6] Cook, C. E., Wright, M., O'Neal, C. "Action research for instructional improvement: Using data to enhance student learning at your institution", (2007). To Improve the Academy, 25, 123-138.

[7] Marie Bienkowski, Mingyu Feng U.S. Department of Education, **"**Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief" , Barbara Means Center for Technology in Learning SRI International October 2012

[8] Johnson, L., A. Levine, R. Smith, and S. Stone. Austin," Institutional Analytics: The 2010 Horizon Report". TX: The New Media Consortium. http://wp.nmc.org/horizon2010/

[9] Ben Daniel and Russell Butson "Foundations of Big Data and Analytics in Higher Education", Higher education Development Centre, University of Otego, Dunedin, New Zealand

[10] L. Romdhae, N. Fadhel, B. Ayeb "An efficient approach for building customer profiles from business data", Expert System with Applications, vol. 37, 2010, pp. 1573-1585.

[11] S. Aher" Data mining in educational system using weka", , L.M.R.J. Lobo International Conference on Emerging Technology Trends,2011, pp. 20-25

[12] Sumam Sebastian, M-Tech Computer and Information Science College of Engineering ,Poonjar Jiby J Puthiyidam, Assistant Professor Dept. of Computer Science and Engineering College of Engineering Poonjar, "Evaluating Students Performance by Artificial Neural Network using WEKA", International Journal of Computer Applications (0975 – 8887) Volume 119 – No.23, June 2015 36

[13] R. R. Kabra ,S.G.R. Education Foundation's College of Engineering and Management, G. H. Raisoni ,College of Engineering and Management "Performance Prediction of Engineering Students using Decision Trees*"*, , Pune, December 2011

[14] Milos Ilic, Petar Spalevic , "Students' success prediction using Weka tool" Electrical and Computing Engineering University of Pristina, Faculty of Technical Science Kosovska Mitrovica, Serbia, Mladen Veinovic,Wejdan Saed Alatresh Singidunum University Belgrade, Serbia, INFOTEH-JAHORINA Vol. 15, March 2016.

[15] Nishchol Mishra, Dr.Sanjay Silakari, Predictive Analytics: A Survey, Trends, Applications, Oppurtunities & Challenges, Nishchol Mishra et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 , 2012, 4434- 4438, ISSN : 0975-9646

[16] James Ogunleye,"The Concepts of Predictive Analytics",Middlesex University, United Kingdom, Received 20 June 2014; received in revised form 1 November 2014; accepted 20 November 2014.

[17] Priyanka Dhamija, Rainu Nandal, Harkesh Sehrawat, "A Review Paper On Prediction   Analysis: Predicting Student Result on the basis of past result. International Journal of Engineering and Technology (IJET, ISSN (Print) : 2319-8613 ,ISSN (Online) : 0975-4024

[18] B.Tulasi Assistant Professor Department of Computer Science, Christ University, Bangalore, India,"International Journal of Computer Applications" (0975 – 8887) Volume 68– No.14, April 2013.

[19] Shikha Anirban,Senior Lecturer, Department of Software Engineering,Big Data Analytics in the Education Sector: Needs, Opportunities and Challenges,International Journal of Research in Computer and Communication Technology, Vol 3, Issue 11, November – 2014,ISSN (Online) 2278- 5841,ISSN (Print) 2320-5156

[20] K.Sharmila, Dr. S. A.Vethamanickam," Survey on Data Mining Algorithm and Its Application in Healthcare Sector Using Hadoop Platform", International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 5, Issue 1, January 2015) R EDUCATION

[21] Manuela Ekowo, Iris Palmer, "Predictive Analytics in Higher Education", Five Guiding Practices for ethical Use.PR

[22] Iciactjournals.in applying big data in higher education A case study.

[23] Louis Columbus "the needs, opportunities and challenges of implementing big data analytics in the education sector" from Power of Big data in education.

[24] Julius Murumba, Elyjoy Micheni, "Big Data Analytics in Higher Education: A Review", The International Journal of Engineering and Science (IJES),|| Volume || 6 || Issue || 6 || Pages || PP 14-21 || 2017 ||,ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805.

[25] Kavya., Arumugam.S, "A Review on Predictive Analytics in Data Mining" International Journal of Chaos, Control, Modeling and Simulation (IJCCMS) Vol.5, No.1/2/3, September 2016,DOI:10.5121/

[26] MeryemOuahilal, Mohammed El Mohajir, Mohamed chahhou, BadrEddine El Mohajir,"The applications of Predictive Analytics: Benefits, Challenges and how it can be improved.(2016)

[27] www.amazonaws.com,rstudio-pubs-ststics3.

[28] Baesens,"Analytics in a Big Data world: The essential Guide to Data Science and its Applications", Wiley India Private Limited, 2014.

[29] Nishchol Mishra, Dr.Sanjay Silakari,Predictive Analytics: A survey, trends, Applications, Opportunities & Challenges, Nishchol Mishra et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012, 4434- 4438

[30] James Ogunleye, "The Concepts of Predictive Analytics", International Journal of Knowledge, Innovation and Entrepreneurship,Volume 2 No. 2, 2014, pp. 82—90