# **EDUCATION ANALYTICS – REPORTING STUDENTS GROWTH USING SGP MODEL**

# Akshay Naik<sup>1</sup>, Paresh Patil<sup>2</sup>, Neha Roygupta<sup>3</sup>, Nikhil Mudhalwadkar<sup>4</sup>, Pankaj Roy Gupta<sup>5</sup>, Vilas Mankar<sup>6</sup>

<sup>1</sup>Dept. of Computer Engineering, International Institute of Information Technology, Pune, India <sup>2</sup>Dept. of Computer Engineering, International Institute of Information Technology, Pune, India <sup>3</sup>Dept. of Information Technology, International Institute of Information Technology, Pune, India <sup>4</sup>Dept. of Computer Engineering, International Institute of Information Technology, Pune, India <sup>5</sup>Chairman & Managing Director, Parardha Infotech Pvt Ltd, Pune, India <sup>6</sup>Dept. of Information Technology, International Institute of Information Technology, Pune, India

#### Abstract

Every part of the education sector is struggling to produce actionable data favorable for their growth. The primary stakeholders of this sector are unable to take effective and productive decisions as the huge amount of data collected is not being processed properly. A lot of striking data are lost in the process as there are no schemas available for extracting the intelligence from them. Various external factors affecting the student's growth are not identified and thus the parents and teachers fail to understand the real reason behind the student's performance. Hence to measure student's growth SGP (Students Growth Percentile) can be used. It is also necessary to keep tab on student's future marks so as to take precautionary measure in case of negative growth. Here, time series analysis and forecasting can be used. Regression is use to calculate impact of any external factors on overall performance. When all these identified external myriad data along with the academic data is captured, processed using analytical models such as R. The stakeholders will be able to understand the core reasoning behind progress rate and thus take decisions accordingly. This is the fundamental idea behind Education Analytics.

Keywords - Analytics, Education Analytics, Student's Growth Percentile, Marks Forecasting, Student's growth

\*\*\*\_\_\_\_\_\_

# **1. INTRODUCTION**

The paper attempts to develop a plug-in for an ERP System that will provide Analytical Solution to the Education Sector by calculating Student Growth Matrix on multiple transactional as well as external factors. The project will enable the stake holders to take better and reliable decision based on the reports generated. The data generated in school usually does not result in any meaningful evaluation. Parents, students as well teacher take a look at the progress card, calculate the percentage and decide upon the rank. None of the schools calculate the growth of student with the help of proven mathematical and scientific techniques. And thus students are left wondering where they actually lack. One interesting part of the student life is comparing one's marks with the topper or with friends, etc. Sometimes, there exists comparison with students belonging to other educational board. However, this comparison is irrelevant. This paper gives a relevant method to calculate the student's growth mathematically, comparing with peers having same intellectual level. Few years back, schools lacked the digital facilities like computers and an ERP system. It was therefore difficult to make data digital. This was the most important issue while proposing analytics on school data. As ICT improved, more and more schools upgraded their data on a digital platform which made it easier to extract intelligence from data.

# **1.1 Analytics**

Analytics deals with analyzing both the enterprise data along with the external data thus making the intelligent information being extracted out to be more efficient and effective for the decision making process. Analytics is strongly associated with predictive analysis. Analytics and its predictive modelling functionality when explored in the educational domain will result in a new revolution in the Education Industry.

#### **1.2 Preliminaries**

The "Education Analytics Student Growth Model" tries to capture all possible factors affecting a student's growth and create a model to analyze the positives and negatives affecting the growth so as to provide better decision making power to the stakeholders absolutely based on data and facts. The solution that is being proposed is based on the availability of an ERP system as a pre-requisite. There should be a system in place which is capturing the data from all relevant aspects of all stakeholders. This module which provides insights about student's performance to the stakeholders takes input data from an ERP system implemented in the school. There is always a unique trend that would be found in every student's performance. To capture that trend huge amount of data is needed for each student. The scale of data for each student should be large in order to make credible and reliable decisions. Since the system is based on the idea of improving student's appraisal and then taking right decisions to move forward, the results are needed to be converted into visually palpable for the stakeholders like parents, teachers, management, etc. There is a questionnaire to capture other external factors such as behavioral factors, mindset of the student, confidence level of the student, etc. The correlation between these factors and the student's performance is found and informed to the stakeholders to make better decisions for the student's future.

#### **1.3 Forecasting**

Forecasting is predicting the trend in future based on some statistical technique. It requires capturing of different components of data sets like seasonality, irregularity, trend and cycle. These components are applied to a suitable model that suits the context of the data. Time series forecasting is done in case we have data points in increasing chronological order over a time period. Time series analysis is different from regression as in case of regression the temporal aspect is not captured. This is an important feature in the assessment of a student since it quantifies the expectations that are fair and achievable with the abilities he/she possesses.

#### 2. EXISTING SYSTEM

Existing system, which is used in most of the part of the world, gives a snapshot of the student's performance in the last exam. There can be myriad reasons that he/she has excelled or flunked the exam. So the subject where relatively lesser marks were scored does not mean that the student is weak in that subject. This cannot be judged based upon single test. Nevertheless, the human tendency of making connections overlooking the credibility of the inference they derive, parents incorrectly analyze the wards capabilities. This system fails to provide in-depth analysis of the student's performance. This system lacks credibility in reporting and does not help parents to take necessary steps for their ward's development. To increase the credibility one must consider previous performances as well. The previous grades give a fair insight about the child's competence in all different subjects. Another shortcoming of the existing system is that it tends to compare every student with same yardstick. It fails to understand that every child has different set of capabilities and it compares tomatoes with watermelons.

# **3. PROPOSED SYSTEM**

In the proposed system, the previous years' marks are taken into consideration along with the current grades. As there is huge amount of data, different statistical techniques can be applied for better understanding different aspects of student's potential. By taken into account the preceding grades one can make credible assessment about the students' academic and overall performance. Another new feature in the proposed system is to group the students in cohorts. The cohorts are based on the previous grades. The students are group on the basis of their intellectual level. Here the students that are not so good academically are placed in the same group. Similarly students who excel in academics are placed together. Having segregated the students into cohorts, SGP (Student Growth Percentile) is calculated for each cohort. It indicates the position of the student in the cohort. This way the students can know how they are doing with respect to others with similar capabilities. This division of students by no means demeans the academically weaker section, but gives the opportunity to assess their performance with similar students and as being compared to similar ones is a reasonable competition for each student and will drive the student to try harder.

#### 3.1 System Architecture

The data to be collected for our module will belong to either structured, semi structured or unstructured format. The data will be gathered from the ERP System and other sources like appraisals by students, teachers and parents, surveys, MBTI personality test by the students, etc. The data to be used for analytics must be in structured format. However, the data collected may be in any format. Thus, it is necessary to convert all data in a structured format as specified in our model. The fields that are irrelevant with respect to our model ought to be dropped. These data are to be fed into the mathematical model as per the requirement of each growth model. Each factor in the model will be assigned some weight based upon the level of role it plays in the student's growth. This mathematical model will calculate the weights and generate growth matrix for each student. This will be presented to the parents, teachers and the administrators as per their requirement and interests. The reports generated by the growth model will be represented using graphs, charts or visual reporting aids that are easily understood by the stakeholders.



Fig.1 System Architecture

# 3.2 Methodology

# 3.2.1 Time Series Forecasting

This paper tries to help students to know their own capabilities by forecasting their scores based on the previous years' marks. This way the student will know how he could perform and set a standard for the upcoming exam. It will also help teachers and guardians to assess the student's growth trend. The forecasting is done on the basis of time series analysis. Predictions are done on different verticals like exam-wise marks, yearly, subject-wise, etc. to get extensive view from each angle. To do this the marks or grades are arranged into time series. The time series can be quarterly or half-yearly or every four months, it depends on the amount of data gathered and frequency of exams held. Once the data is gathered, a suitable model is chosen that fits the data and it is used for forecasting. The forecast () function from forecast package authored by Rob Hyndman in R has been used for the purpose. It automatically detects suitable model based on the data set given as an argument. It separates different components like seasonality, irregularity, trend and cycle.

# 3.2.2 Student Growth Percentile

Student Growth Percentile is calculated by segregating the students with same intellectual level and finding percentile amongst that group. The groups formed are called cohorts. Cohorts can be formed based on the previous performance pattern of the student. Consider a group of schools, following same education methodologies, across the state/region. The students' previous two or three years' marks are taken into consideration. Those students with similar performance pattern are grouped together. The scheme of making cohorts depends on the number of students of same class that are available. After cohorts are made, SGP is calculated using the formula...Equation 1....

Percentile Rank = (Number of students below Score + (.5 \* Number of students at Score))/ (Number of students in the academic peer group)

# 3.2.3 Weakness and Strengths

To point out weaknesses and strengths is very crucial task for a student so that he can take corrective actions on time for the weaknesses and enhance the strengths from beginning and develop it to be best about that. Weaknesses can be found by analyzing previous performances in the different subjects. The subject in which one has scored less marks consistently becomes his/her weakness. Same strategy can be applied for discovering strengths. The subjects in which he/she has scored considerably high consistently can be said to be his strong hold.

# **3.2.4 External Factors**

Various other factors also affect the growth or performance of the student. The identified factors affecting student's performance are as follows: - -

SR.	DOMAIN	FACTORS	NO. OF	
NO.	DOMAIN	FACTORS	FACTORS	
1	Academic Achievement	Academic Grades, Co-Curricular Participation, Achievements, Skills, Aptitude, Knowledge, After School Coaching	7	
2	Background Characteristics	Family Income, Home Environment, Parental Control over Child, Mental Health of Parent, Family Size, Parent Education, Parent Occupation, Parenting Style, Sibling, Guardian/Guide,	9	
3	Perception of Self	Idol/ Role Model, Strengths, Weakness, Interests/Inclination, Adaptable, Self-Management, Moral Values, Mental Stability	8	
4	Institutional Information	Location, Infrastructure, Faculty Quality, Attendance, Past Grades	5	
5	Faculty Contribution	Time Management, Power of Concentration, Discipline, Participation in Class/ Enthusiasm	4	
6	Extra - Curricular and Co- Curricular Activity	Participation, Achievement, Volunteering	3	
7	Personality	MBTI Traits, Optimistic/Pessimistic, Agreeableness, Conscientiousness, Neuroticism, Openness to Experience, Extraversion	7	
8	Social Factors	Peer Pressure, Friend Circle, Learning Environment at Home, Attachment, Interpersonal Relation, Religion	7	
9	<b>Biological Factors</b>	Age, Gender, General Health, Mental Health, Maturity Level	5	
10	Capability/ Qualities	Memory, Reasoning, Perception, Analytical Skills, Creativity, Leadership, Communication Skills, Confidence	8	
11	Behavioral Pattern	Reading Interest, Inquisitive, Attentiveness, Behavior, Aggressiveness, Grasping Power, Learning Speed, Response to Situation, Decision Making Power	9	
12	Travel	Travel Time from Home to School, Travel Time from School to Back Home, Distance of Home from School, Travel Mode	4	
		TOTAL	76	

# 4. RESULTS AND EXPERIMENTAL SECTION

 Table 2: Student Marks table format for Academic year 2011

	Exam					Social			
Student ID	ID	English	Maths	Science	History	Studies	Geography	%	year
aqrq11021999	ut1	80	63	76	49	100	84	40	2011
swva20012000	ut1	60	58	85	94	61	73	79	2011
senz16032000	ut1	34	6	51	27	89	54	67	2011

Table 3: Student Marks table format for Academic year 2012											
Exam Social											
Student ID	ID	English	Maths	Science	History	Studies	Geography	%	year		
aqrq11021999	ut1	34	82	35	94	46	66	77	2012		
swva20012000	ut1	95	96	78	35	85	39	83	2012		
senz16032000	ut1	92	87	51	99	46	50	74	2012		

 Table 4: Student Marks table format for Academic year 2013

	Exam					Social			
Student ID	ID	English	Maths	Science	History	Studies	Geography	%	year
aqrq11021999	ut1	61	34	40	89	93	72	66	2013
swva20012000	ut1	82	37	56	77	83	82	89	2013
senz16032000	ut1	33	68	71	85	93	51	74	2013

Tuble 5. Student Marks tuble format for Fleatenne year 2011											
	Exam					Social					
Student ID	ID	English	Maths	Science	History	Studies	Geography	%	year		
aqrq11021999	ut1	35	69	91	70	82	64	34	2014		
swva20012000	ut1	55	33	43	64	70	43	50	2014		
senz16032000	ut1	77	37	62	39	53	91	59	2014		











Fig. 4 Subject wise SGP for a Particular Year and a Particular Exam

PREDICTION



Fig. 5 Student's Predicted Marks for academic 2017 and 2018

# 5. CONCLUSION

The conventional reporting system is not able to fulfil the expectations of the education industry stakeholders. It is also bundled with many shortcomings such as no growth reporting, no dashboard representation, and comparison of a student with students having different intelligence level. Thus, education analytics with SGP model helps in overcoming the shortcomings of the traditional system. Students, parents and teachers can analyze the actual growth of the student and this will help them in taking preventive

measures. Education Analytics can also be used to forecast the growth reports of the students which in turn help every stakeholder in this sector to take necessary actions. These trends help in optimizing the working of the student and helps in the decision making process. This also enlists various external factors that can hamper or boost the growth of a student. Using this process one can also find the strength and weakness of a student which can be very beneficial to improve one's career graph.

#### REFERENCES

[1]. Katherine E. Castellano, Berkeley Andrew D. Ho, 'A *Practitioner's Guide to Growth Models', Technical Issues in Large-Scale Assessment and Accountability Systems & Reporting, TILSA and ASR*, Feb 2013

[2]. Damian W. Betebenner, 'A Technical Overview of the Student Growth Percentile Methodology: Student Growth Percentiles and Percentile Growth', The National Center for the Improvement of Educational Assessment Dover, New Hampshire August 12, 2011

[3]. '*Georgia Student Growth Model Report*', Georgia Department of Education, 2012 – 2013

[4]. Kimberly J. O'Malley, Stephen Murphy, Katie Larsen McClarty, Daniel Murphy, Yuanyuan McBride, '*Overview of Student Growth Models', White Paper, Pearson*, Sept 2011

[5]. Wayne W. Eckerson, '*PREDICTIVE ANALYTICS* -Extending the Value of Your Data Warehousing Investment', Tdwi Best Practices Report, First Quarter 2007

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

[7]. A Microsoft Platform for Education Analytics', White Paper for Primary and Secondary Education, July 2011

[8]. Charles Elkan, 'Predictive Analytics and Data Mining', 2013

[9]. Beth Dietz-Uhler, Janet E. Hurn, 'Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective', Journal of Interactive Online Learning, 2013.

[10]. M.A. Chatti, A.L. Dyckhoff, U. Schroeder, and H. Thüs, 'A Reference Model for Learning Analytics', International Journal of Technology Enhanced Learning, 2013

[11]. O'Reily, '*How data analytics can improve education'*, *7th edition*, 2011

[12]. Roger Koenkar, 'Quantile Regression in R: A VIGNETTE', http://cran.r-project.org

[13]. Microsoft, 'Student Individualized Growth Model and Assessment (SIGMA) A Microsoft Education Analytics Platform Approach to Students at Risk', May 2010

[14]. varc.wceruw.org/what-we-do/ed-analytics.aspx

[15]. edanalytics.org