Application of Data Science in School Education Sector

Yash Aswani(C21022)
Geetha Joseph(C21003)
Syed Mirak Wajahat Kirmani(C21018)
Shahrukh Gouhar(C21013)

Praxis business School, Kolkata

Abstract

The objective of the study is to employ data science in an attempt to find solution to some of the hard pressing and long pending problems of school education sector. Four high priority problems have been chosen and various machine learning algorithms have been employed to find solutions to the issues. In spite of the fact that some of the techniques (use of images and data) used for data collection are highly ambitious, we could come up with a practical path to solution for each of the problem and a desired course of action.

The four problems we have analyzed are i) High dropout rate of students at various levels of school education ii) Inability of current education system to identify learning disability in a child iii) One size fits all pedagogy & ineffective traditional method of evaluation and iv) Lack of focus on evaluation of teacher's performance and teaching methods. The various machine learning algorithms used in the attempt to solve the problems are classification, clustering and regression. The results show that data science has much scope in finding innovative solutions to many of the issues in school and higher education sector.

Introduction

India has been considered as a developing nation for a few decades now. Though we are counted amongst the fastest growing economies in the world, our advancement on the 'Education' front hasn't been promising. High student drop-out rates, outdated pedagogical approaches, focus on creating trained minds rather than curious minds – these and more comprise the multitude of problems that need to be tackled and set right when it comes to education, especially school education. It is this thought that prompted us to analyse the education sector and the role technology, specifically 'data science' could play in finding solutions to some of the pressing issues in this sector

While customer focus is the mantra of every other sector, education sector especially school education has least focus on the needs and aspirations of individual students. Also, this is one area where data science hasn't made much inroads.

India was ranked 72 out of 73 countries in PISA (Programme for International Student Assessment) ranking of 2021. The constant low place for India year after year among world nations in PISA ranking is a cause of great concern. With the use of technology and data science in the appropriate areas, we believe that a progressive shift can be brought into the field of school education, thereby making a great contribution to the future advancement of the nation.

The four problems we have identified and the approach to solve them are described in the paper.

I   High dropout rate of students at various levels of school education:

The national school dropout rate at primary school level is 4.34 percent and it is even higher at the secondary school level, at 17.86% during 2019-2020. The state-wise dropout rate in schools is the highest in Assam, at both primary and secondary level. Over 62% of dropouts in education happens at school level. [1]

According to National Sample Survey, the major reasons for students dropping out are (according to their order of importance):

1. Engaged in economic activities.

2. Engaged in domestic activities.

3. Lack of interest in education.

4. Early Marriages.

5. Far off schools.

A child dropping out of schools is a huge waste of resourses for our country. Finding an approach to bring down the drop-out rate will add to the betterment of the society and the country.

With the help of data science an attempt has been made to predict the students who will drop out in an academic year and also to predict the no of students who drop out in a particular school/region which would in turn help the govt to deal with the reasons for drop out. For the same, data has to be collected on the various factors that contribute to drop out of a student/student

Proposed Data
Financial status of the family: (APL or BPL) It is found that poor financial conditions are highly effective in promoting child labour, thus resulting in children dropping out.

Educational Details of the family: If the parents are educated, it is very likely of them to motivate their children to pursue education.

Family health background: Presence of any terminally ill or any physically disabled parent is more likely to push a child into labour to meet the financial needs.[2]

Gender of the child: A girl child is more likely to drop out before the age of 15 years for either the school being too far, or she may indulge into house chores, or she may even be subjected to child marriages. A male child is more likely to drop out at the age of 17-18 due to financial problems and may start working, and if the child is a transgender, it may result in the child being outcasted by the family or the society.[3]

Student performance: At times due to lack of interest in studies many of the children drop out of schools. For more details refer [Table 9].

Challenges while collecting the data:

i) Reasons like child marriage, or child labour might not be reported ii) Families would be reluctant to disclose health data iii) Human error and bias since data collected by human iv) Families might not disclose the exact reason for dropping out v) Collecting data is time consuming.

Potential Data Security Issues:

i) Data regarding health conditions is sensitive in nature and has to be kept secure ii) Student details collected from the school might be collected under various terms and conditions and need to be kept secured.

Modelling

Classification algorithm is used to predict whether a student would drop out or not in the current academic year. Possible models are Decision trees, Naïve Bayes ad KNN.

Regression algorithm is used to predict the number of drop outs in a school/region in the upcoming year. Possible models are Linear regression, polynomial Regression.

Deployment and Evaluation

The model will be deployed in one of the regions chosen in Assam as it has the highest rate of drop outs in India. In order to evaluate the model, predict the drop out in an interval of three months from the beginning of an academic year and compare it with the actual drop out happened in that particular academic year.

Course of Action

Based on the outputs of the model, the following actions can be taken by the authorities.

i) The various reasons due to which a student is highly likely to drop out can be figured out. ii) If a girl is likely to drop out due to child marriage, govt. can take actions. iii) If a child is dropping out due to child labor, the govt. can provide scholarships to the child to pursue education. iv) If an area is found to have high no of drop outs, the reason for the same (lack of infrastructure, enough no of schools) can be traced back and actions can be taken.

II Inability of current education system to identify learning disability in a child at his early phase in primary school or at preparatory school.

Learning Disability is one of the most ignored, unattended and unpredictable problem in children. This is mostly because of (i) lack of awareness of such problems among elders, (ii) inability a child suffering from such disability to express his situation properly (iii) lack of appropriate mechanism to observe or measure directly such disabilities in children (iv) the issue of over identification in the current method faculties observation and various check lists [4]

Early identification and intervention is important for the long-term success of individuals with learning disabilities. Early identification includes the evaluation and counselling provided to families and their children under 3 years old who have, or are at risk of having a disability.

Some of the learning disabilities are: Dyscalculia, Dysgraphia, Dyslexia, Non-Verbal Learning Disabilities, Oral / Written Language Disorder and Specific Reading Comprehension Deficit, ADHD, Dyspraxia [4]

Proposed Data:

Student handwriting, Capturing Student behaviour, Student reading skills, Hereditary. [Table 6]

Challenges while collecting data:

(i) Huge Data Storage required, (ii) huge cost involved in collecting video clips and images (iii) getting permission from parents and school to take video and image footage of a student (iv) chances of uunwillingness of parents to share medical data and hereditary data of their child.

Potential Data Security Issues:

(i) Medical data is a very private data and involves huge security issue. Leak of such data holds potential harm for the family and students, (ii) the

output of model is carries private information. If a student is predicted to have certain learning disability, they may not want their fellow mates and society to know about it.

Algorithm:

Using multiclass classification algorithm, students can be classified into different disability classes. Some examples of classes are: (i) Dyslexia (ii) Dysgraphia (iii) Dyscalculia etc.

Other Challenges:

(i) High computational power required to model the data collected (ii) Analysing data containing images and videos is done by applying complex algorithm and model which are not transparent. Since the school authorities are answerable to the parents if their wards are detected with learning disability, lack of transparency poses an issue.

Deployment and Evaluation:

The model will be deployed in a few schools in Kerala assuming parents and teachers in the state are more aware of such learning disabilities considering the fact that education standards are quite high in the state and the state has the highest literacy rate.

In order to evaluate the model, the students with learning disability will be predicted in the schools were model was deployed. Students who were predicted to have learning disability will undergo a medical check-up and the model output will be compared with the medical report.

Course of Action:

If a child has been detected with a learning disorder, then the following would be of help for the student [5] (i) Extra help: A reading specialist, math tutor or other trained professional [5], (ii) Individualized education program (IEP): The IEP sets learning goals and determines strategies and services to support the child's learning in school [5], (iii) Accommodations: Use of computer applications that support writing, including fewer math problems in assignments, or providing audiobooks to supplement reading [5], (iv) Therapy: Occupational therapy might improve the motor skills of a child who has writing problems. A speech-language therapist can help address language skills [5]

III Problem Statement: One size fits all pedagogy and ineffective traditional method of evaluation

Some of the flaws of current education system in pedagogy and evaluation are: i)Evaluation of a student solely based on examination marks ii)Lack of methods to identify curious and innovative minds iii)Least weightage for extra-curricular activities and soft skills iv)Inability to use teaching methods based on student's aptitude and interest or student's strength and weakness and v) Least weightage for extra-curricular activities and soft skills All of this has put unnecessary burden on the shoulders of student, where students only focus on scoring good marks by any means and miss out on the primary aim of education. The aim of education system should be development of student's skills, aptitude, knowledge and perception all combined in a single frame.

The New Education Policy 2020[11] gives emphasis for a shift from mark-based evaluation method to 'Continuous and comprehensive evaluation'. Also, lot of focus is being given for critical thinking and more holistic, inquiry-based, discovery-based, discussion-based, and analysis-based learning. The policy also gives emphasis on building character and creating holistic and well-rounded individuals. Keeping this as the context, the role data science can play in improving student evaluation and there by designing student tailored pedagogy is being attempted in this paper.

Data Collection

Instead of evaluating a subject as a whole in a written examination, each subject is evaluated on multiple parameters [12] which is done regularly. The parameters [Table 1] would be evaluated in a scale of 1-5 by the teacher based on a scientifically prepared evaluation rubric [Table 2]. The data is collected from various sources like written tests, practical sessions, projects, group discussions, class participation etc [Table 1].

In order to evaluate the students on their behaviour, the same approach as mentioned above is followed. The parameters are evaluated by teachers on a scale of 1-10 based on an evaluation rubric. The various sources of data collection are group activities, sports, debates etc [Table 3].

Challenges while collecting the data:

Chances of teacher being partial to certain students [13] ii) Collecting data manually is time consuming and prone to error

Potential Data Security Issues:

The output of modelling student behaviour is highly confidential and has to be kept secured

Modelling

Clustering algorithm is used for both academic and behaviour evaluation

Academic Evaluation

The structured numeric data [Table 4] is used for modelling academic evaluation. The clustering algorithm creates multiple clusters of students with similar interests and competitiveness, for example, 'students good at logic & reasoning and average at arts'. Once such clusters are formed, based on the requirements of the students in the group, pedagogy can be designed by the teacher to best suit their needs. This would help the students to strengthen their weak areas and to go an extra mile in their strong areas.

Unlike in the traditional method of evaluation, students wouldn't be graded low and banded as poor performers for not doing well in one or two areas. Also use of data science in evaluation, helps to identify certain qualities in students like curiosity, reasoning etc. i and nurture them which would have been nearly impossible in the traditional method of evaluation

Behaviour Evaluation

Here again the structured numeric data [Table 5] is used for modelling behaviour Evaluation. Using clustering algorithm various clusters of students are formed who have similar behaviours. For eg: 'students good at individual work but poor at gender sensitivity' could be a cluster. This helps the teachers to formulate games or plays appropriate for each student/student group to instil a particular quality in them.

In the traditional method of teaching, moral science classes were conducted to bring about behavioural changes in students. Using data science in behaviour evaluation helps to cater to the needs of individual students unlike the traditional generic approach.

Deployment and Evaluation of the model

The model will be first deployed in a slected school in Assam (dropout rate = 10.1%, 2017-2018). The model will be evaluated by comparing the performance of the students of a school where the model was deployed with that of the students of a school were the model was not deployed.

Challenges:

i) Small Classes with a maximum strength of 25-30 students is required for the teacher to evaluate various aspects of student ii) High teacher to student ratio is required iii) Requirement of better infrastructure to handle multiple classes of small strength iv) Highly trained and able teachers - Teacher training institutes in India are not at par with international standards

IV Lack of focus on evaluation of teacher's performance and teaching methods.

We find that today, as never before, teachers are dangerously overloaded. Their traditional functions of instruction, socialization, evaluation and classroom management are not regarded sufficient to make them effective. Present time possess challenges that were never faced by traditional school teachers. They are facing a flux in the educational scenario which contains innumerable and complex situations. [8]
Problem details:
  - Evaluation of a teachers solely based on problem solving skills
  - In Evaluation of teacher Experience is given more weightage without considering if there any self-improvement or not.
  - No evaluation on teachers' character traits.

Teacher evaluations are often designed to serve two purposes: to measure teacher competence and to foster professional development and growth. A teacher evaluation system will give teachers useful feedback on classroom needs, and will provide the opportunity to learn new teaching techniques, so to make changes in their classrooms accordingly. Thus the purpose of teacher evaluation in bringing about change for betterment is now widely recognized and accepted. The main aim of the evaluation is continuous improvement of the educational scenario. [9]

Strategies for evaluating teaching quality:

With regards to in-service teacher evaluation, below given are the numbers of teacher evaluation strategies that are described in the literature, utilising multiple sources of evidence and which have been adopted to varying degrees in different contexts around the world [6]
  - Classroom observations
  - Teacher portfolios
  - Student evaluation of the teacher
  - Student learning outcomes
  - Teacher self-evaluation
  - Teacher tests (e.g., of subject knowledge)
  - Professional conversations
  - Peer evaluation
  - Parent feedback

Proposed data:

First the data will be collected in the form of feedback of students, principal of the school and colleagues[10]. For more details refer [Table 8]

Secondly the data is collected from teacher itself. During this process a situation is given to teacher in which he/she has to describe how he/she would respond to the situation, in text format. For more details refer [Table 7]

Challenges while collecting the data:

i) To figure out which students are fit to give the feedback for a teacher ii) Chances of human error and bias iii) Collecting data is time consuming

Potential Data Security Issues:

i) Feedback from students has to be kept anonymous. ii) The data collected is highly sensitive and have to be kept secured iii) Model output should be shared only to the respective teacher and not to the students.

Modelling:

Using multiclass classification algorithm, teacher's performance is classified into different classes. Some examples of classes are: Excellent, Good, Moderate, Poor, Very Poor.

Enablers for the implementation of an effective teacher evaluation system: [Table7]

Deployment and Evaluation:

The model will be first deployed in a selected school in Assam (dropout rate = 10.1%, 2017-2018).

In order to evaluate run the model for three months and compare the outputs of the model with the current performance of the teachers. And check for an improvement in performance and character traits of the teachers in he school

Course of Action:

Training should be given to the teachers based on in which class they fall.

- If classified as 'moderate', then basic confidence boosting session can be given.

- If classified as 'poor' and 'very poor', then conceptual training should be given.

Scope of further improvement and Future Plan

❖ Improve the efficiency and speed of data collection by using electronic devices;
❖ We can extend this project to higher education institutes,
❖ Application of data science in further areas like choosing the appropriate location for govt. schools

Conclusion

In this paper we have employed data science to try and find solutions for the four problems we have identified from education sector. We could see that, if multiple factors leading to school dropout can be identified and data can be collected using appropriate mechanism, students who have high probability of being dropped out in an academic year can be predicted. Similarly learning disability in a student can be correctly identified if the appropriate machine learning algorithm is used. Even though the evaluation of teachers and students using data science has certain limitations, further study in the area and use of more advanced and scientific methods of data collection could bring in more insight into solving those problems. Through this paper, we have attempted to showcase the potential of data science in solving some of the issues in education sector and this need to be treated as a precursor to more in-depth studies that would happen in the near future in the area of education.

References

[1] Syed Rooquiyya Tabassum - "A Study of School Dropouts in India: Sensitivity at Display,"

[2] Yogendra Nath Tiwari, Dr. Ali Imam – "A Study of Potential Drop-Out and its Causal Factors in Elementary Schools of Central U.P."

[3] S.V. Halawar - "A Study on School Dropouts and Student Strength of Dharwad District of Karnataka, India"

[4] Training Module On Learning Disability (Paschim Bang Sarva Shiksha Missiom Bikash Bhaban, Kolkota, West Bengal in collaboration with Distance Education Programme- Sarva Shiksha Abhiyan (DEP-SSA) (An IGNOU-MHRD Govt of India Project) IGNOU, Maidan Garhi, New Delhi 110 068)

[5] https://www.mayoclinic.org/healthy-lifestyle/childrens-health/in-depth/learning-disorders/art-20046105

[6] Michelle Bambawale, Jemima Hughes and Amy Lightfoot – "Exploring teacher evaluation processes and practices in India:A case study"

[7] http://www.educationindiajournal.org/home_art_avi.php?path=&id=359#

[8] http://egyankosh.ac.in/bitstream/123456789/46996/1/Unit-8.pdf

[9].http://www.vkmaheshwari.com/WP/?p=200#:~:text=A%20teacher%20evaluation%20system%20will,now%20widely%20recognized%20and%20accepted.

[10] RENUKA AGRAWAL, JYOTI SINGH, A.S. ZADGOANKAR – "FORMATIVE ASSESSMENT FOR PERFORMANCE EVALUATION OF FACULTY USING DATA MINING"

[11] New Education Policy 2020

[12] Applying Creativity in CS High School Education – Criteria, Teaching Example and Evaluation, Ralf Romeike, University of Potsdam, Department of Computer Science

[13] Evaluation and Assessment Frameworks for Improving School Outcomes, Common Policy Challenges, OECD

Annexure

Table 1:

	Feature / Variable	Data Source	Data
Science	Understanding of concept: Reasoning ability Spirit of Inquiry Application	Written Tests, Continuous evaluation by teacher using rubric Practical sessions, projects, group discussions, class participation (nature of question raised by students)	Structured numeric data
Maths	Understanding of concept: Application Thinking Application		Structured numeric data
Language	Reading Writing Speaking Story Telling	Written Tests, Continuous evaluation by teacher using evaluation rubric	Structured numeric data
Arts	Application of art concepts Use of art materials Creativity	Continuous evaluation by teacher using evaluation rubric	Structured numeric data

Table 2: Sample Evaluation Rubric

Category	5	4	3	2	1
Understanding	Completely understood the problem and solved it	Understood the problem and mostly solved it	Understood part of the problem but could not solve it	Did not understand the problem but put some effort	Did not understand the problem and no effort
Reasoning	Had a good plan for solving the problem and could explain it well	Had a plan for solving the problem and could mostly explain it well	Had a plan for part of the problem and but needed help	Had some idea about the problem	Did not know how to figure out the problem
Questioning and discussing	Discusses and ask questions and helps direct the group in solving problems	Discusses and ask questions	Discusses in length but fails to come up with relevant questions	Discussions with least relevance to the topic	Least involved in discussions
Problem Solving	Actively seeks and suggests right solution to the problems	Actively try out solution and suggests various solution to the problems	Does not offer solution, but willing to try solution	Does not try to offer solution	Does not understand the problem

Table 3:

Feature	Value
Concept	1-5
Reasoning	1-5
Spirit oof Enquiry	1-5
Application	1-5
Reading	1-5
Writing	1-5
Speaking	1-5
Story Telling	1-5
Creativity	1-5
Use of art materials	1-5

Table 4:

Feature	Data	Data Source
Gender sensitivity	Structured numeric data	
Responsibility	Structured numeric data	Group Activities
Team work	Structured numeric data	Debates
Self-discipline	Structured numeric data	General conduct in class
Empathy	Structured numeric data	Sports
Inclusiveness	Structured numeric data	

Table 5:

Feature	Value
Gender sensitivity	1-5
Responsibility	1-5
Team work	1-5
Self-discipline	1-5
Empathy	1-5
Speaking	1-5
Inclusiveness	1-5

Table 6:

DATA	Data Type	Source	Used to Identify
Student's Notebook Images	Unstructured	From Family	Dyslexia, Dysgraphia, Dyscalculia
Student's video and images footage	Unstructured	Installing Camera in School	Dyslexia, dyspraxia, ADHD, Non-verbal Learning Disability
Hereditary	Structured	From School	Dyslexia, Dysgraphia, Dyscalculia, ADHD, Non-verbal Learning Disability
Medical Records of student Already identified to have any disability	Structured	Hospitals	Dyslexia, Dysgraphia, Dyscalculia, ADHD, Non-verbal Learning Disability

Table 7

	Areas for development to facilitate an effective system of teacher evaluation in India		
	Thinking	**Knowledge and skills**	**Action**
Teachers	Understanding of roles and responsibilities	Valuing the multi-faceted nature of the education ecosystem	Acting on feedback
	Inclusive attitudes towards the child – holistic and democratic	Exposure to different assessment or evaluation models	Active engagement with peers (communities of practice)
	Appreciation of why and how evaluation processes will help them (and their learners)	Skills for accepting and giving constructive feedback	Active pursuit of professional development opportunities
		Confidence to judge the quality of their own work	Regular and structured reflection
School leaders	Understanding current reality of the school including current practices of teacher evaluation	Skills for effective monitoring and evaluating of classroom practice	Advocating self-reflection as an integral part of school dayto-day activities
	Understanding the school context in the larger education ecosystem	Provisions that enable all students and staff to be able to realise school goals including student learning as well as CPD	Ensuring collective decisionmaking for development of the teacher evaluation process
	Understanding of teachers' insecurities around evaluation	Links between evaluation and professional development opportunities at the school level	Agree expectations and roles of the teacher, i.e. within the school and with other stakeholders

Table 8

Business Data For Teacher Performance Evaluation			
Teacher Name	Text	Explanation with practical Examples	Yes, No
Teacher ID	Text	Friendly Attitude	Yes, No
Speed of delivery	Very low, low, Moderate, Fast, Very Fast	Encouraging Behaviour	Yes, No
Presentation	Very low, low, Moderate, Fast, Very Fast	IQ Score	Numeric
Subject Knowledge	Very low, low, Moderate, Fast, Very Fast	Attendance %	Numeric
Communication	Very Poor, Poor, Average, Good, Very Good	Average Score of class in his subject	Numeric
Doubt Clearing /Sessions	Yes, No	Principal Evaluation	Very Poor, Poor, Average, Good, Very Good
Assignment regularity	Yes, No	Rating from experts	Very Poor, Poor, Average, Good, Very Good
Over All completion of syllabus	Yes, No	Overall Evaluation of Teacher	Very Poor, Poor, Average, Good, Very Good
Revision Class	Yes, No		

Table 9

Business Data For Drop Out		
Attributes	**Value**	**Source**
Student Name	Text	School
Student ID	Text	School
Location	Text	School
Gender of the child	M,F,T	School
Student performance	Poor, Satisfactory, Good, Very Good, Excellent	School
Distance of school	Within 5km, 5km-10km, above 10km	School
Educational Details of the family:	No Education, Below 10th, 10th, 12th, Graduate, Post Graduate	Family
Financial status of the family	APL, BPL	School/Family
Family health background	Yes, No	Family
Family history of marriage before 18 or 21	Yes, No	Family
Involved in Economic Activities	Yes, No	Family
Involved in Domestic Activities	Yes, No	Family
Parent's Earning type	Fixed, Variable	Family
Religion Constraint	Yes, No	Family
Fully Or Partially Orphend	F, P, N	School
Dropout or Not	Yes, No	School/Board