# Dropout PredictorA Statistical Model to Predict Dropouts in Schools 

Siddhant Sanghi<br>The Cathedral and John Connon School, Mumbai, India


#### Abstract

One of the major issues facing Municipal schools in Mumbai is that of children dropping out of school before completing their schooling, an alarming $15 \%$ and rising. These children are at a significant disadvantage compared to those who complete their education. The purpose of this paper is to predict potential dropouts ahead of time, such that remedial measures can be taken to ensure that they continue their education.


In order to identify likely dropouts, data was gathered from multiple municipal schools across Mumbai. Binary Logistic regression was run on this data to develop a statistical model called "Dropout Predictor". The initial set of variables for the "Dropout Predictor" included age, gender, attendance and grades, household income, education level of parents, number of children, birth order, working mom or homemaker and child's behaviour. The results of running the model on the sample data set revealed that the most indicators of students dropping out were household income, followed by academic grades. The results of the Dropout Predictor indicated that $11 \%$ of the children were likely to dropout if preventive measures were not taken immediately. Running it on past data further validated the model.

The Predictor was successful with the children that were identified as likely to dropout. Results are encouraging since they show that preemptive action via counseling and other techniques can prevent a significant number of dropouts.

A limitation of the model was that while it was reasonably accurate in predicting the students that did dropout, it could not predict all the students that had dropped out. Further work is therefore required to analyze complex causes, such as rampant urban migration or inadequate educational infrastructure, to better predict dropouts, perhaps using machine learning.

## 1. BACKGROUND

## Dropout Situation in India

A major issue facing schools in India was that of children dropping out before completing high school: the dropout rate at the secondary level last year was an alarming 15\% last year.

## Why a Project on Dropouts? - Impetus

Education is the bedrock of a thinking mind, and educated youths are the wealth of a nation. Since education is the only route to imbibing knowledge and skills, breaking the shackles of poverty and leading a meaningful life, it is imperative to stem the flow of dropouts.

A Student Dropout Predictor, a statistical model that predicts students that were likely to dropout, was conceptualized. The advantages of a statistical model are that it could be applied to all schools (scalable), responsible for achieving targets (accountable); and results show if remedial steps were effective in stemming dropouts (measurable).
The statistical model based on data from a few Municipal Schools In Mumbai, is presented in this research paper.

## 2. PROJECT EXECUTION

## Phase 1: Preparing Data for Regression

## Data Gathering

With the support of the principals of the Municipal School adopted for saving dropouts, extensive data for students from grades 5 to 10, across English, Hindi and Marathi mediums was gathered.

Since many of the variables were not in the school database, class teachers were personally spoken to, for getting indicative data on household incomes and child's behaviour. Children were also asked for the number of siblings, their birth order and whether the mother was a homemaker or working mom. In case any of these soft variables (not in the school database) become important for predicting dropouts, perhaps a system to capture those in the future can be worked.

## Selection of Variables, Converting Variables to Numbers

Variables selected for analysis were the ones thought to be most relevant for predicting dropouts. Variables were converted to either numerical ( N ), categorical (C) or interval (I). Interval variables were classified as: Good $\mathrm{G}=3$, $\mathrm{A}-$ Average $=2$ and $\mathrm{P}-\mathrm{Poor}=1$.

| 1. | Age | (Numerical) |
| :--- | :--- | :--- |
| 2. | Gender | (Categorical-2) |
| 3. | Attendance | (Interval-3) |
| 4. | Academic Grades | (Interval-3) |
| 5. | Child's Behaviour | (Interval-3) |
| 6. | Number of Kids | (Numeric) |
| 7. | Child Birth Order (younger/older) | (Categorical-2) |

8. Family Income
9. Parents' Education Levels
10. Working Mom or Home Maker
11. Medium of Instruction (English, Hindi, Marathi-English)

## (Interval-3)

(Interval-3)
(Categorical-2)
(Categorical-3)

## Data on Kids who had dropped out:

The process that the school followed was that if the child was absent for 3 months, it was considered that he has dropped out. Six months into the school year, from approximately 1000 children, $10 \%$ had already dropped out (those children are marked in Bold in the attached table).

Sample Data Set (Grade 6) Key: $\mathrm{G}=\mathrm{Good}$; $\mathrm{A}=$ Average; P=Poor;

| $\begin{aligned} & \dot{8} \\ & z \\ & \dot{Z} \\ & \ddot{Z} \\ & \hline \end{aligned}$ | $\stackrel{8}{\ll}$ | $$ |  |  |  | $\begin{aligned} & y \\ & 0 \\ & 0 \\ & 0 \\ & \# \\ & \# \end{aligned}$ | $\begin{aligned} & \dot{0} \\ & 0.0 \\ & 0 \\ & \tilde{\#} \\ & \end{aligned}$ | 訔 若 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1. | 11 | Girl | G | G | G | 3 | 3 | A | A | Y |
| 2. | 11 | Girl | A | G | G | 4 | 2 | P | P | N |
| 3. | 11 | Girl | G | G | G | 2 | 1 | A | A | Y |
| 4. | 11 | Girl | P | P | G | 4 | 2 | P | P | Y |
| 5. | 11 | Girl | G | G | A | 6 | 3 | A | P | N |
| 6. | 11 | Girl | G | G | G | 4 | 1 | A | A | Y |
| 7. | 11 | Girl | A | A | G | 4 | 2 | A | A | Y |
| 8. | 11 | Girl | A | A | G | 3 | 3 | P | P | N |
| 9. | 11 | Girl | P | P | G | 3 | 3 | P | P | Y |
| 10. | 11 | Girl | G | P | A | 4 | 1 | P | P | N |
| 11, | 11 | Girl | A | A | G | 3 | 1 | P | P | N |
| 12 | 11 | Girl | G | A | G | 4 | 3 | P | P | N |
| 13, | 11 | Girl | G | G | G | 3 | 3 | A | A | N |
| 14. | 11 | Girl | G | G | G | 3 | 1 | A | A | Y |
| 15. | 11 | Girl | A | A | A | 3 | 1 | A | A | Y |
| 16. | 11 | Girl | G | G | G | 2 | 2 | A | A | Y |
| 17. | 11 | Girl | G | G | G | 8 | 7 | P | P | N |
| 18. | 11 | Girl | A | G | G | 3 | 3 | P | P | N |
| 19. | 11 | Girl | A | A | G | 5 | 4 | P | P | N |
| 20. | 11 | Girl | G | G | G | 3 | 1 | A | A | Y |
| 21. | 11 | Girl | G | G | G | 2 | 1 | A | A | N |
| 22. | 11 | Girl | A | G | G | 3 | 2 | A | A | Y |
| 23. | 11 | Girl | G | G | G | 4 | 2 | P | P | N |
| 24. | 11 | Girl | G | G | G | 3 | 3 | P | P | Y |
| 25. | 11 | Girl | G | G | G | 6 | 4 | A | P | N |
| 26. | 11 | Girl | G | G | G | 2 | 2 | A | A | N |
| 27. | 11 | Girl | G | A | A | 6 | 5 | A | P | N |
| 28. | 11 | Boy | G | G | G | 3 | 1 | A | A | N |
| 29. | 11 | Boy | G | A | A | 2 | 1 | A | P | N |
| 30. | 11 | Boy | A | A | G | 3 | 2 | A | P | N |
| 31. | 11 | Boy | P | P | G | 4 | 3 | A | A | N |
| 32. | 11 | Boy | G | G | G | 3 | 2 | A | A | Y |
| 33. | 11 | Boy | A | A | A | 7 | 6 | P | P | N |
| 34. | 11 | Boy | G | G | G | 3 | 1 | P | P | Y |
| 35. | 11 | Boy | G | G | G | 3 | 1 | P | P | Y |
| 36. | 11 | Boy | A | P | A | 2 | 2 | A | P | Y |
| 37. | 11 | Boy | G | G | G | 2 | 1 | A | A | Y |


| 38. | 11 | Boy | A | A | G | 7 | 7 | P | P | Y |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 39. | 11 | Boy | A | A | A | 3 | 1 | P | P | Y |
| 40. | 11 | Boy | G | G | G | 4 | 4 | A | P | N |
| 41. | 11 | Boy | P | P | A | 3 | 3 | A | P | N |
| 42. | 11 | Boy | A | A | A | 4 | 4 | P | P | N |
| 43. | 11 | Boy | P | G | G | 2 | 1 | A | P | N |
| 44. | 11 | Boy | A | A | A | 2 | 1 | P | P | N |
| 45. | 11 | Boy | G | A | G | 3 | 3 | A | A | N |

## DROPOUTS

| 46. | 11 | Girl | P | P | A | 3 | 3 | P | P | N |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 47. | 11 | Boy | P | A | A | 3 | 2 | A | P | N |
| 48. | 11 | Boy | P | P | A | 5 | 3 | P | P | N |
| 49 | 11 | Boy | P | G | G | 3 | 2 | A | P | N |
| 50. | 11 | Girl | P | P | G | 5 | 5 | P | P | N |
| 51. | 11 | Boy | P | P | A | 7 | 3 | A | P | N |

Phase 2: Conducting Statistical Analysis

## Method

Binary logistic regression was used to assess the predictors of dropout. Analysis was performed in a stepwise fashion. Exploratory data analysis was first performed to assess whether some pre-processing is needed prior to the analysis. Univariate binary logistic regression was then performed to identify variables that may prove useful in predicting the dropout status. A model was constructed for each variable. The significance of coefficients was assessed using the likelihood ratio test Chi-square (likelihood ratio test). A receiver-operating curve was constructed for each variable and the area under the curve (AUC) was used to assess the predictive power of the model constructed for each variable. AUC close to 1 indicates better performance.
Secondly, useful variables from the $1^{\text {st }}$ step were used in a multivariate logistic regression model to assess the independent predictors of dropout. Two-tailed hypothesis testing was performed. P values less than 0.05 were considered statistically significant. The data included observations for 188 students; 16 (8.5\%) dropped and 172 ( $91.5 \%$ ). Not taking any variables into consideration, these results show that students are less likely to drop.

## Pre-processing

Some variables were recoded prior to the analysis. The number of kids was recoded as $<3$ or $3+$ while the birth order was recoded as first birth or after-first birth due to the small number of observations.

| Table 1: Cross tabulation of birth order and dropout |  |  |  |
| :--- | :--- | :--- | :--- |
| Birth | No dropout | Dropout | Total |
| 1 | $78(97.5 \%)$ | $2(2.5 \%)$ | 80 |
| 2 | $62(86.1 \%)$ | $10(13.9 \%)$ | 72 |
| 3 | $21(87.5 \%)$ | $3(12.5 \%)$ | 24 |
| 4 | $6(100 \%)$ | $0(0 \%)$ | 6 |
| 5 | $1(50 \%)$ | $1(50 \%)$ | 2 |
| 6 | $1(100 \%)$ | $0(0 \%)$ | 1 |
| 7 | $3(100 \%)$ | $0(0 \%)$ | 3 |
| Total | $172(91.5 \%)$ | $16(8.5 \%)$ | 188 |

Results show that the probability of dropping out for the firstborn child is $2.5 \%$ compared to $13.9 \%$ and $12.5 \%$ among $2^{\text {nd }}$ and $3^{\text {rd }}$ births. The number of observations was small in the remaining categories. Thus, the variable was dichotomized into $1^{\text {st }}$ birth and $2^{\text {nd }}$ birth or more. Similarly, the number of kids was dichotomized into $<3$ and $3+$ kids.

## Exploratory data analysis

| Table 2: Cross tabulation of various demographic factors and dropout |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | No dropout |  | Drop out |  |
|  |  | Count | \% | Count | \% |
| Gender | Boy | 74 | 94.9\% | 4 | 5.1\% |
|  | Girl | 98 | 89.1\% | 12 | 10.9\% |
| Medium | MarathiHindi | 27 | 90\% | 3 | 10.\% |
|  | Hindi | 102 | 91.9\% | 9 | 8.1\% |
|  | English | 43 | 91.5\% | 4 | 8.5\% |
| Attendance | Poor | 11 | 40.7\% | 16 | 59.3\% |
|  | Average | 29 | 100.0\% | 0 | 0.0\% |
|  | Good | 132 | 100.0\% | 0 | 0.0\% |
| Child's Grades | Poor | 14 | 56.0\% | 11 | 44.0\% |
|  | Average | 72 | 98.6\% | 1 | 1.4\% |
|  | Good | 86 | 98.9\% | 1 | 1.1\% |
| Behaviour of Child | Poor | 18 | 72.0\% | 7 | 28.0\% |
|  | Average | 99 | 93.4\% | 7 | 6.6\% |
|  | Good | 55 | 96.5\% | 2 | 3.5\% |
| Income ofParents | Poor | 29 | 78.4\% | 8 | 21.6\% |
|  | Average | 125 | 94.0\% | 8 | 6.0\% |
|  | Good | 18 | 100.0\% | 0 | 0.0\% |
| Education of Parents | Poor | 40 | 80.0\% | 10 | 20.0\% |
|  | Average | 65 | 94.2\% | 4 | 5.8\% |
|  | Good | 60 | 96.8\% | 2 | 3.2\% |
| Mother status | Working | 146 | 93.0\% | 11 | 7.0\% |
|  | Housewife | 26 | 92.9\% | 2 | 7.1\% |
| Number of kids | $<3$ | 65 | 95.6\% | 3 | 4.4\% |
|  | 3+ | 107 | 89.2\% | 13 | 10.8\% |
| Birth order | 1st | 78 | 97.5\% | 2 | 2.5\% |
|  | 2nd or more | 94 | 87.0\% | 14 | 13.0\% |

Results show that the dropout was slightly higher in girls compared to boys ( $10.9 \%$ vs. $5.1 \%$ ). Attendance showed a strong association with dropout. Students with good or average attendance did not dropout ( $0 \%$ ) compared to students with poor attendance (44\%). Results show that $28 \%$ of children with poor grades dropped subjects compared to $1.4 \%$ and $1.1 \%$ of children with average and good grades, respectively. Children with average and good grades can be combined into 1 category since $\%$ was similar in both groups. Dropout was higher across parents with poor income (21.6\%) compared to parents with average or good income ( $6 \%$ and $0 \%$, respectively).

Good and average income can be combined as one category. Parents' education also showed a significant association with dropout. The dropout rate was higher in parents with poor education ( $20 \%$ ) compared to parents with average or good
education. Thus good and average education was combined since $\%$ was similar in both groups. A similar trend was observed with behaviour. Thus the good and average categories were merged for the following variables: behaviour, income, education, and grade. Attendance will not be included in the model to avoid complete separation since no students with average or good attendance dropped out. The medium did not seem to affect the dropout as the $\%$ were equal across the three mediums.

## Univariate analysis

Table 3: Univariate analysis results (Likelihood ratio test)

|  | LR chi- <br> square | $\mathbf{d f}$ | $\mathbf{P}$ | AUC |
| :--- | :--- | :--- | :--- | :--- |
| Attendance | 50.374 | 2 | $<0.001^{*}$ | 0.968 |
| Gender | 2.578 | 1 | 0.108 | 0.615 |
| Child's Grades | 27.124 | 1 | $<0.001^{*}$ | 0.859 |
| Behaviour of Child | 5.433 | 1 | $0.02^{*}$ | 0.618 |
| Income of Parents | 5.053 | 1 | $0.025^{*}$ | 0.666 |
| Education of Parents | 24.526 | 1 | $<0.001^{*}$ | 0.863 |
| Mother's status | 0.159 | 1 | 0.69 | 0.524 |
| Younger/Older | 0.002 | 1 | 0.963 | 0.507 |
| Medium | 3.784 | 2 | 0.151 | 0.622 |
| Birth order | 5.678 | 1 | $0.017^{*}$ | 0.677 |
| Number of kids | 3.807 | 1 | $0.05^{*}$ | 0.639 |

Results show that gender ( $\chi 2=2.578, \mathrm{P}=0.108$ ), mother's work status $(\chi 2=0.0 .159, \mathrm{P}=0.69)$ and age of the child $(\chi 2=$ $0.002, \mathrm{P}=0.963$ ) were not significantly associated with dropout. Attendance was the strongest predictor of dropout ( $\chi 2$ $=50.374, \mathrm{P}<0.001$ ). Child's grade $\left(\chi^{2}=27.124, \mathrm{P}<0.001\right)$, parents' education ( $\chi 2=24.526, \mathrm{P}<0.001$ ), income ( $\chi 2=$ 5.053, $\mathrm{P}=0.025$ ) were strong predictors of dropout. Birth order $(\chi 2=5.678, \mathrm{P}=0.017)$, and number of kids $(\chi 2=3.807$, $\mathrm{P}=0.05$ ), and child's behaviour ( $\chi 2=26.497, \mathrm{P}<0.001$ ) were also predictors of dropout but the $\chi^{2}$ values were lower compared to those seen with the first four variables.
These results show that seven variables can be used to predict the drop especially the attendance which showed the highest likelihood ratio $\chi 2$. Examining the AUC for these variables show that the AUC was highest for education (0.863), attendance ( 0.968 ), grade ( 0.859 ). This indicates that these variables can be used as predictors of dropout. The AUC was lower for birth order and number of kids.

## Multivariate analysis

Backward stepwise logistic regression was used. Variables were entered into the model if the significance for the score statistic was less than 0.2 and were kept in the model if the significance for Wald-statistic is lower than 0.1 .

|  | B | S.E. | Wald <br> $\chi^{2}$ | df | P | OR |
| :--- | :---: | :--- | :---: | :---: | :---: | :---: |
| Grade <br> (Good/average) | -3.826 | .859 | 19.824 | 1 | .000 | .022 |
| Income <br> (Good/average) | -1.813 | .824 | 4.834 | 1 | .028 | .163 |
| Gender (Girl) | 1.920 | 1.021 | 3.538 | 1 | .060 | 6.820 |
| Constant | -2.646 | 1.730 | 2.339 | 1 | .126 | .071 |

## Conclusion

Results show that grade was the strongest predictor of dropping out. The odds of dropping out in students with good/average grades are 0.022 the odds in students with poor grades holding the remaining variables constant $(\mathrm{OR}=0.022$, $\chi^{2}=19.824, \mathrm{P}<0.001$ ). Income was also associated with the odds of dropping out.

The odds of dropping out in families with good/average income were $16.3 \%$ the odds in families with poor income holding the remaining variables constant ( $\mathrm{OR}=0.163$, $\chi 2=$ $04.834, \mathrm{P}=0.028$ ). This means that students are more likely to drop if they come from a family with poor income compared to families with low income.

Finally, girls were more likely to dropout compared to males ( $\mathrm{OR}=6.82, \chi 2=3.538, \mathrm{P}=0.06$ ). This means that the odds of dropping out in girls are 6.8 times the odds in boys. All these results assume that the remaining variables are held constant in the model. Attendance was not included in the model since it showed a complete separation. Thus, including it in the model will affect the calculation of estimates.

## Phase 3: Analyzing Results of the Model

## Low Household income was Strongest Indicator of Dropouts:

The model showed that kids coming from very low-income household were most likely to dropout. Ms. Nischint Hora, Project coordinator of one of the NGO's dedicated to the girl child, revealed various causes for kids dropping out due to financial reasons.

According to Ms. Hora, most of the children enrolled in municipal schools were children of migrant workers: their families had moved from the villages to the city to fulfil their growing aspirations. They had limited finances and lived from hand-to-mouth. If their slums were getting demolished to make way for high-rises, or the earning member of the family had lost his job, or a family member had contracted serious health condition or their lease rental had run out, the family would have no choice but to shift faraway to a cheaper residence or go back to their village.
Furthermore, since compulsory education was only mandatory until age 14 (grade 8), after which children were legally allowed to work, parents often pulled the kids out of school to help them with the trade or with house work. Unless government policies changed and economic environment
improved in uplifting the poor, there is little that can prevent dropouts due to financial reasons.

## Poor Academic Performance was the $2^{\text {nd }}$ Strongest Indicator:

Children lacked motivation to study and therefore fared poorly in exams. Many also often complained of being "bored" in school. If teachers are unable to inspire children, then dropouts cannot be far behind. With a severe shortage of teachers, the same teachers were made to teach multiple subjects, often subjects in which they had no formal teacher training. One municipal school had four teachers managing six classrooms. Teachers also had their hands full: over-crowded classrooms often poorly behaved children always remained a challenge.

Furthermore, children often came from very difficult backgrounds. It was easy for these children on the fringes to get drawn into the toxic circle and trade of alcohol and drugs. Given the poor and illiterate backgrounds that most of the children came from, they lacked the motivation to study; they had no role models to emulate. It was also common for these children to leave home for school, but get distracted with other activities on the way and never reach school, most times unbeknown to parents.

## Phase 4: Limitations of the Model

## False Positives could lead to Biases

It is possible that the model wrongly classifies some children as Likely to Dropout. Labelling kids as likely to dropout (whether they will or not) could create systemic biases and aggravate the situation. If wrongly classified, it could also lead to a self-fulfilling prophecy.

## Labeling kids could create Stigmas and Biases

Kids identified by the Predictor model as 'Likely to Dropout' could face a stigma, a bias against them from their classmates and perhaps even some teachers. These biases could exacerbate the situation even more.

## Complex Causes making it difficult to Procure Data

A limitation of the model was that while it was reasonably accurate in predicting the students that did dropout, it could not predict all the students that had dropped out. The reason is that there are many causes to complex, either to predict, or to quantify, or to solve.

## Next Steps - To use more Advanced Techniques

Further work is therefore required to analyse complex causes, such as rampant urban migration or inadequate educational infrastructure, to better predict dropouts, perhaps using machine learning.

## Phase 5: Final Step-Implementing the Solution

Once potential dropouts had been identified, the next step would be to undertake preventive action.
As discussed earlier, one of the main reasons that children dropped out was because they were not motivated to study. Teachers were of the view that counselling session that highlighted the value of education and showed them all the opportunities that could open up in the future would help them stay back in school

## Counseling Sessions

Counselling sessions with the Municipal schoolchildren commenced with the aim to motivate them: one session before their exams was devoted to study tips, while another was on future career options and making the "Dream Big", while yet another on tips to happiness. Children were extremely happy with these sessions. For the first time they felt that somebody was looking for them, for their interests.

## Save-A-Dropout.org

The counselling sessions were a success, but in order to roll it out to other schools, additional counsellors would be required. Since there was already a shortage of counsellors in municipal school, support was found in neighbouring communities. A web portal called "Save-a-Dropout (.com) ...Build the Nation" was created, where individuals could volunteer their time for providing moral and inspirational support to these children.

They could also ensure that the families of these children likely to dropout are using all the services that the government offers, in the form family counselling in cases of strife, financial help from relevant NGOs, career counselling, etc. The marketing efforts have just begun. My plan is to have each of the volunteer counsellors connect to 4-5 children from the neighbouring schools, and be their mentors.

## 3. ACKNOWLEDGEMENTS

The entire exercise is a collaborative effort and I would like to thank the various individuals who have helped me with this project, without whom the project would surely not have been possible.

Prof. Hema Santhanam, Adjunct Professor (Research) for Statistics at KJ Somaiya, for being a mentor to me and guiding me with the mathematics in this paper: in running regression models on the data and analysing the results
Mr. Kaustubh Dhavse, Officer on Special Duty to the Chief Minister of Maharashtra, for his encouragement to undertake this project.

Shri Milind Sawant, Deputy Municipal Commissioner (Education), for whole-heartedly supporting this proposal and giving a chance to implement this ideas.

Colaba Municipal School Principals for helping me with data gathering, inviting me to conduct counselling for kids, and offering support for Save-a-Dropout website.

Ms. Nischint Hora, Project Coordinator -Vachha Foundation (NGO dedicated to the girl child),for her invaluable inputs on the lesser-known causes of dropouts

## 4. FURTHER SCOPE

## Replicating Model to other Cities in India

This in-depth data mining could bear fruit if similar models for other cities are created. If successful in Mumbai, the predictor along with the best practices package could be shared with Municipal authorities across the country.

This Dropout Predictor would be the first of its kind. Besides being scalable and adaptable to local conditions, it would not only bring measurability to the number of potential dropouts in each school, but also hold schools accountable in reducing dropouts. The Predictor could then be coupled with counselling sessions to ensure greater retention of students.

Some of the prospective organizations that can help implement these projects on a large scale include: The BMC (Bombay Municipal Corporation), Education Department, Sarva Shiksha Abhiyan Authorities and The Praja Foundation.

## Sharing Best Practices with other Countries

The research in this paper shows that this issue is not just India-specific. Dropouts are a big concern in most countries; especially those having a large economically disadvantaged population.
The Dropout Predictor and the Save-a-Dropout website could be used in other countries to first identify children likely to dropout ahead of time, and then to hold counselling sessions with help from private individuals, and save dropouts.

In order to market this concept in other countries, partnerships with NGOs around the world will be formed, to spread awareness in their countries.
Furthermore, The International Institute for Educational Planning (IIEP) under UNESCO could make the Save-aDropout make it a case study for neighbouring countries facing similar issues and further popularize the model.

## References:

[1] A Dropout Prediction Model That Highlights Middle Level VariablesD. Christopher Belcher \& Richard V. Hatley,Published in Journal 'Research in Middle Level Education', Volume 17, 1994 - Issue 2.
[2] Data-driven system to predict academic grades and dropoutSergiRovira, EloiPuertas, and Laura Igual,Published February 14, 2017 https://doi.org/10/1371/journal.pone. 0171207.
[3] Foundation, Praja. "White Paper: State of Municipal Education in Mumbai." PrajaFoundation: Making Democracy Work, 2015.
[4] Juneja, Nalini. "Primary Education for All in the City of Mumbai, India: the Challenge Set by Local Actors2001."unesdoc.unesco.org, 2001.
[5] Predicting Student Dropout: a Machine Learning ApproachLorenz Kemper, Karlsruhe Institute of Technology, 2018.
[6] Statistical Learning for the Prediction of School Dropouts J.H.C. Bunk, Utrecht University, 2016.
[7] Choudhary, and Hammayun. "Economic Effects of Student Dropouts: A Comparative Study." OMICS International, OMICS International, 15 Apr. 2015.
[8] Kattan, Bentaouet, et al. "Patterns, Consequences, and Possible Causes of Dropout in Upper Secondary Education in Mexico." Advances in Decision Sciences, Hindawi, 18 Mar. 2015, www.hindawi.com/journals/edri/2015/676472/.

## Appendix 1: Other Causes of Dropouts in India

## Systemic:

Location - school being far (they cannot afford buses)
After moving to high school (medium of choice not there)
Sanitation (no toilets for girls)

## Migrancy:

Migration/slum demolition so forced to move
Very high rental costs (half salary goes in rent)

## Teaching Quality:

Parents also not convinced that kids will get jobs (very poor education quality)
Teachers lacking motivation
Often overwhelmed with unruly behaviour

## Labour Laws:

Child labour law at age 14 (free compulsory only until $8^{\text {th }}$ by government)

## Cultural Issues:

Moms very busy with daily household work, no facilities, no gas, no pressure cooker, where are the clothes are getting washed, where do they dry, collecting water takes a lot of time (sometimes an hour minimum)
Gender component - girls required to work at home
Boys help father in tea stalls or garages

## Financial Issues:

Daily wage earners (hand to mouth)
Kids or family member contracting long-suffering illnesses (like TB) requiring costly treatment

## Child Issues:

Drugs and immoral activities
No role models, so poor grades
Often poorly behaved because of frustrations

## Appendix 2: Dropout Situation in Other Countries

The dropout situation is not just unique to India; it is prevalent in most developing countries among the lower economic strata of the society.

## Pakistan

Absence of essential facilities, poor quality of the education, over-populated classrooms, remissness of instructors and security issues in young girls' schools are found as significant reasons for student dropouts in Pakistan. Guardians are not interested to educate their kids; rather they need their kids to work and earn. Early marriage and security issues have made numerous youngsters dropout.

## Nepal

As indicated by Child Workers in Nepal (CWIN), a national NGO, the dropout rate is over $40 \%$, because of poverty, absence of educators, corporal punishment and inadequately managed schools. Corporal punishment, particularly by male instructors, including physical and mental maltreatment is a contributing component to dropout rates, as indicated by the Center for Victims of Torture (CVICT).

## China

Youngsters in provincial regions, where around half of the populace lives, have far lower intellectual and social abilities contrasted with their urban partners, showing them a way of dropping out of school before they can even say their very own name. One factor slowing down a child's advancement is the absence of parents.

## Malaysia

NGO My Skills Foundation chief S. Pasupathi said an investigation uncovered that almost 7,000 students were dropping out from school yearly, for reasons like dysfunctional guardians, low confidence, peer group pressures, living in a deplorable environment and feeling lost in the school environment. This, he stated, prompts chronic drug use, drinking, immoral activities committed by those somewhere in the range of 13 and 26 years old.

## Mexico

The dropout rate in Mexico is at $14.5 \%$. The principle purpose behind dropping out was financial constraints, for which the children were expected to work. Alternate reasons included boredom, redundancy of class and being over-age.

