Project Proposal – Examining Drop-out rates in India

(Submitted by Shwetha J Parvathy)

Right to education Act (RTE) passed in 2009 in India guarantees free and compulsory education for all children up to the age of 14. RTE has also played a pivotal role in India achieving universal enrollment (~98%), as indicated by the Annual Status of Education Report. While kids are in school till the age of 14, data from Ministry of Human Resource and Development show a jump in the drop-out rate in 9th grade and beyond, once they are out of the RTE umbrella. Dr. T.V Sekher, in his paper "Factors Leading to School Dropouts in India: An Analysis of National Family Health Survey-3 Data" examines the factors leading to drop-out of children by examining household characteristics like religion, literacy of parents, income etc. Another main reason for this drop-out is that children move up the ladder till 8th grade regardless of whether they are actually learning or not, because of the no-retention policy RTE mandates (nobody can be held back in the same class for an extra year). Once they reach 9th grade, they are too far behind to catch up with the grade level curriculum and are faced with the task of taking exams. This is reinforced by ASER data which shows that 50% children in Grade 5 cannot read a Grade 2 level text. A combination of social, economic and children's performance in primary grades are factors responsible for drop-out rates as examined by the paper.

For the purposes of this project, I would like to examine how the social, economic and academic factors affect drop-out rates across states and whether some factors play a more important role in determining drop-out than others by looking at data from 2012-2015 for all states.

Datasets for the project

- Dataset for this project will be created by scraping open data from the Ministry of Human Development India website, Wikipedia and Census 2011 website.
- All the data to be scraped are made available publicly as tables in the official websites.
- I will also use ASER data which will have to extracted from a pdf format.

Final dataset structure, Analysis and Visualisation

The below table represents a snapshot of the final dataset that will be used for analysis after scraping and joining data from various sources. Not all variables that will be considered for analysis have been included here.

State	Year	Drop-out rate	religion	literacy	housing	Ownership of house	Net domestic product

I will conduct preliminary data exploration to determine any correlation between variables and provide plots for any interesting observations. I will try k-nearest neighbors, linear regression and random forest model to see which one gives the best model and which one makes most sense in terms of understanding the factors affecting drop-out rates. I will provide an overview of drop-out rates across states in a visualization and provide relevant figures for the most successful model based on which the results will be explained.

Defining Success

The success of this project will be in identifying the model that best predicts drop-out rates, identifying the important factors affecting drop-out rates across states. Further, summarizing the findings with relevant visualizations will also count as a success of the project.

References

- Sekher, T.V et.al, 2014. "Factors Leading to School Dropouts in India: An Analysis of National Family Health Survey-3 Data", Journal of Research & Method in Education
- Annual Status of Education Report 2014
- Census of India 2011