PPOL 670 Project Proposal

Primary school drop-out rates across Indian states

Ruchika Bhatia

India passed the Right to Education Act in 2009 which promises free and compulsory education to children between the age of 6 and 14 years. Today,India has achieved a near 100% enrollment rate in primary school. However, this does not show the true picture of the Indian education system [1]. Even though students get enrolled in schools, they tend to drop out within a couple of years, even before completing their primary school education. In 2014-15, the primary school drop-out rate was 4.13% with large variations across states. Some states like Assam had a drop-out rate of more than 15% while others like Goa had a drop-out rate of 0.2% [2].

I aim to study the factors that influence primary school drop-out rates and build a model that can explain the majority of variation in drop-out rates across Indian states using 2011-2015 data. The unit of my analysis will be state-year and my variable of interest will be drop-out rates in primary school (class I-V).

The independent variables that influence the drop-out rates can be broadly categorised into the following three groups:

1. School-level factors : Average distance of the nearest school from household, percentage of students enrolled in government schools and private schools, pupil-teacher ratio, percentage of children availing mid-day meals, percentage of single-teacher and single-classroom schools, percentage of schools with basic facilities like electricity, drinking water, girls' toilet
2. Household-level/State-level factors : Net enrollment rate in primary school, enrollment in pre-primary education, state GDP per capita, poverty rate, number of children per female, literacy rate, literacy rates of females, prevalence of child labor, budgeted education expenditure per child
3. Cultural Factors : Prevalence of child marriage, ethnic and linguistic fractionlization index, sex ratio

I found these variables from the preliminary literature review. I still need to cover more literature on this topic to get a more exhaustive list of variables that affect primary school drop-out rates.

I aim to use data from 4 broad areas:

1. Census Data of India (http://censusindia.gov.in/2011-Common/CensusData2011.html)
2. Open Government Data Platform of India (https://data.gov.in)

3. Ministry of Human Resource Development, India (https://mhrd.gov.in/statistics-new)
4. National Institute of Educational Planning and Administration (http://udise.in/)

The data is available to download and is also present on the websites in form of tables and reports for use. Some tables will need to be directly scraped from the reports. State level statistics from the data will have to be all merged in the master dataset. Further, certain variables like prevalence of child labour (as a percentage of total children in the state), ethnic and linguistic fractionlization index will have to be calculated using variables from different data sources.

Preparation of data will involve scraping the data, merging multiple datasets and creating new variables. I also plan to use ggplot2 for multiple data visualisations like variations in drop-out rates across states and their interaction with some of the key variables like state GDP/capita, enrollment rate in pre-primary education etc. Further, I plan to explore multiple linear regression, k-nearest neighbors, and random forest to build a machine linear model.

A successful project for me would have easy to interpret data visualisations which can explain the problem being addressed in the project and also a model that can explain significant variation in my independent variable i.e the reasons for wide variability in drop-out rates in primary schools across states in India.

References:
1.https://data.worldbank.org/indicator/SE.PRM.NENR?locations=IN
2.http://udise.in/ElementarySRC-2013-14.htm?ay=2015-16