TETRA TECH
International Development

Machine learning in education:
Identifying girls at risk of dropping out of school

How we applied machine learning techniques to help identify which girls might not return to school following Covid-19

We trained prediction models to look for patterns in historical data to learn how to correctly classify girls as being at risk of drop-out, which we tested in Kenya and Nepal.

School closures: a risk for girls' education

The Covid-19 pandemic brought new urgency to the UK's Girls' Education Challenge (GEC), which has supported marginalised girls since 2012. When Covid-19 hit, and schools closed, we developed a machine learning model to identify which girls might not return to school , which we tested when schools reopened in Kenya and Nepal – with a view to improving targeting of responses for the future.

How machine learning works

Training: what are the characteristics of girls who dropped out? Were there any patterns among those girls who dropped out in the past?

Predict: what are the characteristics of these girls, and which of these matter?

Validate: go to girls' schools to find out which girls did not return and why

How we did this

Our GEC project partners – EDT in Kenya and Mercy Corps in Nepal – provided us with information about the girls they supported. We combined this with country-wide household survey data to train a model. Put simply, we used existing data about girls who dropped out of school to train the model to classify girls into those who are likely to progress, and those who are likely to drop out of school. So, if historically girls who repeated school grades a lot dropped out, then any girl who repeated a lot in the project data was likely to be predicted to drop-out. We tested these predictions by visiting schools and talking to the girls and their teachers.

How accurate was it?

Our study finds that 8.6% of girls did not return to school in Kenya, while 26.2% of girls did not return in Nepal. Our model predicted that girls were more at risk of dropping out of school if they: Had already repeated a year of schooling ; came from less-educated families; were part of a bigger household or family; were involved in farming more than their peers; were married.

Our initial predictions (on historical data) was accurate for eight out of every ten girls in Kenya. In Nepal, this dropped to seven out of every ten, with higher numbers of girls getting married and dropping out – a predicted risk factor, suggesting that a future model which can capture 'learning' from the first set of predictions could have high accuracy.

Prediction Model	Kenya				
	Broad	Strict	Broad & Strict, only Secondary	Broad, only females	Strict, only females
Overall model accuracy	73.4%	76.9%	79.8%	75.7%	77.5%
Accurate prediction of dropout	40.0%	26.7%	6.7%	33.3%	26.7%
Accurate prediction of progression	76.6%	81.6%	86.7%	79.7%	82.3%

The broad model includes whether children had dropped-out previously and the strict model does not.
The first-stage prediction models were better at classifying girls who will progress than those who will drop out. This was in part due to the reasons for drop-outs, which were often driven by changes in the girls' circumstances - many girls who had been unmarried when the project data was last collected had since married and dropped out (i.e. the risk had been realised).

Prediction Model	Nepal			
	All grades	Grade 10	All grades, only females	Grade 10, only females
Overall model accuracy	65.8%	49.7%	70.8%	64.6%
Accurate prediction of dropout	18.1%	34.9%	1.7%	25.6%
Accurate prediction of progression	82.8%	54.5%	95.4%	77.3%

Key learning points

Our findings suggest predictive algorithms **can be a useful tool** for classifying girls who may be at risk of not progressing – but require learning loops to offer greater accuracy. Given the magnitude of the Covid-19 shocks, having 70-80% accuracy based on old data (and correctly identifying the risk factors which drove drop-out) means that this work has early promise. However, the initial model was far better at classifying girls who progress than girls who drop out, due to the changing nature of girls' circumstances. This means that improvements to the models will require more regular data collection, to allow projects to more accurately monitor the girls' situations.

The models can also improve with increased samples, and further exploration of the trade-offs between progressing in school and dropping out. The more data we use to 'feed' and train the model (up to a certain point), the better it gets. But this initial work suggests that machine learning has the potential to be used as part of an 'early warning system' to help projects and schools identify and provide early targeted support to students at risk of dropping out.

For more information

This research was carried out by the Independent Evaluation Team of the Girls' Education Challenge Fund Programme.

If you would like to find out more about our research, please contact:
Sophie Amili – Tetra Tech International Development Europe
Sophie.Amili@tetratech.com I +44 (0)207 837 2881
www.intdev.tetratecheurope.com
Tetra Tech International Development Europe I 107 Gray's Inn Road, London, UK, WC1X 8TZ I +44 (0)20 7837 2881

TETRA TECH
International Development

107 Gray's Inn Road, London WC1X 8TZ, UK
+44 (0) 20 7837 2881 | intdev.tetratecheurope.com

Tetra Tech International Development has regional offices in
United Kingdom | Poland | Kenya | Nigeria | The Netherlands | Turkey

Learn more at intdev.tetratecheurope.com