Nicole Abraham

Machine Learning: Who Trains Whom

Machine Learning in Child Welfare Call Centres

When a Child Welfare Call Centre is called, call operators have a limited amount of time to decide whether or not to send someone to check in on the reported child. They do not have the time needed to complete a thorough investigation of each child's living situation, and because of this, some children who need help may be passed over. To try to solve this problem, Allegheny County in Pennsylvania chose to invest in machine learning algorithms that will ideally help determine a child's risk factor.

When an operator receives a call, they must choose whether to screen a case in, allowing a Child Protective Services worker to be sent to further investigate a situation and decide what, if any, next steps to be taken, or to screen a case out, where nothing further is done. Reports of physical violence, sexual abuse, and other reports of abuse are automatically screened in, however other cases are often much harder to determine. Cases of potential child neglect are often made more difficult because in many cases the family is trying their hardest to provide for the child, but they simply do not have the resources. Because of the limited time an operator has to research a child's situation, often only an hour at most and more frequently half that, important details of a child's life may be missed (Hurley). This can lead to both not sending help to a child who needs it and to sending help to a child who does not, when resources are limited and a visit from a Child Protective Services worker can interfere negatively with family dynamics. The machine learning algorithm created to help will evaluate a child's risk of being reported again or being removed from their home. From this, a score representing the risk factor is shown to the operator to help them determine if a case should be screened in or out.

To develop this algorithm, researchers looked at referrals to the Allegheny call centre between 2010 and 2014. This allowed for a significantly large data set, as well as at least 2 years of follow-up data for each child referred. The data set is publicly available upon request, deidentified for privacy, a good practice for machine learning. Both algorithms, predicting re-referral and removal from home, looked at a variety of factors about the child, any siblings, others living with the child, legal guardians, and their significant others where applicable. These variables include but are not limited to: time spent in county jail or juvenile probation, whether they have received public welfare in the past 2 years, any behavioural health programs they have been involved with, and census neighbourhood poverty indicators (Vaithianathan). Models were tested with and without race as a factor, however, race was not included in the final models as it was heavily correlated with other variables and did not improve the models' ability to predict.

After the algorithms are developed, a easily useable program was created for use by the call centre operators. After the information is input into the system, an operator will press a button that generates the risk score for both the algorithm predicting the risk of a rereferral and the algorithm predicting the risk of a child being placed in a different home. Whichever score is highest will be displayed to the operator on a coloured scale, and if the risk is high enough the child will be marked as a mandatory screen-in. While operators do have the power to override this, any case of an override will be reviewed closely to ensure the right decision was made. These practices should help increase accuracy while still ensuring a human has the final call, particularly because these cases are so difficult to classify.

In addition to having publicly available data, other good machine learning practices shown in the development of these models include a publically available list of variables used in each model. This allows people other than those who worked on the project to understand what is analysed to create the risk score. Additionally, both models are validated externally using data of critical events, such as fatalities and near fatalities, in children, in order to determine the accuracy of the models. This is particularly important, as it shows that a source not involved in the development of the models and that has no stake in the success of their project can vouch for their accuracy. As the models are actively used in the field, regular quality assurance analyses will be performed to ensure that the model remains accurate as society changes. If they receive poor performance reviews, the weights of variables in the models will be adjusted, or the models will be removed from use and other solutions will be explored. Finally, the research and development behind these models were funded publicly by the Centre for Social Data Analytics. This means that there isn't corporate pressure to profit off of this model, and allows for the model and data to be publicly available without corporate confidentiality issues.

One challenge provided by these models is providing consent in how a person's information is used. Much of the information used is available as public record, however, some of it is only available through the call centre's records. In order to develop this model using the provided data, the researchers assumed that a call to the centre provides grounds for further research on cases (50 Vaithianathan). Additionally, this information has been de-identified for privacy reasons, particularly because of the social stigma of being identified as "at risk". As with any machine learning model that deals with classification, there is also a chance for false negatives and false positives. False negatives are a failure to identify a high-risk situation as high risk, which could lead to a child not getting the help they need. A false positive is identifying a low-risk child as high risk, which can lead to the misuse of valuable resources and may also place the pressure of a Child Protective Services inquiry on a family. Even if a human operative may have correctly identified the situation, because the algorithm is giving contrasting results, it may influence operators to act against their instincts. An alternative option to this model is to hire more call operators to ensure that each referral gets an appropriate amount of time for the case to be reviewed. Unfortunately, because funding is fairly limited, Child Protective Services does not currently have the resources to make this concept reality. Until their budget is increased drastically, something fairly unlikely in the current government system, using machine learning algorithms to assist call operators is the best available solution.

Generally, the media perception of this project has been cautiously positive. A *New York Times* article, titled “Can an Algorithm Tell When Kids Are in Danger?”, talks about the challenges and lack of resources of the current system. It does make note of many difficulties that have arisen in past attempts to use machine learning to classify the risk levels of children, including expense, inaccuracy, and a lack of transparency, however, the article also acknowledges that the researchers for this algorithm are taking proper steps to increase transparency and improves the accuracy and implementation (Hurley). The *Pittsburgh Post-Gazette* published an article titled “Allegheny County DHS Using Algorithm to Assist in Child Welfare Screening”, which also touched on the impressive vetting and transparency of the algorithm. It comments on the applicability of this model to other counties. The research and development cost $781,073 along with an additional $224,858 for evaluation, which is not an amount all counties can afford (Giammarise). It suggested adapting this algorithm for other counties, but that requires funding as well and may not be as accurate, if it is accurate at all. In contrast, other media outlets are more critical of the impact of this technology. An article from *Wired*, titled “A Child Abuse Prediction Model Fails Poor Families”, discusses two specific cases, one where operators predict a situation to be high risk while the algorithm predicts it is low risk, and another where the operators believe the situation is low risk while the algorithm classifies it as high risk (Eubanks). While it is entirely possible that this is a common situation, it is also possible that the article writer chose specific cases in order to portray the algorithm in a negative light. It is difficult to tell which is the case without performance statistics from the Allegheny Call Centre.

A machine learning algorithm may be a solution to improve call centre's abilities to best help the children in need of help. Additionally, the practices behind this particular model are widely considered impressive in terms of validation and transparency. However, the expense may be a barrier to widespread implementation, and the model is not always accurate. Using its classification power to assist human operators instead of replacing them is a great compromise, as working together will likely improve the accuracy compared to either working alone. Overall, the development and implementation of this model has followed good and ethical machine learning practices and is helping to increase the accuracy of classification in the Allegheny Child Welfare Call Centre.

Works Cited

Eubanks, Virginia. “A Child Abuse Prediction Model Fails Poor Families.” Wired, Conde Nast, 31 Jan. 2018, www.wired.com/story/excerpt-from-automating-inequality/.

Giammarise, Kate. “Allegheny County DHS Using Algorithm to Assist in Child Welfare Screening.” Pittsburgh Post-Gazette , 8 Apr. 2017, www.post-gazette.com/local/region/2017/04/09/Allegheny-County-using-algorithm-to-assist-in-child-welfare-screening/stories/201701290002.

Hurley, Dan. “Can an Algorithm Tell When Kids Are in Danger.” New York Times, 2 Jan. 2018, www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html?module=inline.

Vaithianathan, Rhema, et al. Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions: Allegheny County Methodology and Implementation. Centre For Social Data Analytics, Apr. 2017.