Can Machine Learning Help Save Baby Lives: A Study on Prenatal Case Management

Juan David Argüello Pazmiño; Manisha Jha, Arshdeep Kaur, Idrishan Kavak, Tanyaradzwa Mushohwe, Yi Zhang

Best Model Obtained (Model C from Code)

- R-squared: 0.720
- Adj. R-squared: 0.711
- Training Score 0.7219
- Testing Score: 0.6753

Introduction

Low birth weight, which is a key determinant for infant mortality and illness (Institute of Medicine,1985) refers to babies born weighing less than 2500g (<5 pounds, 8 bounces) as compared to average newborns that weigh about 3,600g (8 pounds). Over the years, several interrelated maternal and fetal factors have been associated with low birth weight, commonly a result of pre-term birth, a major adverse birth outcome (Ahmadi et al., 2017). The known risk factors today include hereditary impacts such as mother's age, race and genetic predispositions, as well as lifestyle influences such as alcohol consumption and tobacco smoking, poor nutrition, inadequate health care, low social economic status and stress among many other (Pan, 2017).

According to a study by Stanford Children's Health (2019), over 8% of new born babies in the US have a low birth weight and the associated health complications including respiratory, neurological conditions and even lead to sudden death - a prevalence that is on the rise. Compared to other developed countries in the world, the US has significantly worse adverse birth outcomes which have extensive personal, financial, and developmental consequences for the mother and child broad as well as public health financial implications, accounting for 47% of hospital costs in 2001 (Pan, 2017).

18 years later, the adverse effects associated with low birth are still a notable challenge in the US, due to the limited public health resources available for intervention programs. To date, public health administrators tend to provide services on a first-come, first-served basis or use a threshold based on socioeconomic status or other relevant risk factors to determine eligibility for assistance – which does not necessarily target or prioritize the most at-risk groups (Pan, 2017).

The Problem Statement and Opportunity

Low child birth weight is such a major public health challenge and cost in the US today, and on the rise. Therefore, we proposed leveraging the predictive power of Machine Learning algorithms to determine the highest impact factors affecting new born birth weight to help identify the most at-risk mother groups. This presents the opportunity to drive data driven resources allocation and intervention efforts that are better tailored and targeted for pre-natal case management – reducing and potentially mitigating associated negative socioeconomic effects.

Hypotheses:

Based on the evidence shown in previous studies and the available Literature (including Ahmadi et al., 2017, Institute of Medicine,1985, Pan, 2017, Stanford Children's Hospital, 2019) we hypothesize that the following factors have the highest effect on child birth weight:**1. drug usage**: a mother drinking alcohol and/or smoking cigarettes during pregnancy has a significant negative impact on the child's birth weight; **2. age of mother:** if a mother is older than 35 years, this has a significant negative impact on the child's birth weight;3. **race:** if the baby is black (meaning having at least one parent of African-American heritage according to Davis, 1991), they are more likely to have lower birth weight; 4. **pre-natal care:** if the mother starts her doctor in the first trimester and regularly visits throughout her pregnancy, this has a significant positive impact the child's birth weight

Assumptions:

This data set was collected and reported by a skilled and trusted source, therefore it was assumed that there were no major data collection and entry errors made – resulting in good data quality for Machine Learning modelling.

Based on the observation that 95% of the sample size (196 samples) were drinking or smoking cigarettes, and 90% did both, we assume that the samples were derived from a high drug usage population were many of the pregnant women consumed one or both of these drugs during pregnancy. Therefore, the results, insights and recommendation made in this paper are specifically for a high drug usage population.

Model Training and Evaluation

The aim of this study is to evaluate the positive predictive value of machine learning algorithms for early assessment of adverse birth risk among pregnant women as a means of optimizing the allocation of public health resources (Pan, 2017).

Our modelling objective was to compare the performance of 2 machine learning algorithms for predicting child birth weight (namely OLS Linear Regression which is based on the relationship/correlation analysis of risk factors and K-Nearest Neighbors which is based on the similarity among risk factors) which we determined as most appropriate for this problem type.

We set up the prediction problem as a binary classification problem: whether a woman will experience an adverse birth or not. We used 15 binary prenatal risk factors as predictors, called "features" in Machine Learning, to ensure that we included only features present before child birth (Ian Pan, 2017).

Results & Insights

Due to the correlative nature of the risk factors assessed, the OLS Linear Regression Model proved to be the more power algorithm for this challenge.

Based on our model, the risk factors that most influenced a low child birth were smoking, drinking and the mother's age (above 50 years) – and further negatively impacted by the father's age and education, when both parents are black.

Therefore, the highest risk group for having a child with low birth weight are black women who smoke and drink over 50 years old, who's fathering partner is also black and has low or no education.

Looking at our initial hypotheses, we see that for our population where drinking and smoking rates were extremely high, this model predicted that these 2 lifestyle risk factors most negatively impact a child's birth weight (independent of all other risk factors involved) which validates our literature-based hypothesis 1.

Recommendation: This suggests that is it is critical to prioritize intervention among pregnant women (or women to trying to get pregnant) who are smoke and/or drink – whether educative, rehabilitative or both.

Secondly, mother's age also significantly negatively impacts the child's birth weight (independent of other risk factors), which validates our hypothesis 2. A deeper analysis of the predicators using feature engineering showed that mother's age becomes a serious negative risk factor for low child birth weight after the age of 50 as show below:

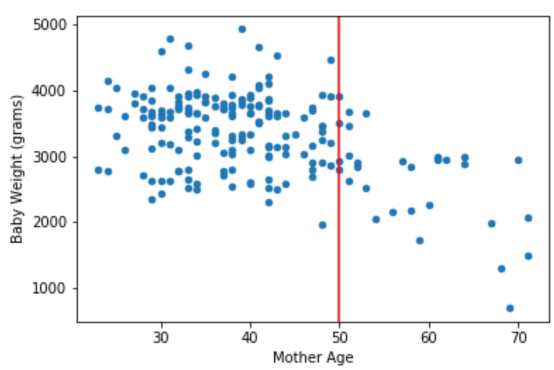

Figure 1: The relationship between birth weight and the mother's weight

This show that age 50+ which is when most have entered menopause, is a very risky age women to become pregnant, and could have adverse effect on the child's health. **Recommendation:** Therefore, it is important to also prioritize education initiatives targeted towards women in this age group. We propose advocating for the use of birth control among women in this age group who no longer want to have children as it is not uncommon for women who are in menopause to get pregnant. Among this risk age group of women who would like to still fall pregnant, we recommend education on extra healthcare, lifestyle and nutrition adjustment needed to reduce this risk factor for the child.

Beyond the scope of our hypotheses, father's age and education also showed to have independent moderate negative impact on child birth weight. This was an unexpected outcome of the study, and likely to be related the genetic influence of older age as we saw with older mother's age on the part of father's age, and lower social economic status associated with low father's age (assuming family structure where the male is the main income earner for the family). **Recommendation:** Considering the moderate negative effect of father's age and low education, we propose that the educative programs we have suggested above target and cater to both the mothers and fathers as both parents contribute risk factors.

Interestingly, hypothesis 3 on the negative risk factor based on race, had insignificant impact on low child birth weight independently, and only only showed impact in combination to smoking, drinking, mother's age (above 50), father's age and father's education – in the instance that both parents are black. This result was contrary to the body of literature that presents race has one of the major independent risk factors associated with birth weight. However, considering that this population was of high drug users, this result may suggest that drug use and the parent's age and socioeconomic status affect birth weight more than race. **Recommendation:** Therefore, in light of the fact that resources allocated to this public health issue are limited, we do not focusing initiatives by racial groups – except for community groups that have high rates of "black women who smoke and drink over 50 years old, who's fathering partner is also black and has low or no education" where there is a noted prevalence of low child birth weight. This may be the case in some low-income neighborhoods with large black people populations – however, this is out of the scope of this study and would have to be validated in a separate study.

Conclusion

With limited resources allocated to public health, it is critical to be able to make data-based decisions on how to best target the groups to prioritize and provide intervention initiatives for to fight the rising challenge of low child birth weight in the US.

This simple study has showed that Machine Learning has the potential to predict more powerful analytics that the traditional human based ones we use today, in a much faster and cheaper way – with high impact results in society and the business world.

Although this study was limited by the low sample size, and use of 2 algorithms – the results derived give us confidence that Machine Learning is a robust and valuable predictive tool, which we recommend the public health sector to serious consider as part of the pre-natal case management strategy.

REFERENCES

- Ahmadi, P., Alavimajd, H., Khodakarim, S., Tapak, L., Kariman, N., Amini, P., & Pazhuheian, F. (2017). Prediction of low birth weight using Random Forest: A comparison with Logistic Regression. Journal of Paramedical Sciences, 8(3), 36-43. doi: https://doi.org/10.22037/jps.v8i3.15412
- Davis, F.J. (1991). Who is Black: One Nation's Definition. Retrieved from https://www.pbs.org/wgbh/pages/frontline/shows/jefferson/mixed/onedrop.html
- Flaxman, A. D., & Vos, T. (2018). Machine learning in population health: Opportunities and threats. PLoS medicine, 15(11), e1002702. doi:10.1371/journal.pmed.1002702
- Institute of Medicine. (1985). Preventing Low Birthweight: Summary. Washington, DC: The National Academies Press. doi: https://doi.org/10.17226/512.
- Pan I., et al. (2017). "Machine Learning for Social Services: A Study of Prenatal Case Management in Illinois", American Journal of Public Health 107(6), 938-944. doi: 10.2105/AJPH.2017.303711
- Stanford Children's Health. (March 2019). Low Birthweight. Stanford Children's Health. Retrieved from https://www.stanfordchildrens.org/en/topic/default?id=low-birthweight-90-P02382