EXAMENSARBETE The use of machine learning to predict adverse birth outcomes: Empirical real world evidence from a human cohort study in Adama, Ethiopia
STUDENT Stephanie Bol
HANDLEDARE Anna Oudin, Andreas Jakobsson
EXAMINATOR Frida Sandberg

Predicting adverse birth outcomes with machine learning in Ethiopia

POPULÄRVETENSKAPLIG SAMMANFATTNING **Stephanie Bol**

The number of adverse birth outcomes has long been a problem in Sub-Saharan Africa. Most of these cases are mainly preventable, however, 84% of all stillbirths around the world still occur in this region. The technology of machine learning has now emerged as an alternative to understanding the connections between environmental factors and adverse birth outcomes. Hopefully, this can help us find the key factors of the region and prevent future cases of stillbirth and neonatal death.

A study that gathered data from the city of Adama in Ethiopia could not find any statistical significance between environmental factors and pregnancy complications. To determine if machine learning could draw other conclusions, an algorithm was built that trained on the gathered data and six selected features. These features were if the mother was hospitalized after birth, the place of delivery, previous pregnancies, age, level of education, and living situation.

The first attempt showed poor results due to the dataset being imbalanced. Meaning that the dataset is only composed of 5.5% cases of stillbirth or neonatal death. This imbalance is common in epidemiological data and can pose problems when not accounted for. To handle this issue, techniques to balance the dataset were used. One of these called oversampling generates new data points. It makes the minority class as large as the majority class with a technique called SMOTE. It increased the result remarkably making it an alternative when dealing with highly imbalanced cases.

After finding the best method to improve the given dataset, four different algorithms were used to classify the data. The classifier random forest yielded the best result with an accuracy of 85% with the given dataset.

Comparing the findings of the study to similar settings, this result was slightly lower than other comparable studies. The notable difference was the number of data points where this study was lacking with only 1615 participants. The features chosen were similar to already common risk factors for stillbirth and neonatal death with age, level of education, and first pregnancy seemingly having a big impact.

In conclusion, with further work, more data, and higher accuracy, machine learning can be of great use to predict adverse birth outcomes. Hopefully, the algorithm can find those who are at risk based on their environmental and preexisting health factors to then put in preventable measures to lower the risk of adverse birth outcomes in countries such as Ethiopia. However, to improve the evaluation scores of these models and identify the most essential characteristics to analyze for this region, additional research is required, particularly with larger study populations.