Who Gets Health Benefits? An (Un)biased Machine Learning Prediction

Ghaida Alsadah, MS[1], Janusz Wojtusiak, PhD[1], Priyanka Anand, PhD[1], Laura Dague, PhD[2], Kathryn Wagner, PhD[3]

George Mason University[1]; Texas A&M University & NBER[2]; Marquette University[3]

Introduction

There is a need to accurately identify who receives health benefits as part of their employment. Currently, few datasets have information about who receives health benefits[1] and all are survey-based. The American Time Use Survey (ATUS) is a nationally representative survey that includes information on how people spend their time and includes information on health benefits and paid leave. A more comprehensive Survey of Income and Program Participation (SIPP) provides information on income, employment, government program participation, and health status (i.e., hospitalization or childbirth) but lacks health benefits. The presented research is part of a project that studies the relationship between health benefits and health status, one needs to estimate the likelihood of a given person receiving such benefits. The presented work reports on the development, testing and potential biases in such models.

Methods

Model development: 2011, 2017 and 2018 ATUS panel data (~15K datapoints) have been preprocessed to include variables and coding that is also found in SIPP (~3M datapoints), including: gender, age, race, US citizenship, marital status, place of birth, home ownership, number of children below 18, disability status, education, income, having multiple jobs, the availability of state paid leave policy, and grouped industry and occupation codes for primary and secondary jobs. The variables were coded with one-hot coding when appropriate. Traditional machine learning models (random forest, gradient boost, logistic regression (plain, L2, L2 and elastic net), k-nearest neighbors, and decision tree) have been constructed on 80% training portion of ATUS.

Internal model evaluation: The remaining 20% of the ATUS data were used for internal testing of the models. Traditional statistical methods were used to evaluate models: area under ROC (AUC), precision, recall and accuracy.[2] Additionally model calibration and sensitivity were evaluated. Finally, LIME algorithm has been repetitively applied to high- and low-probability cases to understand the most important characteristics influencing the likelihood of having health benefits. While several of the models were close in terms of AUC (within 2%), the other metrics (mainly calibration) led to selection of gradient boost models as the best.

External validation: To externally validate the selected models, SIPP data were used. SIPP does not include information about health benefits (which is why it is predicted), thus it was not possible to calculate standard accuracy metrics. Instead, when ATUS and SIPP are weighted to represent overall US population, the distribution of predicted values should closely match. Probability (PDF) and cumulative probability distribution functions (CDF) were constructed for ATUS test set and SIPP. The distributions were first compared visually and then using Kolmogorov-Smirnov test and Lullback-Leiber divergence.

Potential bias evaluation: Data and model bias have been evaluated on protected variables: gender, race age, and disability. Regression models were created to assess societal bias in the data and model predictions. Accuracy measures of models were assessed in subpopulations defined by these variables.

Results

The developed models had reasonably high accuracy of prediction with AUC (0.77±0.04). The models were well calibrated and not sensitive to changes in input. External validation demonstrated that the models led to distribution of predicted probabilities almost identical between testing ATUS and SIPP datasets. There was no difference in accuracy of models within subpopulations. In contrast, the tests clearly indicated societal bias in the data and models: U.S. citizens are more likely to receive health benefits, so are males, and those who identify as white and Asian. Gender and race should not affect the probability of receiving health benefits when controlling for other person's characteristics (e.g., job type). No algorithmic bias has been added to the predictions as compared with bias already existing in data.

Conclusion

It is possible to predict health benefits and use these predictions on independently collected data. The dataset and models are biased, which represents societal bias – the way the current practice of offering benefits is. However, this specific type of bias is not a concern for the specific models developed with the goal of predicting the world as it is, not the world as it should be.[3] If the goal was to develop models for deciding who should receive benefits, if such application is needed, the constructed models would need to be free of the present biases and properly tested for fairness.

References

1. Batchelor A. Paid Family and Medical Leave in the United States: A Data Agenda.
2. Wojtusiak J. Reproducibility, Transparency and Evaluation of Machine Learning in Health Applications. InHEALTHINF 2021 (pp. 685-692).
3. Mitchell S, Potash E, Barocas S, D'Amour A, Lum K. Algorithmic fairness: Choices, assumptions, and definitions. Annual Review of Statistics and Its Application. 2021 Mar 7;8:141-63.