Predicting Intimate Partner Violence in South Asia with Data from the Demographic and Health Surveys

Qiwei Lin[1]

[1]Harris School of Public Policy
University of Chicago

June 4 2021

Motivation

- Intimate Partner Violence (IPV) during Covid-19
  - Concerns about a potential surge in IPV cases
  - "Pandemic within a pandemic" (Evans, Lindauer and Farrell, 2020)
- To inform policy interventions:
  - How to identify potential victims of IPV to provide social support and protection? $\Rightarrow$ Prediction by machine learning
  - What socioeconomic and demographic features help distinguish victims from non-victims? $\Rightarrow$ Examining important predictive features

Data

- DHS: nationally representative surveys of women aged 15 to 49 and currently in union.
  - information is available on women, their partner and households
- Most recent available surveys:
  - India (2016)
  - Nepal (2016)
  - Pakistan (2018)
- 740,000 observations available initially but only 73,000 left after filtering
- Distribution of observations:
  - 90% India
  - 5% Nepal
  - 5% Pakistan
- Unbalanced sample: more women who did not experience IPV (70%)

Computational Approaches

- Implementation: mainly Dask with some functions from sklearn
- Predictive Analytic:
    - Model Building and Tuning:
        - Logistic Regression (LR) and Random Forest (RF)
        - Parallel grid search with five-fold cross-validation
    - Model Evaluation:
        - Prediction Accuracy on Testing Data
        - AUC Score
        - Recall Score
- Feature Importance:
    - Mean Decrease in Impurity for the pooled dataset
    - Permutation Importance for country-specific testing datasets

Results

- Prediction
  - Prediction Accuracy: **76.4%** (RF) vs 75.5% (LR)
  - AUC Score: **0.748** (RF) vs 0.745 (LR)
  - Recall Score: 0.578 (RF) vs **0.584** (LR)
- Feature Importance
  - Top Five Predictive Features in the Pooled Data:
    1. Family history of IPV
    2. Partner's consumption of alcohol
    3. A woman's attitude toward wife-beating
    4. A woman's years of education
    5. A woman's family's wealth level
  - Partner's alcohol consumption is not predictive in Pakistan
  - Women's attitude toward wife-beating is not predictive in India

More can be found at this repo

Reference

Evans, Megan L., Margo Lindauer and Maureen E. Farrell. 2020. "A Pandemic within a Pandemic — Intimate Partner Violence during Covid-19." *New England Journal of Medicine* 383(24):2302–2304.