

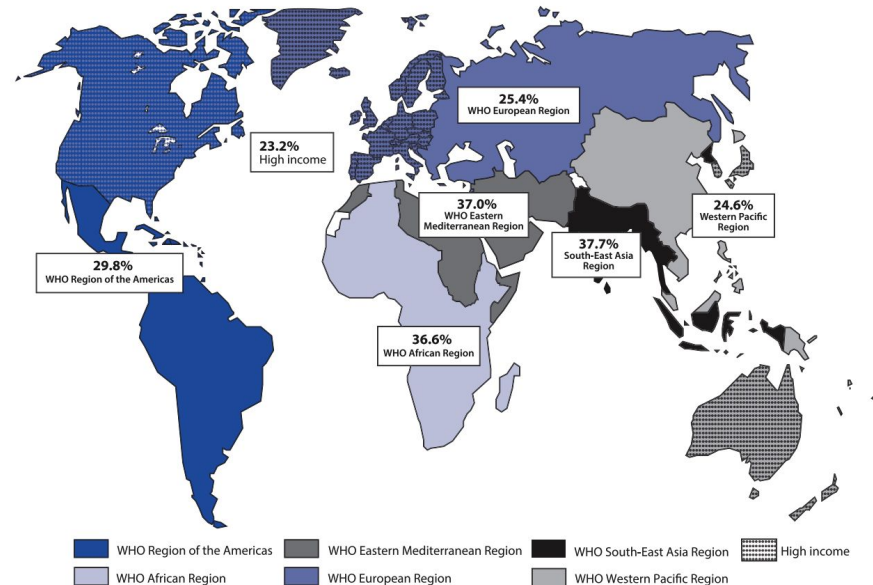


Identify Women's Risks of Intimate Partner Violence with Evidence from South East Asia

Yangzhou Ou (yangzhou) Xiaoting Sun (sunxt)

Background and Solution Overview

- Intimate partner violence (IPV) is one of the most common forms of violence against women and includes physical, sexual, and emotional abuse.
- South East Asia is one of the areas with highest prevalence of physical and sexual IPV of around 37%.
- The goal of this study are:
 - To report the prevalence of IPV among ever married or cohabiting women in Pakistan, Cambodia, Philippines, Maldives and Nepal
 - To develop classification algorithms on individual-level variables and couple-level variables to predict whether a specific woman is prone to IPV in these five countries;
 - To identify important features associated with experiencing physical, sexual or emotional IPV.
- Policy impact

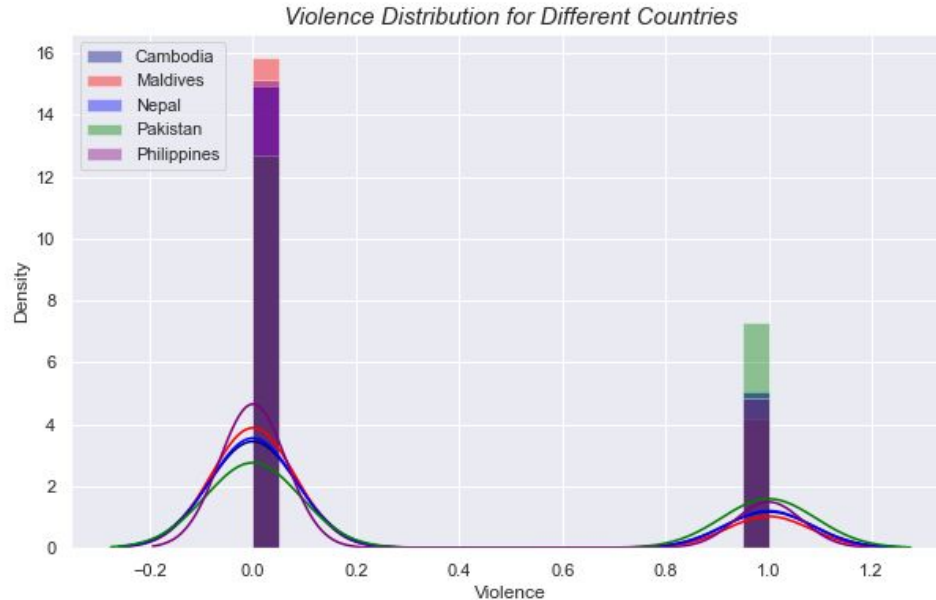


Data Source

- Main data source is Demographic and Health Surveys (DHS) conducted by the United States Agency of International Development
- DHS program had developed a standard module and methodology for the collection of data on domestic violence by 2000.
- The standard module was used in all of the countries examined in this report: Cambodia(2014), Maldives(2016), Nepal(2016), Pakistan(2017), Philippines(2017)
- We filtered data using the variable if_union to only include women who are currently or formerly married (or live with a partner)

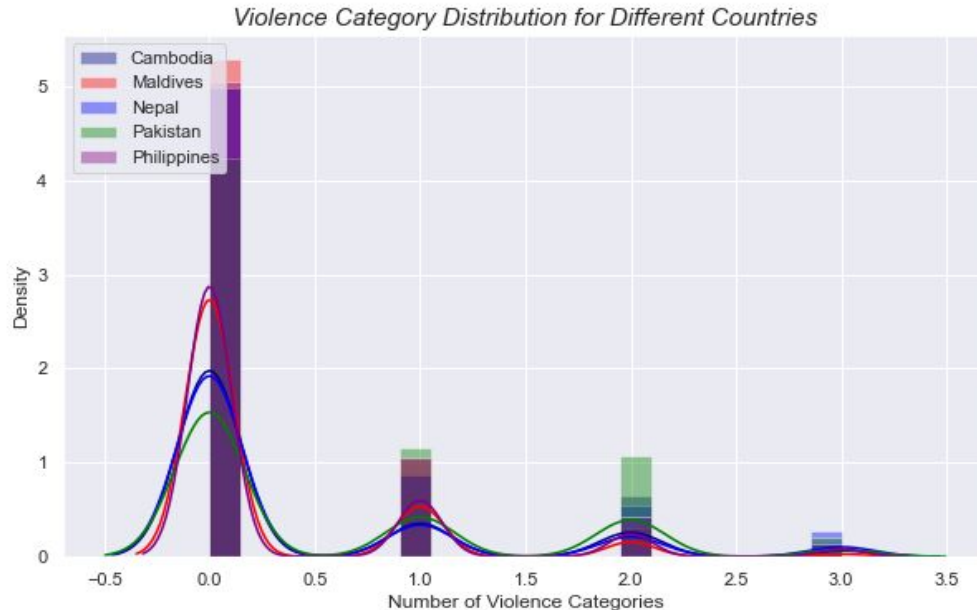


Data Summary: Country Comparison



- If have: about 1/3
- Comparison: Country level
- Pakistan: most severe!

Data Summary: Country Comparison



- Severity: how many kinds of domestic violence?
- Comparison: Country level
- Trend: less people get all kinds of domestic violence
- Pakistan: most severe!



Data Summary: Statistics

Mean Statistics for 5 Countries

country	age	education	wealth index code	if own house	if own land	if employment current	num household
Cambodia	32.98080219	5.084333219	2.161552965	0.761312993	0.666780939	0.725917038	5.468803565
Maldives	33.55747331	8.666370107	1.411032028	0.199644128	1	0.407829181	6.867437722
Nepal	31.71516559	4.341175283	1.901252019	0.087035541	0.125403877	0.606926494	5.459612278
Pakistan	32.11819059	4.462970625	2.005240656	0.026685974	0.019031858	0.129499379	8.45993656
Philippines	34.52010359	10.29744254	1.654969246	0.527290385	0.190935578	0.506701198	5.409906118
country	num child	if emo vio	if phy vio	if phy vio severe	if sex vio	if vio	num vio
Cambodia	2.594446349	0.215017065	0.14179336	0.058641018	0.040645361	0.253490537	0.397455787
Maldives	2.265124555	0.161384217	0.096701164	0.025549806	0.010349288	0.207956016	0.26843467
Nepal	2.532915994	0.123247033	0.216019417	0.097087379	0.074703344	0.252427184	0.413969795
Pakistan	3.353882223	0.302143758	0.242745395	0.064833502	0.043138244	0.365219586	0.587834427
Philippines	2.829135643	0.205830671	0.10399361	0.040015974	0.038738019	0.242412141	0.3485623



Data Summary: Statistics

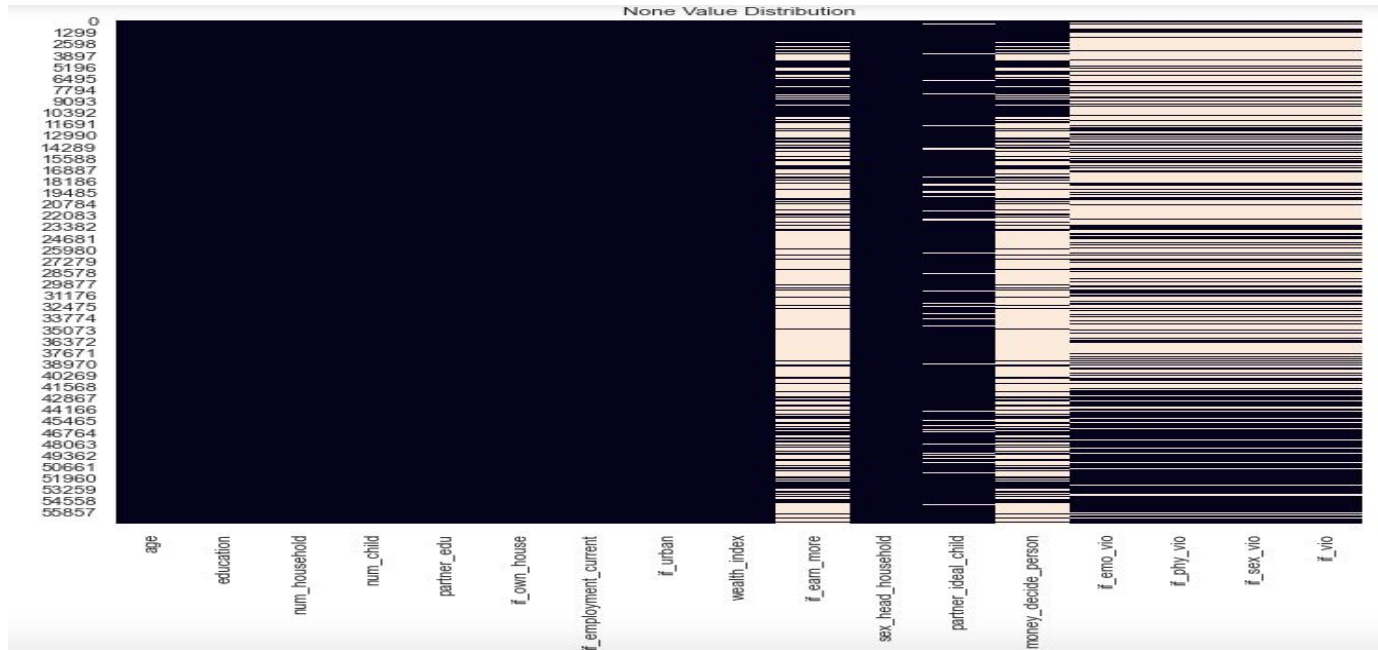
Median Statistics for 5 Countries

country	age	education	wealth_index_code	if_own_house	if_own_land	if_employment_current	num_household
Cambodia	32	5	2	1	1	1	5
Maldives	33	10	1	0	1	0	6
Nepal	31	4	2	0	0	1	5
Pakistan	32	0	2	0	0	0	7
Philippines	35	11	1	1	0	1	5
country	num_child	if_emo_vio	if_phy_vio	if_phy_vio_severe	if_sex_vio	if_vio	num_vio
Cambodia	2	0	0	0	0	0	0
Maldives	2	0	0	0	0	0	0
Nepal	2	0	0	0	0	0	0
Pakistan	3	0	0	0	0	0	0
Philippines	2	0	0	0	0	0	0

Data Summary: Statistics

	age	education	num_household	num_child	if_own_house	if_employment_current	if_emo_vio	if_phy_vio	if_sex_vio	if_vio
age	1.000000	-0.105109	-0.104255	0.559626	0.241121	0.158178	0.012396	0.002700	0.005246	0.008616
education	-0.105109	1.000000	-0.118380	-0.330872	0.075785	0.068526	-0.089336	-0.189951	-0.073587	-0.135853
num_household	-0.104255	-0.118380	1.000000	0.200378	-0.230284	-0.197053	0.051479	0.062059	0.004854	0.058736
num_child	0.559626	-0.330872	0.200378	1.000000	0.092283	-0.010939	0.100992	0.119256	0.053896	0.116031
if_own_house	0.241121	0.075785	-0.230284	0.092283	1.000000	0.251134	0.011014	-0.050794	0.001354	-0.013199
if_employment_current	0.158178	0.068526	-0.197053	-0.010939	0.251134	1.000000	-0.025153	-0.012239	0.024524	-0.022790
if_emo_vio	0.012396	-0.089336	0.051479	0.100992	0.011014	-0.025153	1.000000	0.462482	0.272768	0.856860
if_phy_vio	0.002700	-0.189951	0.062059	0.119256	-0.050794	-0.012239	0.462482	1.000000	0.333684	0.693199
if_sex_vio	0.005246	-0.073587	0.004854	0.053896	0.001354	0.024524	0.272768	0.333684	1.000000	0.350672
if_vio	0.008616	-0.135853	0.058736	0.116031	-0.013199	-0.022790	0.856860	0.693199	0.350672	1.000000

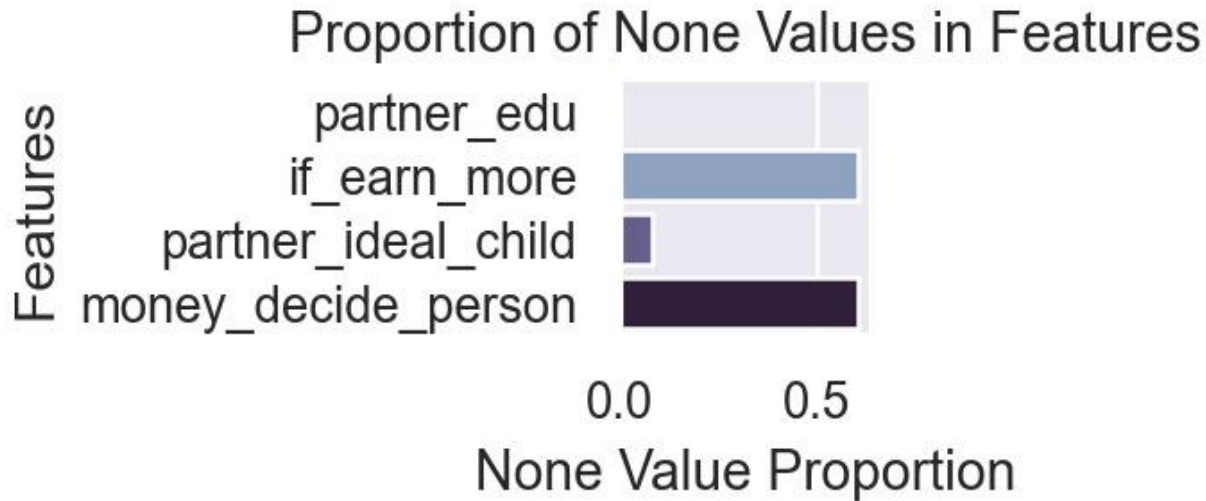
Data Summary : Missing Values



- None Values:

Are they missing randomly?

Data Summary: Missing Values



- Missing Proportions:
4 targets missing half
2 features missing half
(categorical features)
- Dataset:
Original: (57139, 33)
Final: (26505, 33)



Machine Learning Approach

- We developed **binary classification algorithms with individual-level variables and couple-level variables to predict whether a specific ever married or cohabiting woman is prone to each category of IPV in Pakistan, Cambodia, Philippines, Maldives and Nepal.**
- Using ten-fold cross-validation, we trained Balanced Random Forest, Weighted Random Forest, Decision Tree, Random Forest, Logistic Regression, Linear Support Vector Machines and Gaussian Naive Bayes models with Synthetic Minority Over-sampling Technique
- Interpretability, Dealing with imbalanced data
- Feature variables:
 - At the individual level:
 - Categorical: if_urban, wealth_index, if_own_house, if_employment_current
 - Continuous: age, education
 - At the couple level:
 - Categorical: if_earn_more, sex_head_household, partner_ideal_child, money_decide_person
 - Continuous: num_household, num_child
- Target variables to predict for each woman:
 - Binary: if_emo_vio, if_phy_vio, if_sex_vio and if_vio (1 positive outcomes, 0 negative outcomes)



Model Evaluation

Metrics:

- Recall: what proportion of actual positive identifications are correctly labeled.
- Precision: what the proportion of positive-labeled identifications are actual positive identifications.
- F1 score: weighted average of Precision and Recall.

We chose to emphasize on F1 score and recall score to minimize false negatives since we need to detect more potential victims of IPV.

Finetune:

Fine tuned parameters using k-fold on train data to get the best model and evaluate metrics by predicting test data.

Best Model:

Balanced Decision Tree & Weighted Random Forest



Model Evaluation

Five Countries Combined	Target: if_vio			
	Recall	Precision	F1	Accuracy
LogisticRegression	0.00	0.00	0.00	0.75
LinearSVC	0.00	0.00	0.00	0.95
GaussianNB	0.00	0.00	0.00	0.75
Decision Tree	0.20	0.32	0.25	0.68
Random Forest	0.10	0.44	0.17	0.73
Balanced Random Forest	0.59	0.34	0.43	0.59
Weighted Random Forest	0.54	0.33	0.41	0.59

For if_vio:

- Balanced Decision Tree:

F1 score: 43.1%

Recall: 58.6%

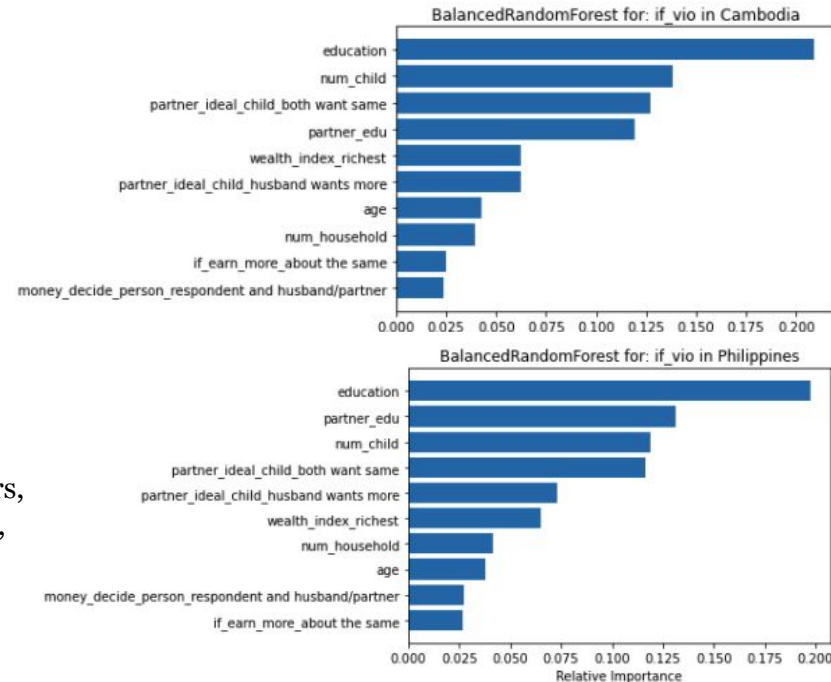
- Weighted Random Forest

F1 score: 41.3%

Recall: 54.4%.

Feature Importance and Policy Recommendation

- **Women's education level** and **her partner's education level**
- **Number of children** and categorical couple-level features - **partner_ideal_child** (whether the respondent believes her partner wants the same number of children or more children than she wants herself)
- Categorical features **wealth_index** are among the top 5 most important features in Cambodia, Nepal and Pakistan IPV predictions.
- With the high feature importance of education and wealth factors when predicting IPV risks, government officials could target provinces that show low education level and economic development. The officials should also take reproductive factors, such as reproductive decision making gap within relationships, into account when developing IPV detection strategies.





Ethics

- DATASET:

Ethics: respect for person and justice

Safety, Confidentiality, Interview and intervention skills.

DHS Programs followed strict interviewing and sampling regulations.

We acknowledged not spreading the individual level data to maintain the confidentiality.

Not guarantee whether there is bias on the societal level or the estimating level.

- ANALYSIS:

Ethics: Justice

Results of our analysis can be misinterpreted and affect resources allocation to certain areas.

May ignored other high-importance features in predicting IPV risks, which might mislead prevention strategies design.

It will put more jeopardy to the physical safety and psychological well-being of both IPV victims or potential victims if research or policy teams do not take adequate precautions.



Limitations and Suggestions for Future Work

- Only based on Pakistan, Cambodia, Philippines, Maldives and Nepal survey data, too population-specific to be extrapolating to the whole of Southeast Asian countries. Causality should not be inferred.
- High proportion of missing values in target variables, IPV is often underreported and our analysis are probably biased.
- Data source: reporting and recall bias, survey methodologies, limits of access and capacity in research.
- The model predictive power could be further improved by introducing more features, including individual acceptance of violence, community and societal factors: gender-inequitable social norms, poverty, weak legal sanctions against IPV within marriage, etc.
- This study also suggests that decomposed analyses should be conducted in other countries as well in order to highlight country specific determinants of global women's risks of IPV rather than employing blanket policies across culturally diverse areas.



Thank you for the watching!