

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

**University of Chicago  
CAPP 30254, Spring 2021, Final Project**

Yangzhou Ou (yangzhou) Xiaoting Sun (sunxt)

June 3, 2021

[GitHub Project Repository](#)

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

## 1 Executive Summary

## 2 Background and Solution Overview

## 3 Data

## 4 Machine Learning Approach

## 5 Evaluation, Results and Policy Recommendation

### 5.1 Model Evaluation

### 5.2 Feature Importance and Policy Recommendation

## 6 Ethics

## 7 Limitations and Suggestions for Future Work

## 8 Reference

## 9 Appendix

### 10.1 Terminology

### 10.2 DHS Survey Questions regarding Intimate Partner Violence

### 10.3 Features, Targets and Dataset Columns Description

### 10.4 Dataset Summary

### 10.5 Model Evaluation Summary

### 10.6 Figures of Feature Importance

# 1 Executive Summary

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 IPV of around 37%. The goal of this study are: To report the prevalence of intimate partner violence (IPV) among ever married or cohabiting women in 5 developing countries in South East Asia: Pakistan, Cambodia, Philippines, Maldives and Nepal; to develop classification algorithms on individual-level variables and couple-level variables to predict whether a specific ever married or cohabiting woman is prone to IPV in these five countries; to identify key characteristics in each country and among countries, including individual-level and couple-level characteristics, associated with experiencing physical, sexual or emotional IPV.

Our main data source is Demographic and Health Surveys (DHS) conducted by the United States Agency of International Development. 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. F1 score and recall score are main evaluation metrics to minimize false negatives since the study needs to detect more potential victims of IPV. The best models of these five countries all achieved F1 score of over 0.4 and recall score nearly 0.6. Women's education level, partner's education level and reproductive intention difference are important features for predicting IPV among all the five countries studied. Government and non-profit officials could target provinces that show low education levels. The officials should also take reproductive factors, such as reproductive decision making gap within relationships, into account when developing IPV detection strategies.

## 2 Background and Solution Overview

Intimate partner violence is one of the most common forms of violence against women and includes physical, sexual, and emotional abuse and controlling behaviours by an intimate partner. The WHO multi-country study on women's health and domestic violence against women showed that IPV is widespread in all countries studied. Among women who had ever been in an intimate partnership: 13–61% reported ever having experienced physical violence by a partner, 6–59% reported sexual violence by a partner at some point in their lives, and 20–75% reported experiencing one emotionally abusive act, or more, from a partner in their lifetime.<sup>1</sup> The highest prevalence of IPV was in Africa, Eastern Mediterranean and South East Asia, where approximately 37% of ever-partnered women reported having experienced physical and/or sexual intimate partner violence at some point in their lives.<sup>2</sup>

IPV affects women's physical and mental health through direct pathways, such as injury, and indirect pathways, such as chronic health problems that arise from prolonged stress. A history of experiencing violence is therefore a risk factor for many diseases and conditions.<sup>3</sup> In light of the health, social, and economic costs of IPV, there have been calls for better data on its prevalence,

---

<sup>1</sup> World Health Organization. (2012) *Understanding and addressing violence against women: Intimate partner violence*. No. WHO/RHR/12.36. World Health Organization.

<sup>2</sup> García-Moreno, C., & Pallitto, C. (2013). Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence. Retrieved from Geneva, Switzerland: [http://apps.who.int/iris/bitstream/handle/10665/85239/9789241564625\\_eng.pdf?sequence=1](http://apps.who.int/iris/bitstream/handle/10665/85239/9789241564625_eng.pdf?sequence=1)

<sup>3</sup> Heise, Lori, and Claudia Garcia-Moreno. (2002) "Violence by intimate partners." *World report on violence and health* 1 : 87-113.

causes, and consequences to improve interventions aimed at addressing of violence against women and girls (VAWG).<sup>4</sup> This global pressure led to the Sustainable Development Goal (SDG) 5.2, which calls for the elimination of all forms of VAWG in both public and private spaces.<sup>5</sup> To resolve IPV issues, it is critical to understand the correlated factors with IPC, including economic and cultural factors.

Our main data source is Demographic and Health Surveys (DHS) conducted by the United States Agency of International Development. In consultation with experts on measurement of domestic violence and survey research, the DHS program developed a standard domestic violence module that was used in surveys of 5 South East Asian countries: Pakistan, Cambodia, Philippines, Maldives and Nepal. Although important policy and programmatic steps have been taken to address violence against women in these five countries over the past decade, there are still gaps to understanding the range of risk and protective factors:

- Pakistan: With 70% of the women in Pakistan having been a victim of IPV, the country is ranked as the sixth most dangerous country in the world for women.<sup>6</sup>
- Cambodia: The report published by Cambodian League for the Promotion and Defence of Human Rights highlighted that no authorities were entrusted with the primary responsibility of enforcing laws to prevent domestic violence. As such, the League for the Promotion and Defence of Human Rights highlighted an urgent need for government programmes to educate the public and law enforcement officials, to redress the perceived inferiority of Cambodian women and recognise domestic violence as a serious problem.<sup>78</sup>
- Philippines: The Republic Act No. 9262 or the “Anti-Violence Against Women and Their Children Act of 2004” is one of the Philippine government’s initiative in addressing the issue on violence against women. Under this Act, violence against women is classified as a public crime and penalizes all forms of abuse and violence within the family and intimate relationships.<sup>9</sup>
- Maldives: The Domestic Violence Prevention Act 2012 (DVPA 2012) experiences small and slow progress, continuing to suffer resource constraints. DV cases are consistently observed where serious systemic failures show victims being unserved and perpetrators released. However, civil society actors inject life to sustain momentum to make the DVPA2012 a source of remedy and redress for violence survivors in a less than favourable environment.<sup>10</sup>
- Nepal: the 2010 National Action Plan Against Gender-Based Violence established an integrated approach to serving survivors and includes a focus on early detection, appropriate referrals and follow-up, and women and children’s legal empowerment.

<sup>4</sup> United Nations Department of Economic and Social Affairs Statistics Division. (2014) Guidelines for producing statistics on violence against women-statistical surveys. New York: United Nations.

<sup>5</sup> United Nations. Transforming our world: the 2030 agenda for sustainable development. New York: United Nations; 2015.

<sup>6</sup> Burki, Navera, Determinants of Intimate Partner Violence: National and Provincial Level Analysis (2020). 3rd International Conference on Gender Research 2020

<sup>7</sup> LICADHO 2006. Violence against women in Cambodia. Phnom Penh. Cambodian League for the Promotion and Defence of Human Rights. <https://www.licadho-cambodia.org/reports/files/105LICADHOREportViolenceWoman2006.pdf>

<sup>8</sup> Cambodian Ministry of Women’s Affairs. Women’s Experience of Domestic Violence and Other Forms of Violence: Secondary data analysis report of CDHS 2015 (2016) <https://asiapacific.unfpa.org/sites/default/files/pub-pdf/Cambodia%20CDHS%20VAW%20study%20secondary%20analysis%20DV%2017%20October%202016%20Eng.pdf>

<sup>9</sup> Authority, Philippine Statistics. "One in four women have ever experienced spousal violence (Preliminary results from the 2017 National Demographic and Health Survey)." (2018).

<sup>10</sup> Society for Health Education, Maldives. Universal Periodic Review of Maldives 36th Session: Apr-May 2020 <https://gh.bmj.com/content/bmjgh/5/1/e002208.full.pdf>

While the strength and enforcement of these laws and initiatives remains underdeveloped.<sup>11</sup>

To address the analysis gap, we analyzed individual-level data within these five countries provided by DHS. The most widely used model for understanding violence is the ecological model, which proposes that violence is a result of individual, relationship/couple, community and societal levels of factors.<sup>12</sup> While much of the work to date on IPV has focused on individual-level characteristics, there is increasing recognition of the need to consider the roles of couple-level characteristics. Use of couples' characteristics also enables examination of the effect of differences within couples' backgrounds and decision making.<sup>13</sup>

The goals of this study are threefold:

- 1) To report the prevalence of intimate partner violence (IPV) among ever married or cohabiting women in 5 developing countries in South East Asia: Pakistan, Cambodia, Philippines, Maldives and Nepal.
- 2) To develop classification algorithms on individual-level variables and couple-level variables to predict whether a specific ever married or cohabiting woman is prone to IPV in these five countries.
- 3) To identify key features in each country and among countries, including individual-level and couple-level characteristics, associated with experiencing physical, sexual or emotional IPV (Definition in Appendix 10.1, 10.2).

We hope that this work could serve to inform policy makers of these five countries who are designing and implementing measures to prevent IPV against women and international and non-profit organizations officers who design intervention strategies and offer assistance to IPV survivors specific to these countries. The country-specific key characteristics identified in our work could serve as a guide and stepstone of an integrated approach to detecting and serving IPV survivors and early detection and prevention intervention.

### 3 Data

Our main data source is Demographic and Health Surveys (DHS) conducted by the United States Agency of International Development. The survey datasets contain individual-level information for women of reproductive age (15-49), enabling us to make classification for individual-level data entry. By about 2000, the DHS program had developed a standard module and methodology for the collection of data on domestic violence. The module helped to increase the validity and comparability of violence data and to ensure that data collection procedures followed the ethical guidelines for the collection of such sensitive information. The standard domestic violence module developed by DHS was used in all of the countries examined in this report: Cambodia(2014), Maldives(2016), Nepal(2016), Pakistan(2017), Philippines(2017) .

---

<sup>11</sup> Office of the Prime Minister and Council of Ministers Government of Nepal.(2012) A study on gender-based violence conducted in selected rural districts of Nepal. Kathmandu, Nepal: Office of the Prime Minister and Council of Ministers.

<sup>12</sup> World Health Organization. (2012) *Understanding and addressing violence against women: Intimate partner violence*. No. WHO/RHR/12.36. World Health Organization.

<sup>13</sup> Hindin, M. J., Kishor, S., & Ansara, D. L. (2008). *Intimate partner violence among couples in 10 DHS countries: Predictors and health outcomes*. Macro International Incorporated.

(See: the Demographic and Health Surveys Standard Recode Manual for DHS-7<sup>14</sup>) We filtered data using the variable `if_union` to only include women who are currently or formerly married (or live with a partner).

We developed classification algorithms on individual-level variables and couple-level variables to predict whether a specific ever married or cohabiting woman is prone to IPV in these five countries. The prediction target is individual-level binary variables of whether a specific woman has experienced IPV: emotional violence, sexual violence, physical violence and any of the three categories. (See Appendix 10.2, 10.3)

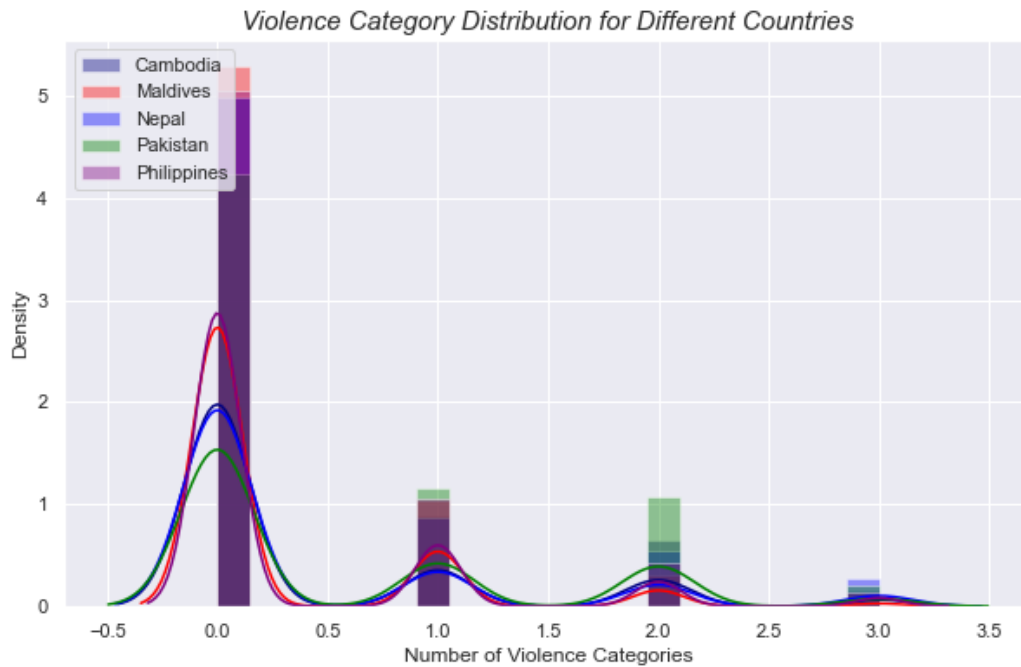
Target	Description
<code>if_emo_vio</code>	Ever experienced any emotional violence
<code>if_phy_vio</code>	Ever experienced any physical violence
<code>if_sex_vio</code>	Ever experienced any sexual violence
<code>if_vio</code>	Ever experienced any of emotional violence, physical violence or sexual violence

\*Figure1: Target Interpretation

Regarding missing data, some features (e.g. `partner_edu`, `partner_ideal_child`) have a relatively small proportion of missing data, so we treated these as Missing Completely at Random: for continuous variables, we impute them with median values; for categorical data, we mark them as 'Missing' and then one-hot encode them. The target variables: `if_emo_vio`, `if_phy_vio`, `if_sex_vio` and `if_vio` all have about 53.6% of missing values. We treated them as Missing at Random and dropped the records with missing target values. However, our data showed that only 25.9% of ever-partnered women have endured any of the three categories of violence (`if_vio` = 1), which is lower than data our literature review reported that 37% of ever-partnered women having experienced physical and/or sexual intimate partner violence at some point in their lives. It is possible that women who did not answer IPV are more likely to suffer from IPV since they might be afraid of further violence or the presence of their partners when conducting the survey. Due to this, the results of our analysis are probably biased.

We counted the categories of violence for each observation (`num_vio`=3 if respondent reported suffering emotional violence, physical violence and sexual violence) to represent the severeness of domestic violence. Plotting the distribution of the severeness, we found that there is a trend that less people get all kinds of domestic violence. However, Pakistan has much higher distribution both on `num_vio` = 1 & 2. This means that female respondents in this country usually faced more than one kind of violence.

<sup>14</sup> USAID. Demographic and Health Surveys Standard Recode Manual for DHS-7. From [https://dhsprogram.com/pubs/pdf/DHSG4/Recode7\\_DHS\\_10Sep2018\\_DHSG4.pdf](https://dhsprogram.com/pubs/pdf/DHSG4/Recode7_DHS_10Sep2018_DHSG4.pdf)



\*Picture1: Violence Category Distribution for Different Countries

## 4 Machine Learning Approach

Given the prevalence of IPV, policy makers of these five countries who are designing and implementing measures to prevent IPV against women and international and non-profit organizations officers who offer assistance to IPV survivors are urged to detect potential IPV victims to provide interventions. 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.**

We also reference similar precedent algorithms to predict likelihood of IPV at individual and community level.<sup>15</sup> Education and wealth are two major individual characteristics associated with IPV: Women's education was significantly associated, after multivariate adjustment, with lifetime spousal violence in Cambodia, Egypt, and India.<sup>16</sup> Higher household economic status

<sup>15</sup> Garcia-Moreno C et al. WHO multi-country study on women's health and domestic violence against women: initial results on prevalence, health outcomes and women's responses. Geneva: World Health Organization, 2005. WHO/LSHTM. Preventing intimate partner and sexual violence against women: taking action and generating evidence. Geneva/London, World Health Organization/ London School of Hygiene and Tropical Medicine, 2010. Hindin, M. J., Kishor, S., & Ansara, D. L. (2008). *Intimate partner violence among couples in 10 DHS countries: Predictors and health outcomes*. Macro International Incorporated.

<sup>16</sup> Kishor, S., & Johnson, K. (2004). *Profiling domestic violence: A multi-country study*. MEASURE DHS+, ORC Macro.

was associated with less physical violence in marriage in Vietnam<sup>17</sup> and Cambodia<sup>18</sup>. Couple characteristics like husband education level need to be considered: Women were also at higher risk of IPV if their husbands had less than college education.<sup>19</sup>

The features and targets for us building binary classification models are as follows: (See Appendix 10.3 for detailed description)

- 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)

Interpretability will be important for our model to enable the audience to develop intervention strategies, so we use Decision Tree and Random Forest. Logistic Regression, Linear Support Vector Machines, Gaussian Naive Bayes are also good fit to binary classification purposes.

Our dataset is imbalanced (if\_vio=1: 25.95%) and the positive cases we care about are the minority classes. To alleviate the problem, we also used Balanced Random Forest, Weighted Random Forest and applied the Synthetic Minority Over-sampling Technique (SMOTE) for imbalance classification to the other models. Since the Random Forest classifier tends to be biased towards the majority class, Balanced Random Forest randomly under-samples each bootstrap sample to balance the data, and Weighted Random Forest places a heavier penalty on misclassifying the minority class.<sup>20</sup> SMOTE is to oversample the minority class and involve duplicating examples in the minority class, although these examples don't add any new information to the model. Instead, new examples can be synthesized from the existing examples.<sup>21</sup> By implementing LogisticRegression, GaussianNB and LinearSVC stepping on SMOTE, both training and testing scores perform better than models without SMOTE.

Since the size of the dataset is limited, we used 10 fold cross-validation at the training stage. We chose to emphasize on F1 score and recall score to minimize false negatives since we need to detect more potential victims of IPV. Thus we evaluated and found the best models based on the highest average test f1 score.

---

<sup>17</sup> Luke, N., S.R. Schuler, B.T.T. Mai, P. Vu Thien, and T.H. Minh. 2007. Exploring couple attributes and attitudes and marital violence in Vietnam. *Violence Against Woman* 13(1):5-27

<sup>18</sup> Yount, K. and J. Carrera. 2006. Domestic violence against married women in Cambodia. *Social Forces* 85(1):355-87

<sup>19</sup> Ackerson, L.K. and S.V. Subramanian. 2008. Domestic violence and chronic malnutrition among women and children in India. *American Journal of Epidemiology* 167(10):1188-96

<sup>20</sup> Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, 110(1-12), 24.

<sup>21</sup> SMOTE for Imbalanced Classification with Python: <https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/>



## 5 Evaluation, Results and Policy Recommendation

### 5.1 Model Evaluation

Our study seeks to correctly identify as many women at risk of intimate partner violence as possible with the purpose of allocating resources towards those individuals and develop early-detection and intervention strategies.

- The recall score tells us what proportion of actual positive identifications are correctly labeled.
- The precision score tells us what the proportion of positive-labeled identifications are actual positive identifications.
- The F1 score is the 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. Since we are dealing with skewed datasets, where the percentages of women with IPV experience is the minority class, accuracy is not the preferred performance measure.

For 5 kinds of classification, we fine tuned our parameters on train data to get the best model and evaluate metrics by predicting test data. The f1, recall and precision are almost zero for models such as LogisticRegression, GaussianNB, LinearSVC, though they have a range of 0.1-0.4 on the k-fold average train scores. This might be caused by the low probability of labels predicted as violence in the test data. If we tuned the threshold of the probability, only when the threshold is less than 0.1 can the model predict positive labels. Though this manual change brings better metric scores, it's not proper to testify that these models are good at predicting in this context.

In contrast, the Balanced Decision Tree and Weighted Random Forest perform much better on predicting. For `if_vio`, Balanced Decision Tree has a f1 score of 43.1% and a recall of 58.6%; Weighted Random Forest has a f1 score of 41.3% and a recall of 54.4%.

With f1 score and recall as the chosen evaluation metrics, the best performing models were the Balanced Random Forest, and Weighted Random Forest the next. This is the same for predicting other targets (shown as pictures in Appendix 10.5). We can see that the precision scores for both models are around 34%, showing the tradeoff between the precision and recall.

Five Countries Combined	Target: <code>if_vio</code>			
	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

## 5.2 Feature Importance and Policy Recommendation

Appendix 10.6 shows feature importance for the good-performing Balanced Random Forest, Weighted Random Forest, Random Forest, Decision Tree models. Within each country, the most important features for predicting emotional violence, physical violence, sexual violence and any of intimate partner violence are similar. While among Pakistan, Cambodia, Philippines, Maldives and Nepal, the feature importance varies a bit.

Women's education level and her partner's education level are the among the top important features for predicting emotional violence, physical violence and sexual violence among all five countries we studied. Women with fewer years of education have a higher risk of lifetime experience of IPV. Women's education facilitates empowerment by expanding their social networks, boosting self-confidence, and improving abilities to use information and resources available in society to become less dependent on their partners. Women are also at higher risk of IPV if their husbands have less education.

Number of children and categorical couple-level features - partner\_ideal\_child is predictive for three categories of IPV risks in all five countries we studied as well, showing that reproductive issues including the number of children, whether the respondent believes her partner wants the same number of children or more children than she wants herself have high feature importance. The couple-level reproductive factors have been underestimated in IPV research and intervention strategies. The discordant answers within relationships regarding reproductive decision making and intentions are also associated with IPV, which may give women's partner excuses of IPV conducts.

Categorical features wealth\_index - Wealth index quintile measuring household's cumulative living standard are among the top 5 most important features in Cambodia, Nepal and Pakistan IPV predictions. Greater wealth and social support are protective against violence.

As intimate partner violence wide-spreads within Pakistan, Cambodia, Philippines, Maldives and Nepal, timely and efficient responses of the government to allocate resources to detect and intervene is essential for improving the situation. Our predictive model could serve as a reference for policy makers of these five countries who are designing and implementing measures to prevent IPV against women and international and non-profit organizations officers who design intervention strategies and offer assistance to IPV survivors specific to these countries. 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.

## 6 Ethics

Various ethical issues may arise from our research on intimate partner violence against women. The nature of the topic and the data collection emphasized importance in issues of safety, confidentiality, interview and intervention skills. The data we used from the DHS Programs questionnaires were obtained by following strict interviewing and sampling regulations. From their manuals, we can tell that they obeyed the ethics of respect for person and justice. We applied for the use of the data and acknowledged not spreading the individual level data to maintain the confidentiality. However, we can not guarantee whether there is bias on the societal level or the estimating level.

Given the sensitivities of reporting intimate partner violence, the estimated prevalence of IPV are likely to underestimate actual prevalence. Due to lack of testing on datasets representativeness among provinces within Pakistan, Cambodia, Philippines, Maldives and Nepal, we are concerned about the possibility that the results of our analysis can be misinterpreted and affect resources allocation to certain areas. Due to lack of features in our analysis, we are also concerned that our results have ignored or failed to learn other high-importance features in predicting IPV risks, which might mislead prevention strategies design. However, out of the justice ethic, we think that 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.

## 7 Limitations and Suggestions for Future Work

The data was only based on Pakistan, Cambodia, Philippines, Maldives and Nepal survey, suggesting that it may be too population-specific to be extrapolating beyond the range of data to the whole of Southeast Asian countries. We need to notice that causality should not be inferred.

The main limitation in this analysis, as mentioned above, is that we have a high proportion of missing values in target variables and we treat them as Missing Completely at Random. It is possible that women who did not answer IPV are more likely to suffer from IPV since they might be afraid of further violence or the presence of their partners when conducting the survey. As has been demonstrated in other contexts, IPV is often underreported as women may fear retribution from husbands and in-laws or wish to avoid social embarrassment.<sup>22</sup> Due to this, the results of our analysis are probably biased.

The data source itself also has limitations including reporting and recall bias, particularly for age or other retrospective data relying on memory of a past event.<sup>23</sup> The survey methodologies will also affect the accuracy of the DHS datasets given the sensitivity of the topic. Since developing countries have limited capacity in research, survey data were collected 3 years ago or more, which causes difficulties in figuring out the latest trends. Some poverty and remote areas might be underreported due to 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

---

<sup>22</sup> Ghimire, A., & Samuels, F. (2017). Understanding intimate partner violence in Nepal: prevalence, drivers and challenges. *Overseas Development Institute, Institute OD*.

<sup>23</sup> Boerma, J. T., & Sommerfelt, A. E. (1993). Demographic and health surveys (DHS): contributions and limitations. *World health statistics quarterly. Rapport trimestriel de statistiques sanitaires mondiales*, 46(4), 222–226.

country specific determinants of global women's risks of IPV rather than employing blanket policies across culturally diverse areas.

## 8 Reference

- World Health Organization. (2012) Understanding and addressing violence against women: Intimate partner violence. No. WHO/RHR/12.36. World Health Organization.
- García-Moreno, C., & Pallitto, C. (2013). Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence. Retrieved from Geneva, Switzerland: [http://apps.who.int/iris/bitstream/handle/10665/85239/9789241564625\\_eng.pdf?sequence=1](http://apps.who.int/iris/bitstream/handle/10665/85239/9789241564625_eng.pdf?sequence=1)
- Heise, Lori, and Claudia Garcia-Moreno. (2002) "Violence by intimate partners." World report on violence and health 1 : 87-113.
- United Nations Department of Economic and Social Affairs Statistics Division. (2014) Guidelines for producing statistics on violence against women-statistical surveys. New York: United Nations.
- United Nations. Transforming our world: the 2030 agenda for sustainable development. New York: United Nations; 2015.
- Burki, Navera, Determinants of Intimate Partner Violence: National and Provincial Level Analysis (2020). 3rd International Conference on Gender Research 2020
- LICADHO 2006. Violence against women in Cambodia. Phnom Penh. Cambodian League for the Promotion and Defence of Human Rights. <https://www.licadho-cambodia.org/reports/files/105LICADHOREportViolenceWoman2006.pdf>
- Cambodian Ministry of Women's Affairs. Women's Experience of Domestic Violence and Other Forms of Violence: Secondary data analysis report of CDHS 2015 (2016) <https://asiapacific.unfpa.org/sites/default/files/pub-pdf/Cambodia%20CDHS%20VAW%20study%20secondary%20analysis%20DV%2017%20October%202016%20Eng.pdf>
- Authority, Philippine Statistics. "One in four women have ever experienced spousal violence (Preliminary results from the 2017 National Demographic and Health Survey)." (2018).
- Society for Health Education, Maldives. Universal Periodic Review of Maldives 36th Session: Apr-May 2020 <https://gh.bmj.com/content/bmjgh/5/1/e002208.full.pdf>
- Office of the Prime Minister and Council of Ministers Government of Nepal.(2012) A study on gender-based violence conducted in selected rural districts of Nepal. Kathmandu, Nepal: Office of the Prime Minister and Council of Ministers.
- World Health Organization. (2012) Understanding and addressing violence against women: Intimate partner violence. No. WHO/RHR/12.36. World Health Organization.
- Hindin, M. J., Kishor, S., & Ansara, D. L. (2008). Intimate partner violence among couples in 10 DHS countries: Predictors and health outcomes. Macro International Incorporated.
- USAID. Demographic and Health Surveys Standard Recode Manual for DHS-7. From [https://dhsprogram.com/pubs/pdf/DHSG4/Recode7\\_DHS\\_10Sep2018\\_DHSG4.pdf](https://dhsprogram.com/pubs/pdf/DHSG4/Recode7_DHS_10Sep2018_DHSG4.pdf)
- Garcia-Moreno C et al. WHO multi-country study on women's health and domestic violence against women: initial results on prevalence, health outcomes and women's responses. Geneva: World Health Organization, 2005. WHO/LSHTM. Preventing intimate partner and sexual violence against women: taking action and generating evidence. Geneva/London, World Health Organization/ London School of Hygiene and Tropical Medicine, 2010. Hindin, M. J., Kishor, S., & Ansara, D. L. (2008). Intimate partner violence among couples in 10 DHS countries: Predictors and health outcomes. Macro International Incorporated.
- Kishor, S., & Johnson, K. (2004). Profiling domestic violence: A multi-country study. MEASURE DHS+, ORC Macro.
- Luke, N., S.R. Schuler, B.T.T. Mai, P. Vu Thien, and T.H. Minh. 2007. Exploring couple attributes and attitudes and marital violence in Vietnam. Violence Against Woman 13(1):5-27
- Yount, K. and J. Carrera. 2006. Domestic violence against married women in Cambodia. Social Forces 85(1):355-87
- Ackerson, L.K. and S.V. Subramanian. 2008. Domestic violence and chronic malnutrition among women and children in India. American Journal of Epidemiology 167(10):1188-96

- Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. University of California, Berkeley, 110(1-12), 24.
- Ghimire, A., & Samuels, F. (2017). Understanding intimate partner violence in Nepal: prevalence, drivers and challenges. Overseas Development Institute, Institute OD.
- Boerma, J. T., & Sommerfelt, A. E. (1993). Demographic and health surveys (DHS): contributions and limitations. World health statistics quarterly. Rapport trimestriel de statistiques sanitaires mondiales, 46(4), 222–226.
- UN Women. Definition of Domestic Violence. From <http://www.endvawnow.org/en/articles/398-definition-ofdomestic-violence.html>
- Breiding MJ, Basile KC, Smith SG, Black MC, Mahendra RR. Intimate Partner Violence Surveillance: Uniform Definitions and Recommended Data Elements, Version 2.0. Atlanta (GA): National Center for Injury Prevention and Control, Centers for Disease Control and Prevention; 2015.

## 9 Appendix

### 10.1 Terminology

Term	Definition
Domestic violence	Used by adults or adolescents against family member. Includes a pattern of assaultive and coercive behaviours. Most domestic violence is perpetrated by intimate partners. <sup>24</sup>
Intimate partner	Current and former husbands, cohabiting partners, fiancés, dating partners etc., whether or not there is or has been a sexual relationship. Does not include parents, neighbours, sweethearts etc. <sup>25</sup>
Intimate partner violence	Physical, sexual, psychological or economic violence perpetrated by a current or former intimate partner.
Emotional violence	Constant belittling, humiliating, scarring or intimidating through verbal aggression and threats of violence with the intention to instil fear and control the behaviour of the victim. See survey questions in 10.2.
Physical violence	Slapping, shaking, beating with a fist or object, strangulation, burning, kicking, threats with an actual harm resulting from use of a weapon. See survey questions in 10.2.
Sexual violence	Coerced or forced sex or sexual acts through physical force or intimidation. See survey questions in 10.2.

---

<sup>24</sup> UN Women. Definition of Domestic Violence. From <http://www.endvawnow.org/en/articles/398-definition-ofdomestic-violence.html>

<sup>25</sup> Breiding MJ, Basile KC, Smith SG, Black MC, Mahendra RR. Intimate Partner Violence Surveillance: Uniform Definitions and Recommended Data Elements, Version 2.0. Atlanta (GA): National Center for Injury Prevention and Control, Centers for Disease Control and Prevention; 2015.

## 10.2 DHS Survey Questions regarding Intimate Partner Violence

Based on the Demographic and Health Surveys Standard Recode Manual for DHS-7<sup>26</sup>

<b>Violence Category</b>	<b>DHS Survey Questions</b>
<b>Emotional violence</b>	<p>Ever-married women were asked if their current or former (for those who were formerly married) husband or intimate partner ever:</p> <ul style="list-style-type: none"> <li>• Spouse ever humiliated her.</li> <li>• Spouse ever threatened her with harm.</li> <li>• Spouse ever insulted her or made her feel bad.</li> </ul>
<b>Physical violence</b>	<p>Ever-married women were asked if their current or former (for those who were formerly married) husband or intimate partner ever:</p> <ul style="list-style-type: none"> <li>• Spouse ever pushed, shook or threw something.</li> <li>• Spouse ever slapped.</li> <li>• Spouse ever punched with fist or something harmful.</li> <li>• Spouse ever kicked or dragged.</li> <li>• Spouse ever tried to strangle or burn.</li> <li>• Spouse ever threatened with knife/gun or other weapon.</li> <li>• Spouse ever attacked with knife/gun or other weapon.</li> <li>• Spouse ever twisted her arm or pulled her hair.</li> </ul>
<b>Sexual violence</b>	<p>Ever-married women were asked if their current or former (for those who were formerly married) husband or intimate partner ever:</p> <ul style="list-style-type: none"> <li>• Spouse ever physically forced sex when not wanted.</li> <li>• Spouse ever forced other sexual acts when not wanted.</li> </ul>

<sup>26</sup> USAID. Demographic and Health Surveys Standard Recode Manual for DHS-7. From [https://dhsprogram.com/pubs/pdf/DHSG4/Recode7\\_DHS\\_10Sep2018\\_DHSG4.pdf](https://dhsprogram.com/pubs/pdf/DHSG4/Recode7_DHS_10Sep2018_DHSG4.pdf)



## 10.3 Features, Targets and Dataset Columns Description

- 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)

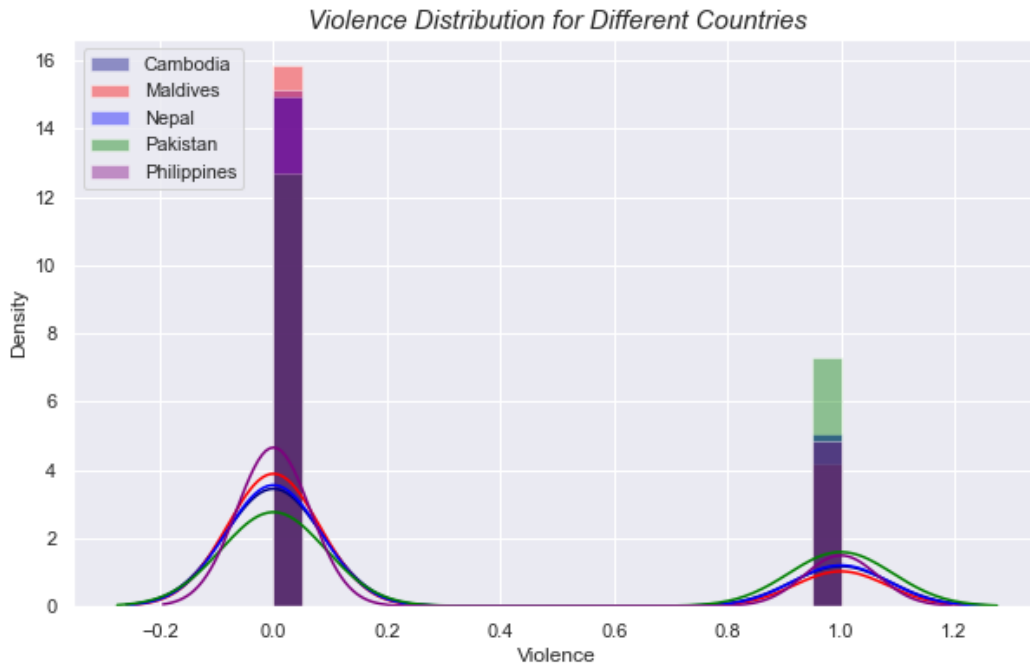
Dataset column	Description
caseid	Case ID
country	Country
year	Survey year
province	Region of residence
if_urban	Urban/rural residence
if_emo_vio	Ever experienced any emotional violence
if_phy_vio	Ever experienced any physical violence
if_sex_vio	Ever experienced any sexual violence
if_vio	Ever experienced any of emotional violence, physical violence or sexual violence
education	Education in single years
if_union	Whether the respondent is currently, formerly or never married (or lived with a partner).

wealth_index	<p>Wealth index quintile</p> <p>The wealth index is a composite measure of a household's cumulative living standard.</p> <p>The wealth index is calculated using easy-to-collect data on a household's ownership of selected assets, such as televisions and bicycles; materials used for housing construction; and types of water access and sanitation facilities.</p> <p>Generated with a statistical procedure known as principal components analysis, the wealth index places individual households on a continuous scale of relative wealth. DHS separates all interviewed households into five wealth quintiles to compare the influence of wealth on various population, health and nutrition indicators. The wealth index is presented in the DHS Final Reports and survey datasets as a background characteristic</p>
wealth_index_code	<p>Based on wealth_index:</p> <p>'poorest': 0,  'poorer': 1,  'middle': 2,  'richer': 3,  'richest': 4</p>
house_ownership	Owns a house alone or jointly
if_own_house	Whether the respondent owns house alone or jointly
if_employment	Whether the respondent worked in the last 12 months.
if_employment_current	Whether the respondent is currently working.
employment_pay_method	Type of earnings for work. The respondent is asked whether they receive cash for their work, they are paid in kind, a combination of the two or not paid.
if_earn_more	Whether the respondent earns more than her partner.
partner_edu	Most recent husband or partner's education in single years
num_household	Total number of household members is the number of usual residents plus the number of visitors who slept in the house the previous night that were listed in the household schedule.
num_child	Total number of children ever born
sex_head_household	Sex of the head of the household

sexual_activity	<p>Recent sexual activity gives the sexual activity of the respondents during the last four weeks coded as follows:</p> <p>0 Never had intercourse</p> <p>1 Active in last 4 weeks</p> <p>2 Not active in last 4 weeks - postpartum abstinence after the birth of the last child</p> <p>3 Not active in last 4 weeks - not postpartum abstinence after the birth of the last child</p> <p>BASE: All women. Respondent who never had sex were coded 0.</p>
ideal_num_child	<p>The ideal number of children that the respondent would have liked to have in her whole life, irrespective of the number she already has.</p>
partner_ideal_child	<p>Whether the respondent believes her partner wants the same number of children, more children or fewer children than she wants herself.</p>
money_decide_person	<p>The person who mainly decides how the money earned by the respondent is used</p>

## 10.4 Dataset Summary

### 10.4.1 Statistics Summary

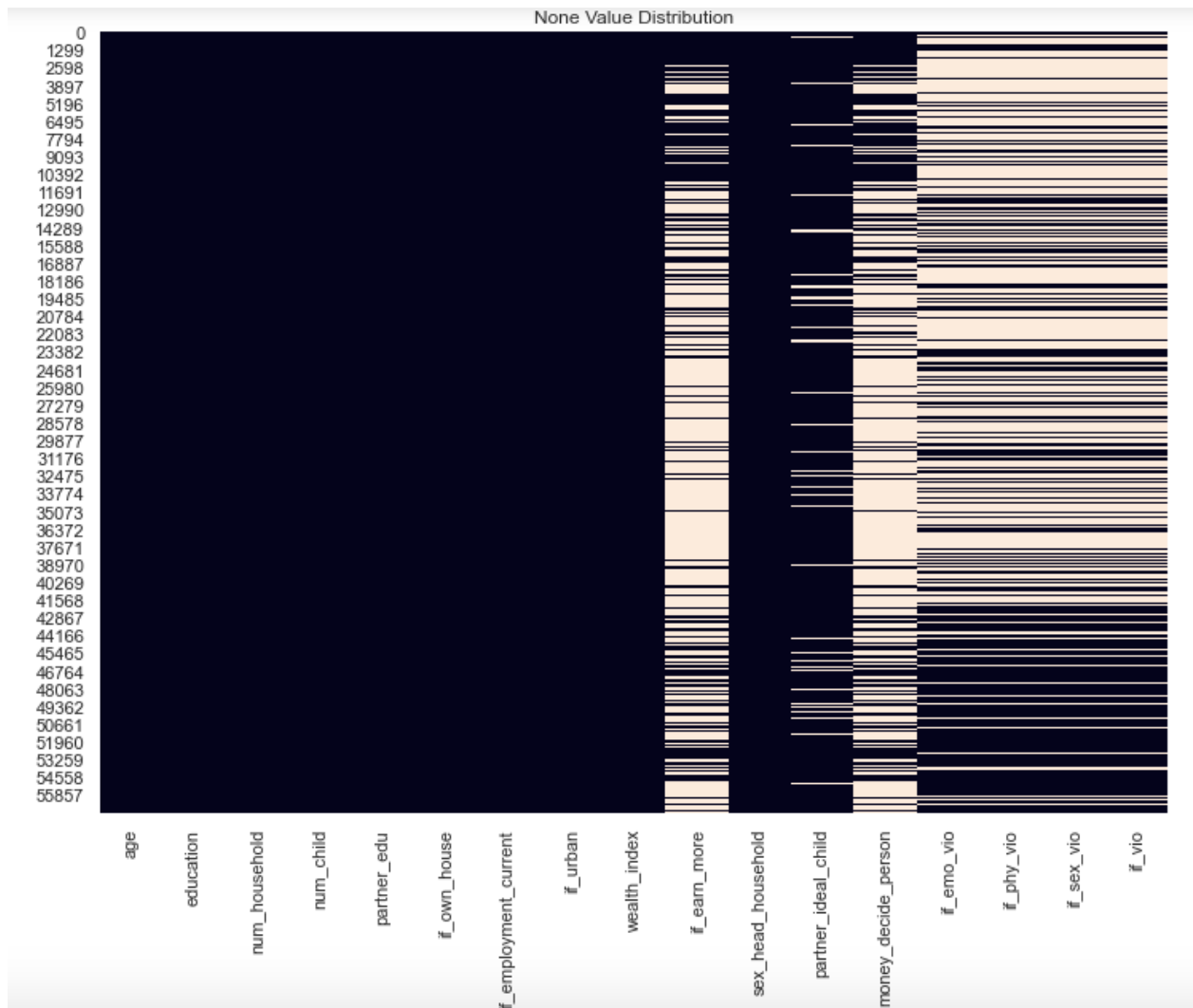


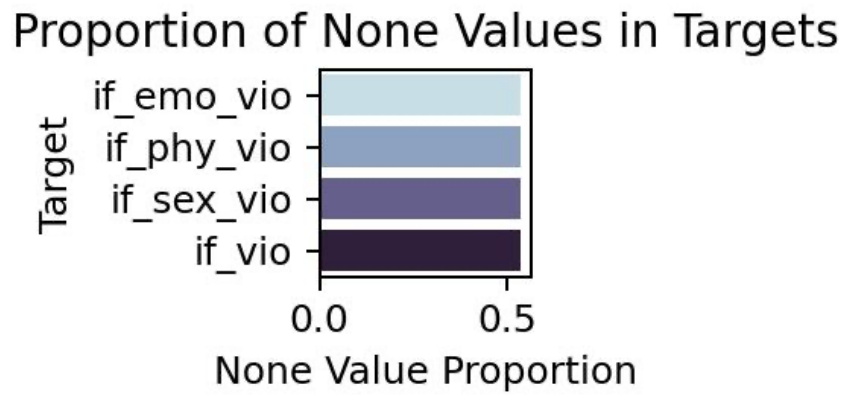
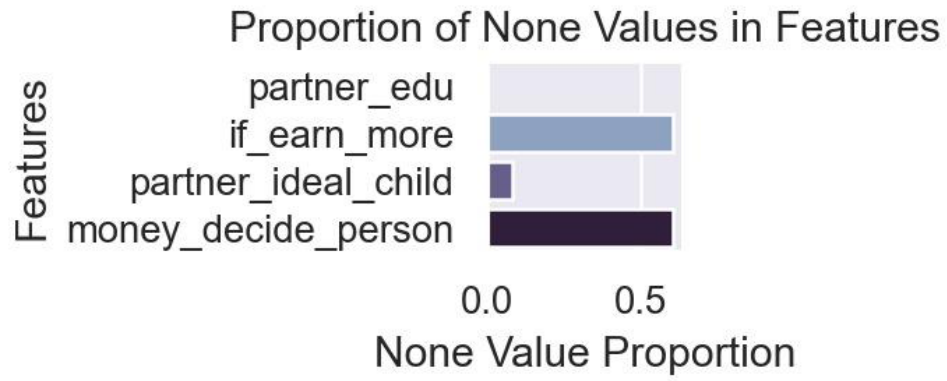
	age	education	num_household	num_child	if_own_house	if_employment_current	if_emo_vio	if_phy_vio	if_sex_vio	if_vio
age	1.000000	-0.105109	-0.104255	0.559626	0.241121	0.158178	0.012396	0.002700	0.005246	0.008616
education	-0.105109	1.000000	-0.118380	-0.330872	0.075785	0.068526	-0.089336	-0.189951	-0.073587	-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

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

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	1	5
Maldives	33	10	1	0	1	0	0	6
Nepal	31	4	2	0	0	1	1	5
Pakistan	32	0	2	0	0	0	0	7
Philippines	35	11	1	1	0	1	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	0
Maldives	2	0	0	0	0	0	0	0
Nepal	2	0	0	0	0	0	0	0
Pakistan	3	0	0	0	0	0	0	0
Philippines	2	0	0	0	0	0	0	0

## 10.4.2 None Values Summary





## 10.5 Model Evaluation Summary

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

Five Countries Combined	Target: if_emo_vio			
	Recall	Precision	F1	Accuracy
LogisticRegression	0.00	0.00	0.00	0.80
LinearSVC	0.00	0.00	0.00	0.78
GaussianNB	0.00	0.00	0.00	0.80
Decision Tree	0.16	0.25	0.19	0.72
Random Forest	0.05	0.41	0.09	0.78
Balanced Random Forest	0.58	0.27	0.37	0.59
Weighted Random Forest	0.55	0.27	0.36	0.59

Five Countries Combined	Target: if_phy_vio			
	Recall	Precision	F1	Accuracy
LogisticRegression	0.00	0.00	0.00	0.86
LinearSVC	0.00	0.00	0.00	0.86
GaussianNB	0.00	0.00	0.00	0.86
Decision Tree	0.12	0.22	0.16	0.81

Random Forest	0.04	0.38	0.08	0.85
Balanced Random Forest	0.62	0.23	0.34	0.65
Weighted Random Forest	0.60	0.23	0.33	0.66

Five Countries Combined	Target: if_sex_vio			
	Recall	Precision	F1	Accuracy
LogisticRegression	0.00	0.00	0.00	0.95
LinearSVC	0.00	0.00	0.00	0.95
GaussianNB	0.00	0.00	0.00	0.95
Decision Tree	0.08	0.08	0.08	0.93
Random Forest	0.00	0.05	0.00	0.96
Balanced Random Forest	0.62	0.06	0.11	0.64
Weighted Random Forest	0.53	0.06	0.11	0.69

Cambodia	Target: if_vio			
	Recall	Precision	F1	Accuracy
Decision Tree	0.20	0.32	0.25	0.68
Random Forest	0.09	0.40	0.15	0.72
Balanced Random Forest	0.58	0.33	0.42	0.58
Weighted Random Forest	0.57	0.33	0.42	0.58

Maldives	Target: if_vio			
	Recall	Precision	F1	Accuracy
Decision Tree	0.20	0.31	0.24	0.67
Random Forest	0.09	0.43	0.15	0.73
Balanced Random Forest	0.57	0.34	0.42	0.59
Weighted Random Forest	0.56	0.33	0.42	0.59

Nepal	Target: if_vio			
-------	----------------	--	--	--



	<b>Recall</b>	<b>Precision</b>	<b>F1</b>	<b>Accuracy</b>
Decision Tree	0.19	0.32	0.24	0.68
Random Forest	0.09	0.45	0.15	0.73
Balanced Random Forest	0.58	0.33	0.42	0.58
Weighted Random Forest	0.57	0.34	0.42	0.59

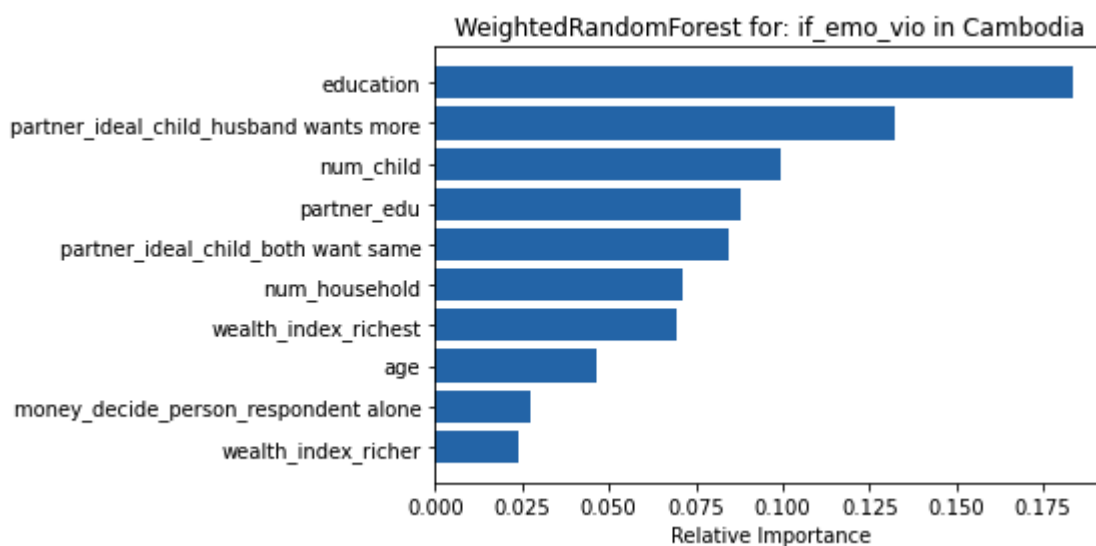
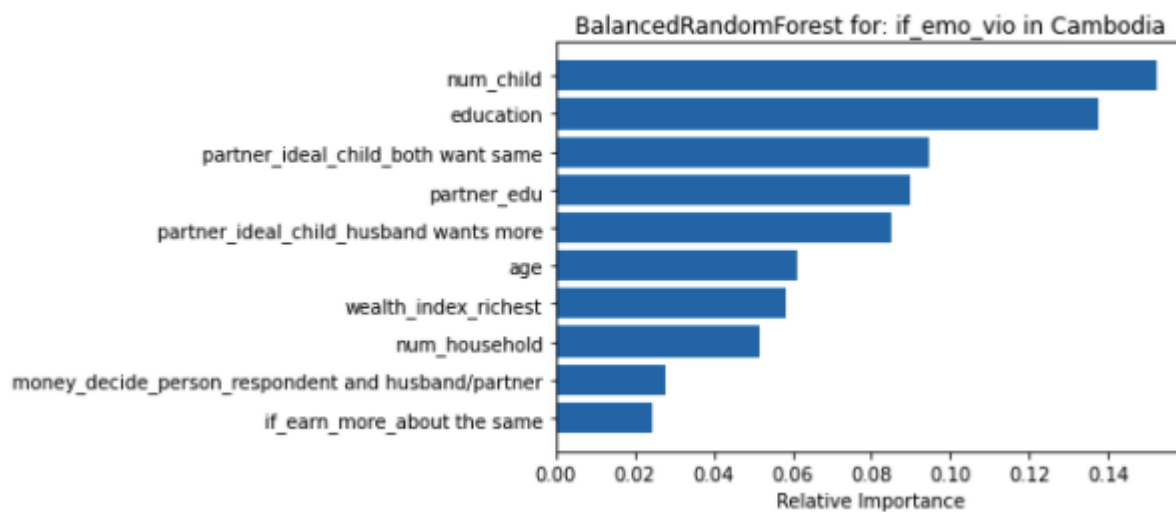
<b>Pakistan</b>	<b>Target: if_vio</b>			
	<b>Recall</b>	<b>Precision</b>	<b>F1</b>	<b>Accuracy</b>
Decision Tree	0.19	0.34	0.24	0.69
Random Forest	0.10	0.43	0.16	0.73
Balanced Random Forest	0.57	0.33	0.42	0.58
Weighted Random Forest	0.56	0.34	0.43	0.60

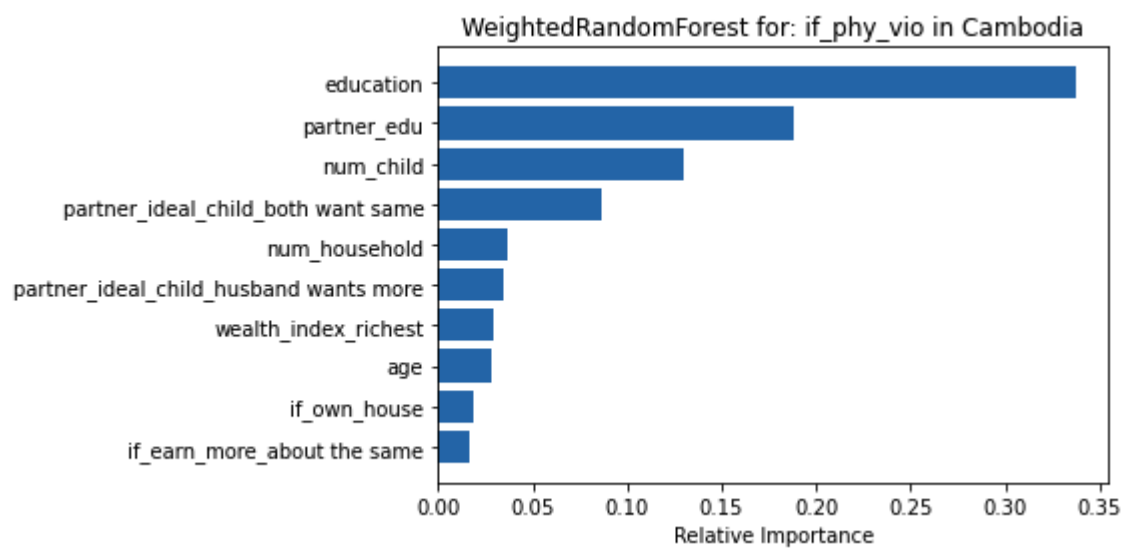
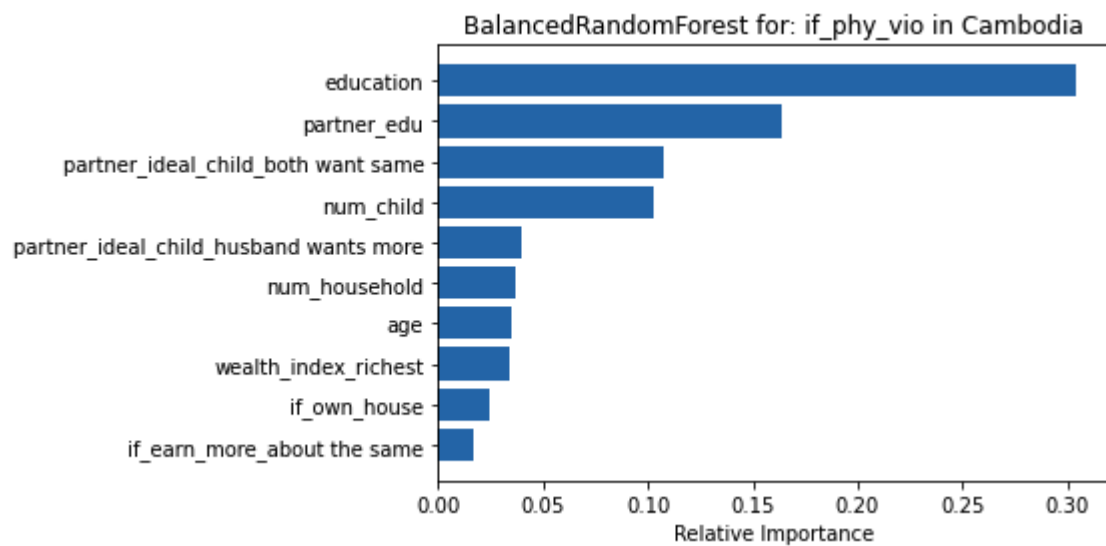
<b>Philippines</b>	<b>Target: if_vio</b>			
	<b>Recall</b>	<b>Precision</b>	<b>F1</b>	<b>Accuracy</b>
Decision Tree	0.20	0.32	0.24	0.68
Random Forest	0.10	0.41	0.16	0.72
Balanced Random Forest	0.58	0.34	0.43	0.59
Weighted Random Forest	0.55	0.33	0.41	0.59

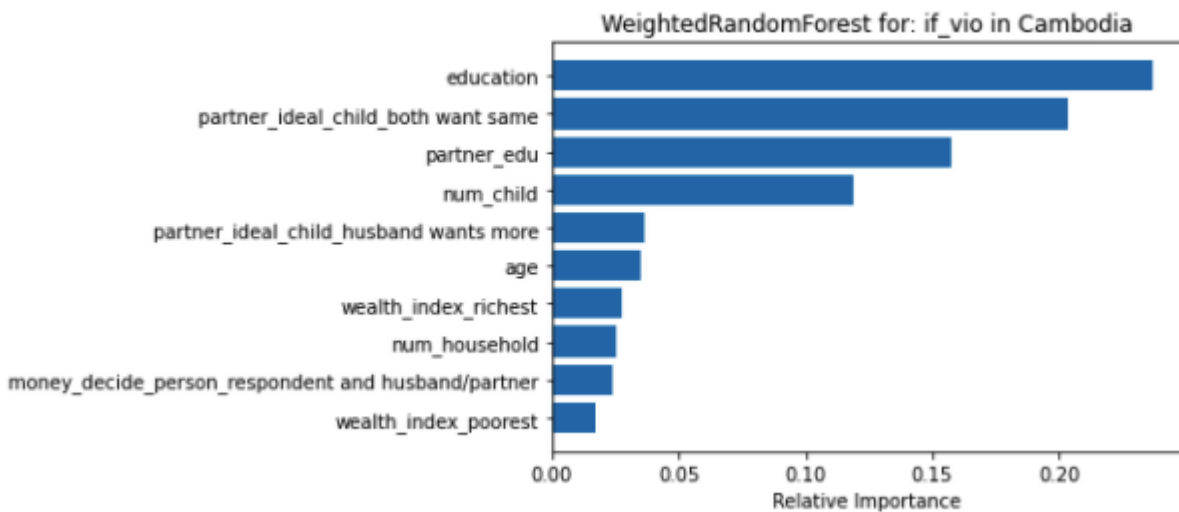
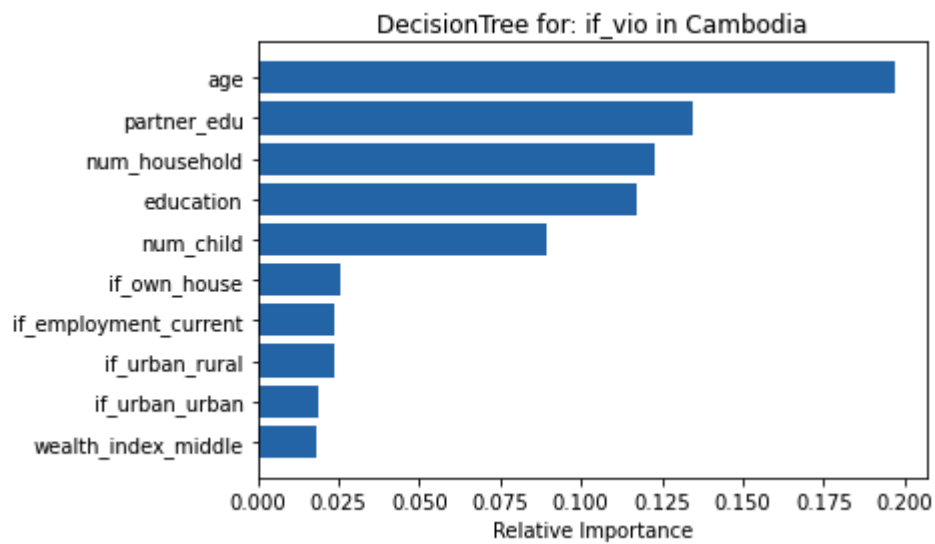
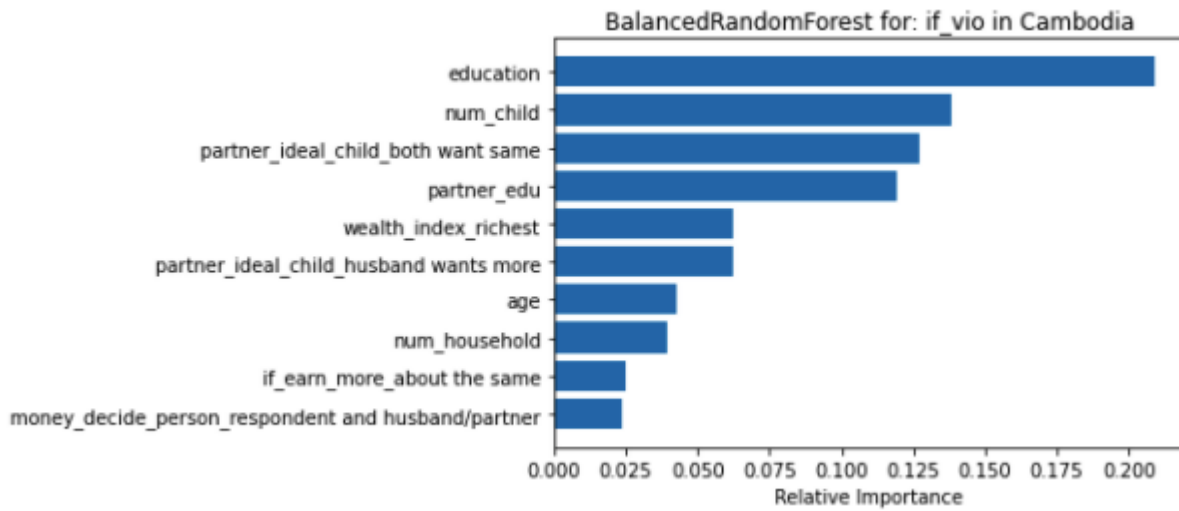
## 10.6 Figures of Feature Importance

Cambodia: Feature importance for IPV risk prediction using Classifier Models

\* Excludes models whose testing f1 scores are below 0.3.

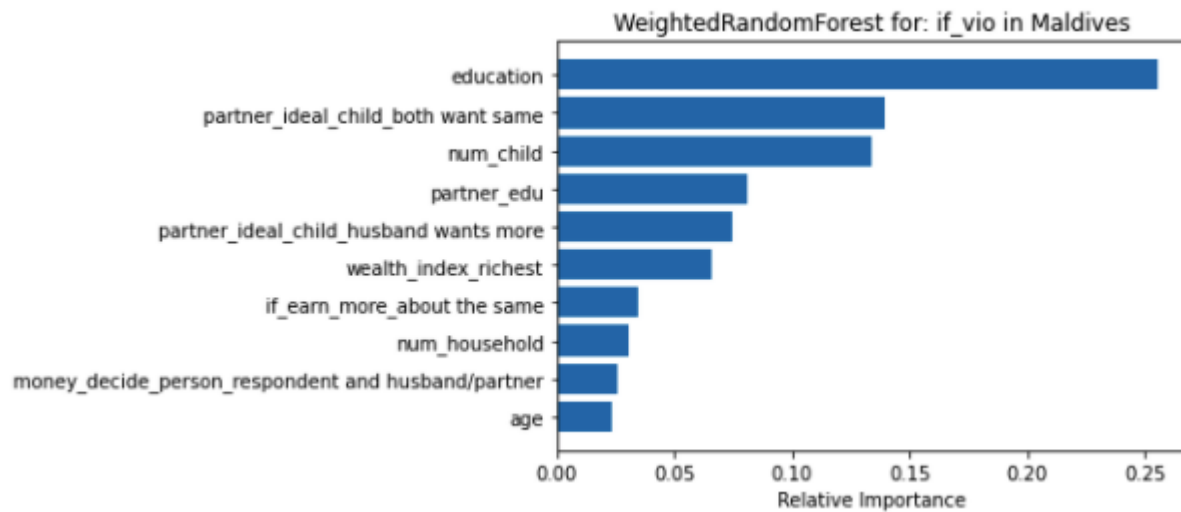
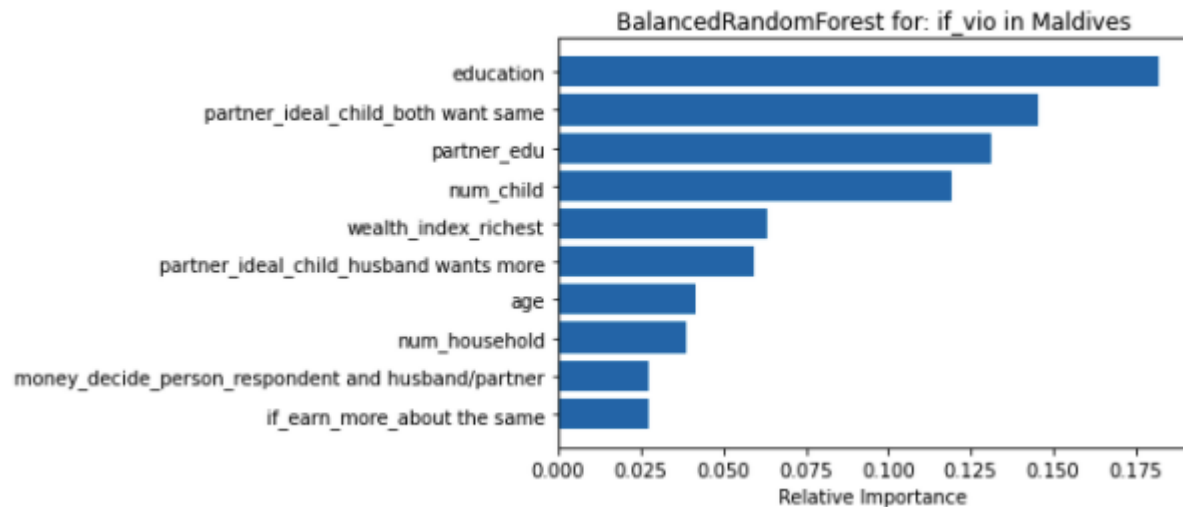






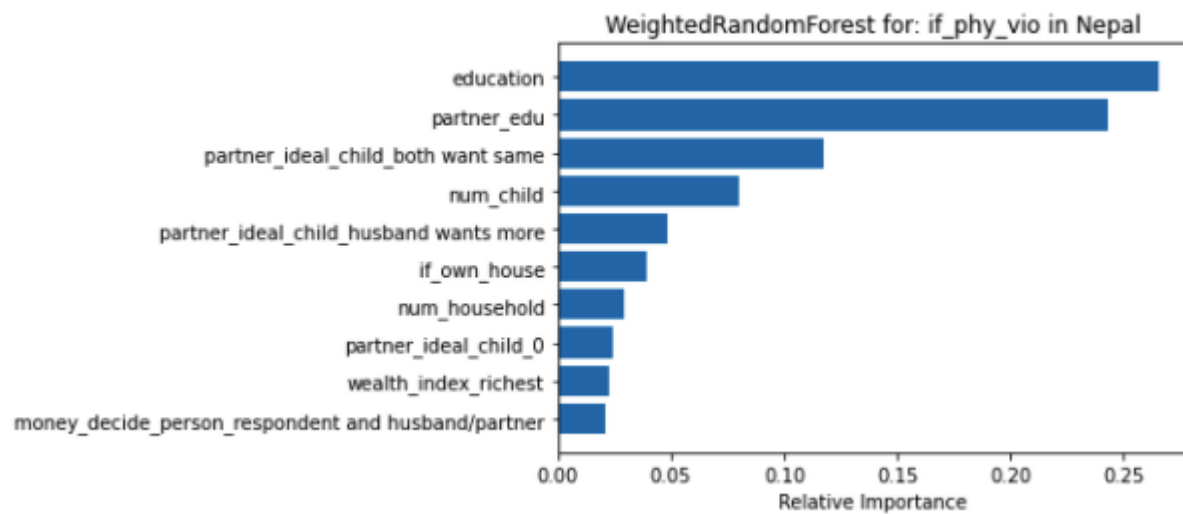
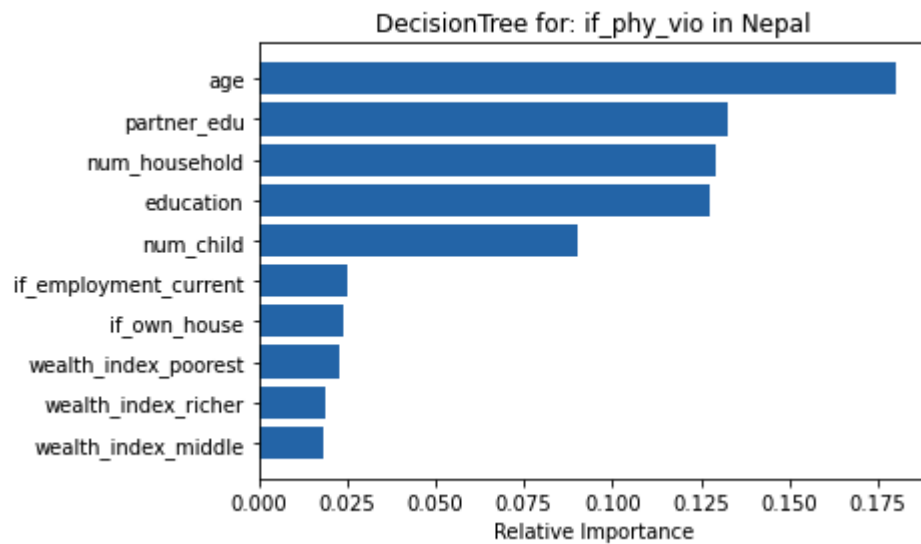
## Maldives: Feature importance for IPV risk prediction using Classifier Models

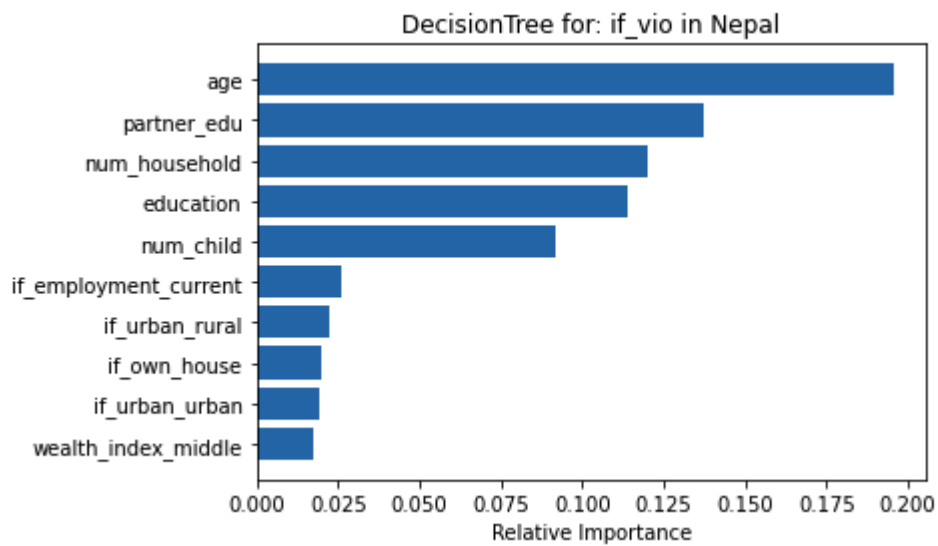
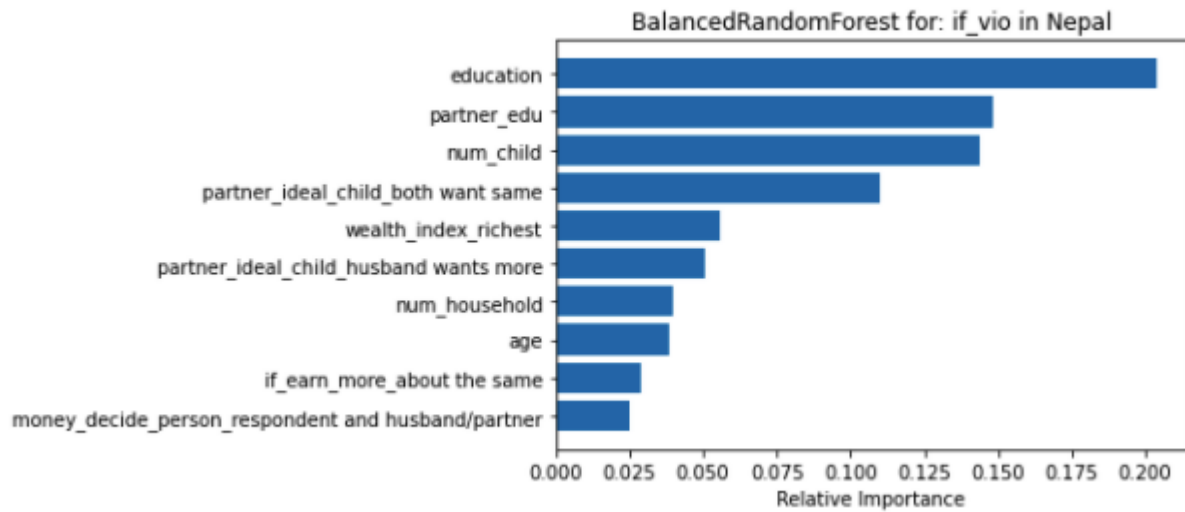
\* Excludes models whose testing f1 scores are below 0.3.

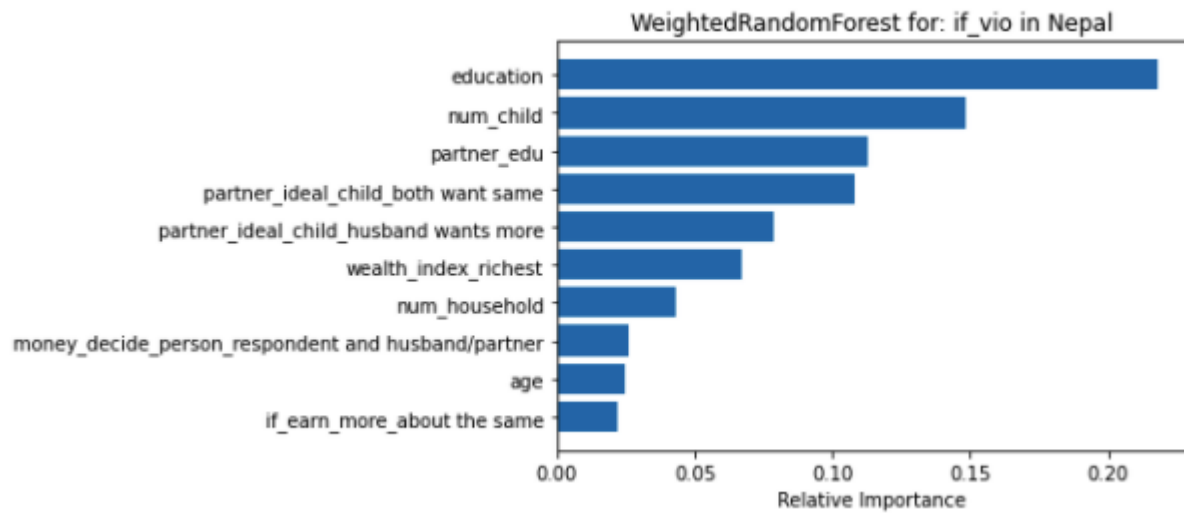


## Nepal: Feature importance for IPV risk prediction using Classifier Models

\* Excludes models whose testing f1 scores are below 0.3.



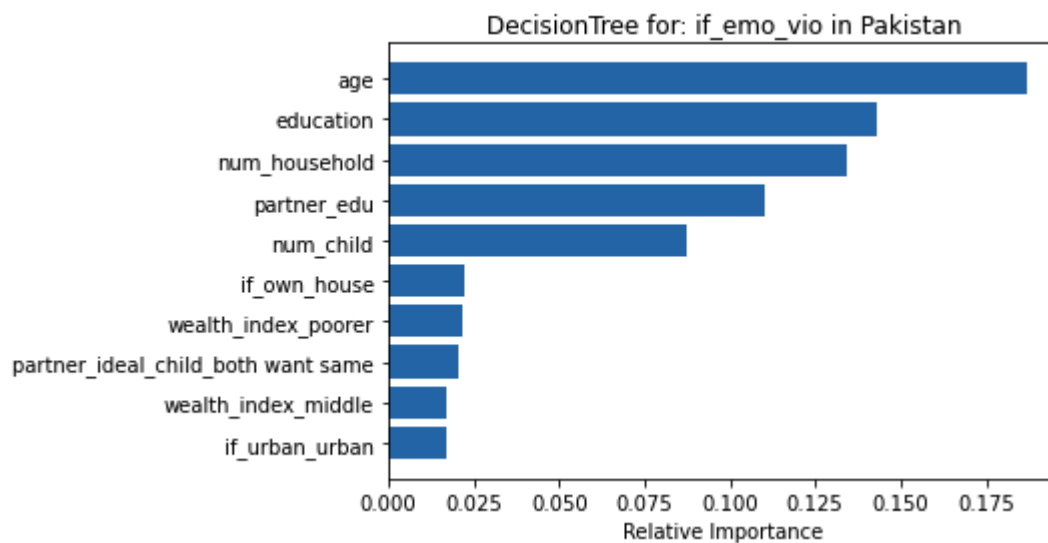
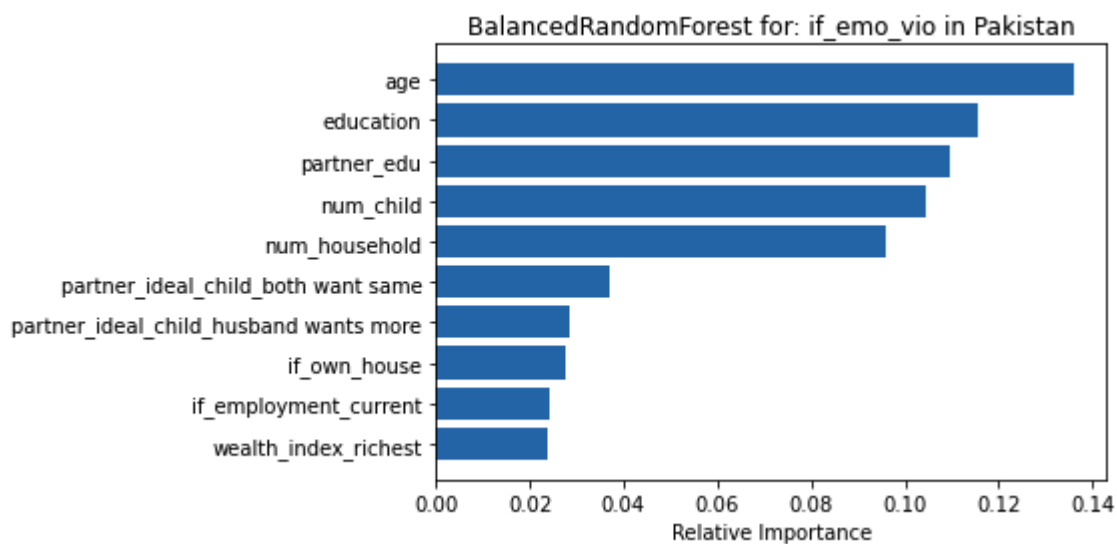


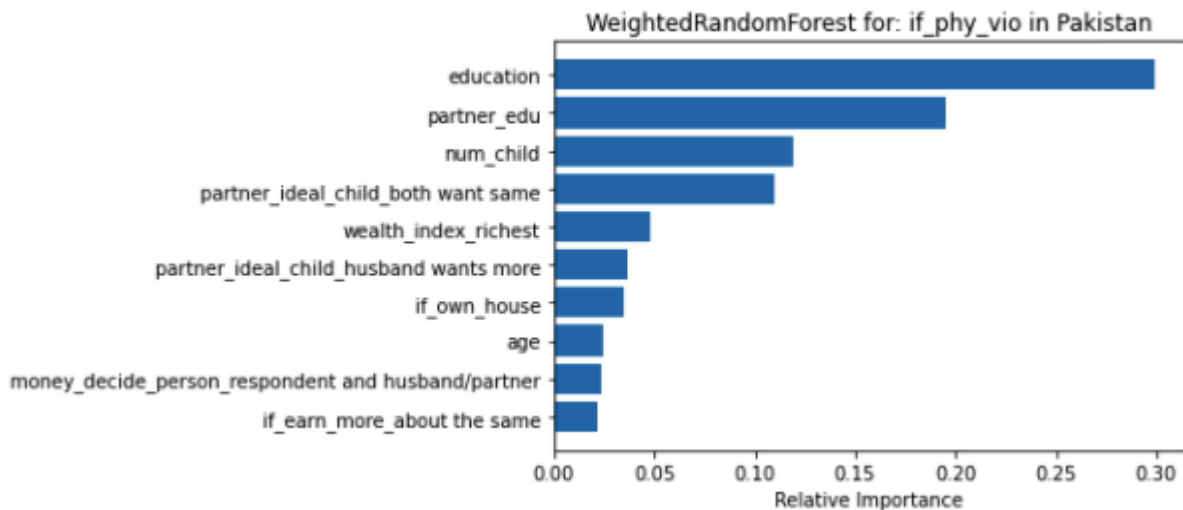
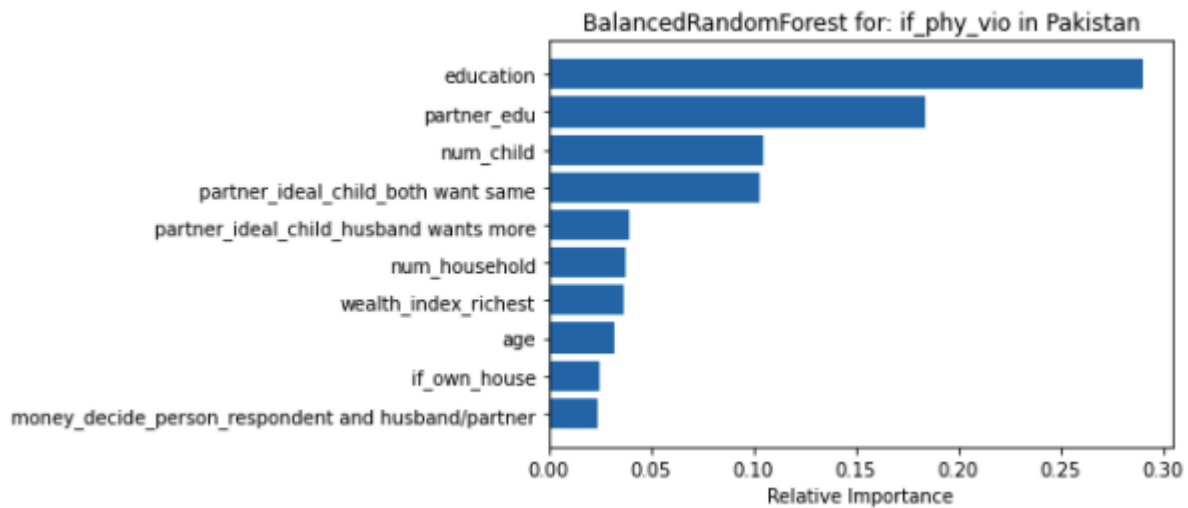
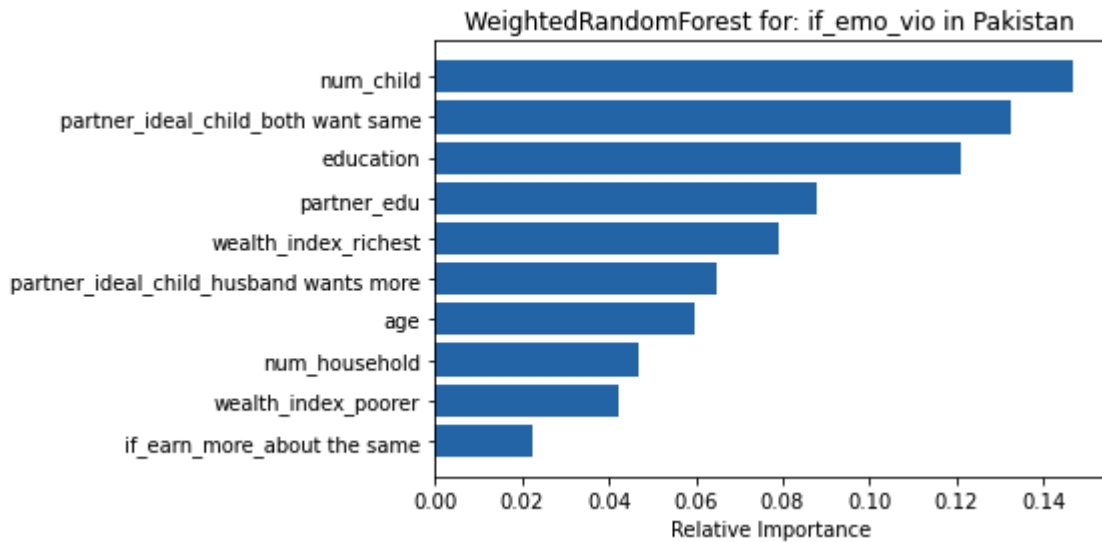


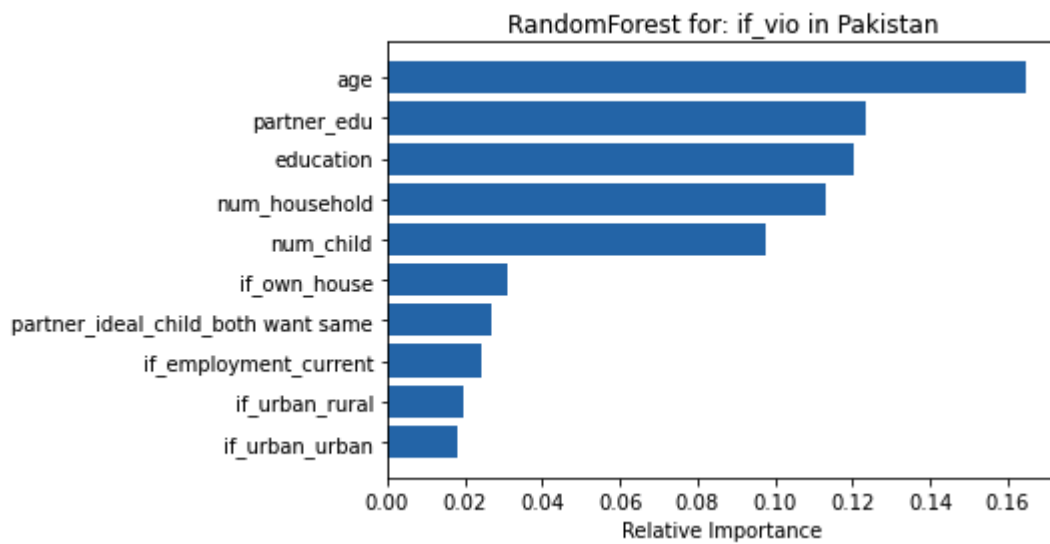
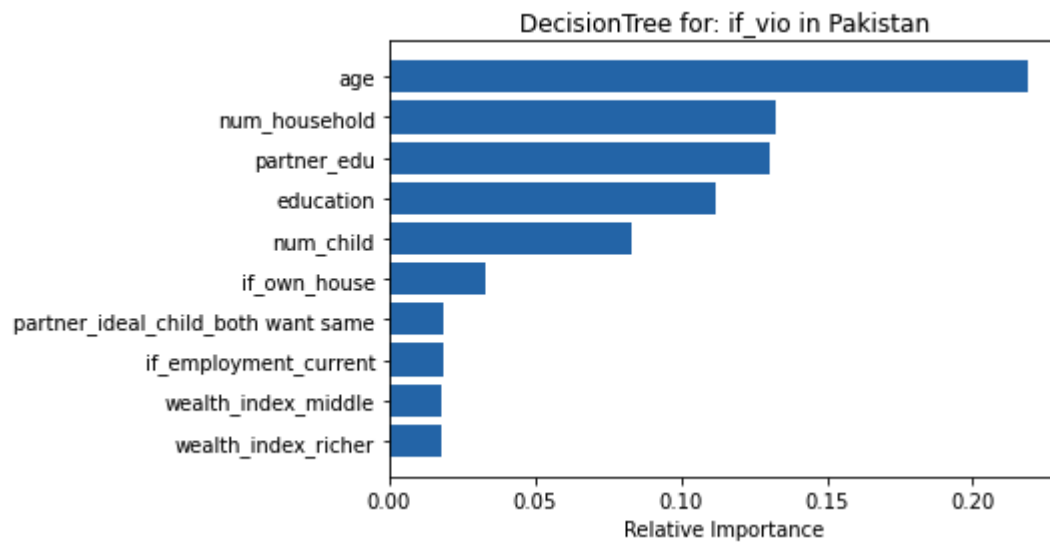
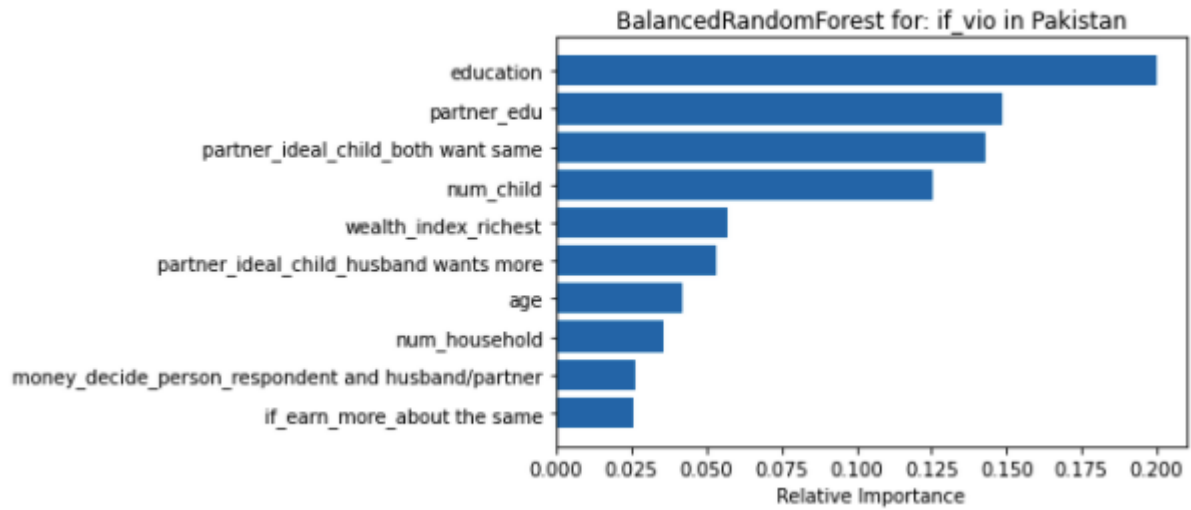


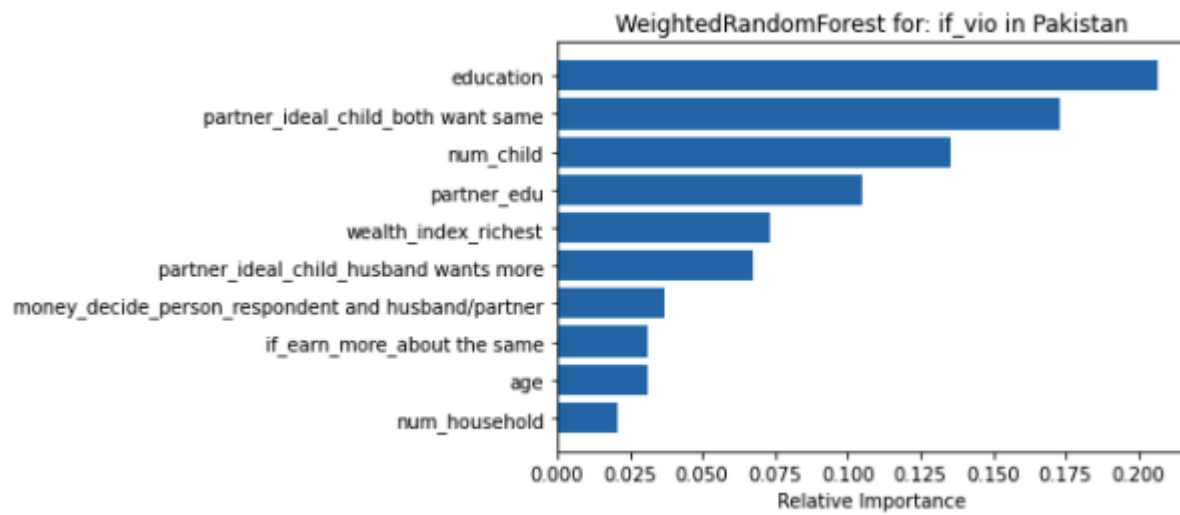
## Pakistan: Feature importance for IPV risk prediction using Classifier Models

\* Excludes models whose testing f1 scores are below 0.3.



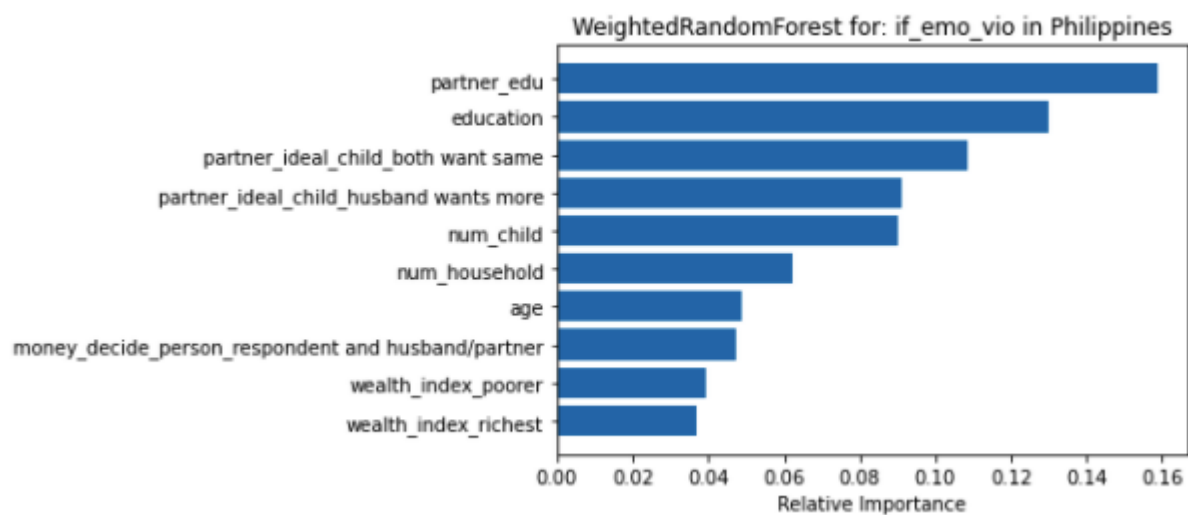
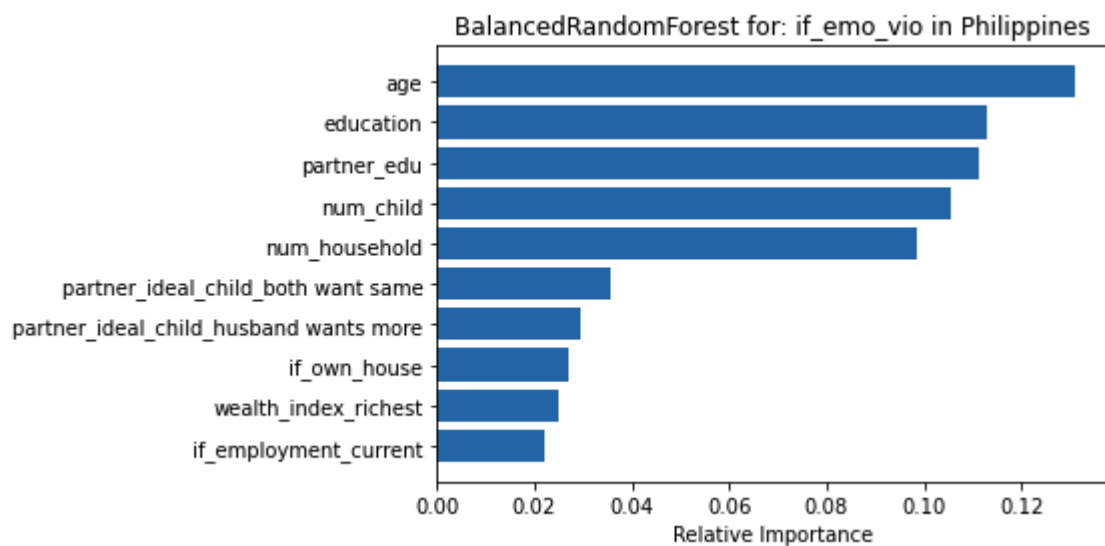


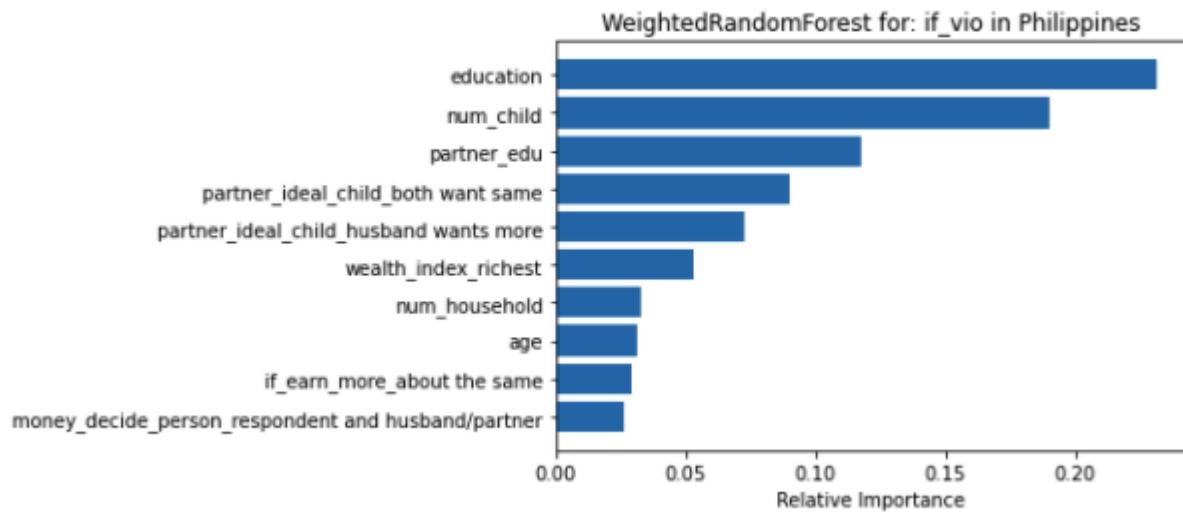
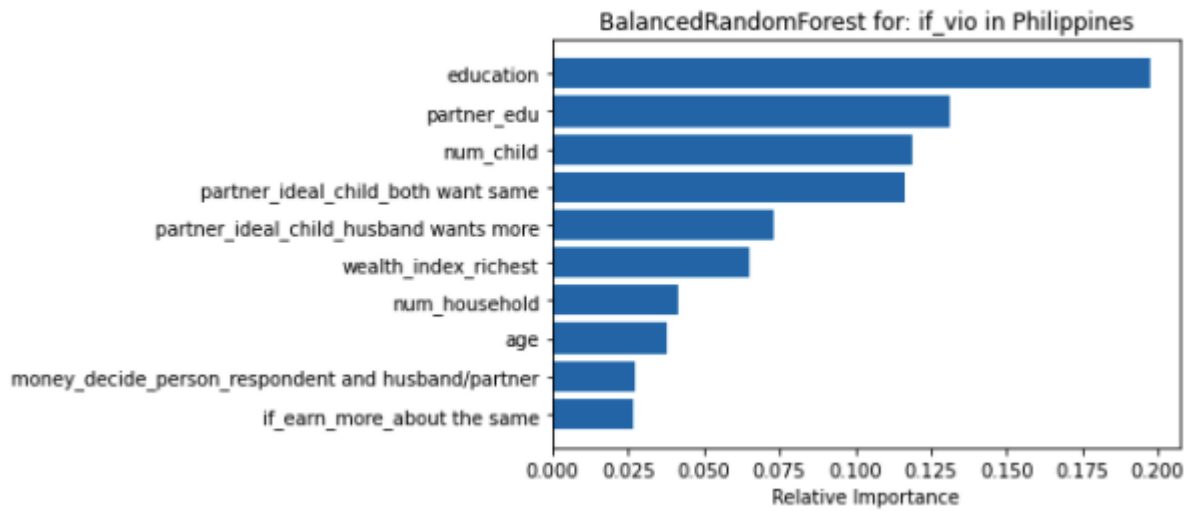




## Philippines: Feature importance for IPV risk prediction using Classifier Models

\* Excludes models whose testing f1 scores are below 0.3.





## 5 countries: Feature importance for IPV risk prediction within 5 countries using Classifier Models

\* Excludes models whose testing f1 scores are below 0.3.

