JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Using Machine Learning to Estimate Sexual Violence Against Women aged 15-24 in Different Countries: A Comparative Analysis

Sirjan Kaur
Henrico High School
302 Azalea Ave, Richmond, VA 23227

Abstract

This study employs TensorFlow Decision Forests, a machine learning framework, to analyze data from 2018 to 2021 and predict the prevalence of sexual violence against women aged 15-24 across 195 countries. Results reveal a high accuracy in predicting sexual abuse rates, with an F-score of 0.75, and highlight geographic variations, particularly in Central and West Africa. Additionally, an analysis of income inequality suggests a correlation with higher rates of sexual violence, while South Africa stands out as an outlier. Despite methodological limitations, such as underreporting biases, the research provides insights crucial for evidence-based policymaking and intervention strategies to combat this pervasive human rights issue. These findings hold promise for informing United Nations initiatives aimed at addressing sexual violence globally, facilitating resource allocation and targeted interventions to support survivors and prevent further instances of abuse.

Introduction

Sexual abuse has significant effects on the lives of women worldwide. The World Health Organization (WHO) reports that one in three women globally has experienced physical or sexual violence in their lifetime, indicating the scale of the problem (“Devastatingly pervasive: 1 in 3 women globally experience violence”). Sexual abuse is linked to adverse reproductive outcomes for women and girls and can have severe effects on their physical, emotional, and psychological well-being. Research conducted over the past two decades has shown that sexual abuse is associated with adverse reproductive outcomes for women and girls. In some studies, women who experience sexual abuse are twice as likely to report an unintended pregnancy than women who do not experience violence in their relationships (“The impact of sexual abuse on female development: Lessons from a multigenerational, longitudinal research study”). Additionally, sexual abuse is associated with an increased risk of sexually transmitted infections (STIs) and HIV (“Sexually Transmitted Diseases Among Adults Who Had Been Abused and Neglected as Children:A 30-Year Prospective Study”). Women who have experienced sexual abuse may also experience reproductive tract infections, pelvic inflammatory disease, and chronic pelvic pain. These conditions can have long-term consequences for women's reproductive health..

Sexual abuse can have significant physical and emotional consequences for survivors. Victims of sexual abuse are at higher risk of developing a range of mental health problems, including depression, anxiety, and post-traumatic stress disorder (PTSD) (“Sexual Abuse and Lifetime Diagnosis of Psychiatric Disorders: Systematic Review and Meta-analysis”). These conditions can have an impact on a survivor's daily life, including their ability to work, attend school, and maintain healthy relationships. In addition to mental health problems, survivors of sexual abuse may also experience physical health problems. These may include chronic pain, gastrointestinal problems, headaches, and fatigue. Survivors may also have difficulty sleeping and may experience nightmares and flashbacks related to their abuse. These physical symptoms can be debilitating and can further exacerbate the survivor's emotional distress (Tull).

Sexual assault carries a significant amount of stigma, particularly in lower-income areas and developing countries. Women living in countries such as South Asia and Sub-Saharan Africa may face numerous barriers to speaking out against injustices like sexual abuse, including cultural norms, societal expectations, and fear of retaliation (Kimani). For instance, in India, some women do not come forward due to fear of hostile reactions from the public, which can further victimize them. In addition to cultural factors, women in these areas may lack access to necessary healthcare services and legal recourse, further perpetuating cycles of abuse and making it more challenging to seek help and break free from abusive situations (“The stigma and blame attached to rape survivors in India”).

Recent advances in machine learning techniques provide new opportunities for identifying patterns of sexual abuse and predicting the likelihood of future instances. When it comes to reporting instances of sexual violence against women, data typically lags a year behind (e.g 2023 data is released in 2024 due to processing time), which can hinder policy and change-makers from recognizing trends in a timely manner. However, machine learning algorithms can be trained on large datasets to estimate the prevalence of sexual violence among women. By using machine learning to identify patterns of sexual abuse, policymakers and organizations can develop more targeted and effective prevention strategies. This is especially crucial in areas where sexual abuse is stigmatized, as comprehensive sex education programs, community-based interventions, and policies promoting gender equality and women's rights may face more significant barriers to implementation.

Data Dictionary:

Field	Description	Data Type	Example
Country	The name of the country.	Text	Afghanistan
WomenAffected_2018	The number of women affected by sexual assault in the year 2020, disaggregated by country.	Text	12
WomenAffected_2019	The number of women affected by sexual assault in the year 2019, disaggregated by country.	Number	20
WomenAffected_2020	The number of women affected by sexual assault in the year 2020, disaggregated by country.	Number	15
WomenAffected_2021	The number of women affected by sexual assault in the year 2021, disaggregated by country.	Number	35
WomenAffected_2022	The number of women affected by sexual assault in the year 2022, disaggregated by country.	Number	19
Pred:2023	The predicted value in the year 2023	Number	25.5619
gini_coef_feature	A measure of income inequality in a country calculated as the ratio of the area between the Lorenz curve and the line of perfect equality to the total area below the line of perfect equality.	Number	.45

Purpose:

The purpose of this study is to use machine learning techniques, specifically TensorFlow Decision Forests, to provide a comprehensive understanding of the prevalence of sexual violence against women aged 15-24 across all 195 countries. The aim is to develop a model that can accurately predict the incidence of sexual abuse and identify trends over time, by analyzing data from previous years.

The insights derived from this study can help contribute to the development of evidence-based policies and programs that enhance the well-being of survivors of sexual abuse and mitigate the incidence of this pervasive issue. Policymakers, healthcare providers, and other organizations can use the information to make informed decisions and implement effective prevention strategies to support survivors of sexual abuse and prevent further instances of sexual violence.

Moreover, we aim to identify regions where the issue of sexual violence against women is most prevalent. The insights from this can help organizations like the United Nations, governments, and NGOs set up sex education programs for younger girls and create dedicated resources to help women in areas where the issue is most common.

Methodology:

TensorFlow Decision Forests, a machine learning library developed by Google, was employed to predict the prevalence of sexual violence against women aged 15-24 in various countries for the current year. TensorFlow Decision Forests is a versatile machine learning library that enables the construction of decision tree-based models, allowing for the development of accurate predictions of sexual abuse incidence and identification of trends over time.

To build the models, data was obtained from reputable sources such as the United Nations International Children's Emergency Fund, the United Nations, and the World Health Organization. The data was collected on the percentage of women who reported experiencing sexual abuse in previous years, spanning from 2018 to 2021. Prior to building the models, the data underwent a cleaning process to remove any missing or inaccurate values, ensuring reliable and accurate data was used.

The cleaned data was used to develop decision trees using the F-score as a performance metric to predict the likelihood of sexual abuse in various countries. The F-score is a common evaluation measure used in machine learning that combines precision and recall assessing the overall accuracy of a model. The developed models were evaluated on a separate dataset to ensure their ability to generalize well to new data and to measure their accuracy using the F-score.

Google Sheets was utilized throughout the process. Google Sheets played a crucial role in data collection, cleaning, analysis, and modeling, providing a collaborative platform for the researchers to work with the data and develop machine learning models.

Results:

As mentioned in the methodology, to test the accuracy of the model, the researchers used the F-score, a performance metric that combines precision and recall. The F-score is a value between 0 and 1, where 1 represents perfect precision and recall, and 0 represents poor performance. In this study, the F-score was calculated as 0.75, indicating that the model has a relatively high accuracy in predicting the prevalence of sexual violence against women aged 15-24 across all 195 countries. A high F-score indicates that the model has a high level of precision and recall.

Moreover, a world map was generated to visualize the predicted rates of sexual violence against women aged 15-24 across all regions for the years 2021 and 2023, and to determine whether any changes have occurred. Figure 1 represents the map for 2021, and Figure 2 represents the map for 2023. The map was color-coded to represent the varying levels of sexual violence incidence, with green shades indicating higher rates.

Figure 1.1: World Map of Reported Sexual Abuse in 2021

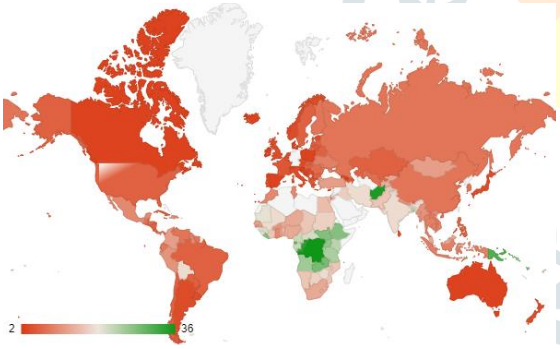
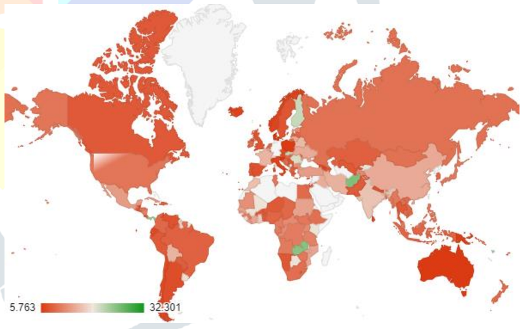


Figure 1.2: World Map of Reported Sexual Abuse in 2023

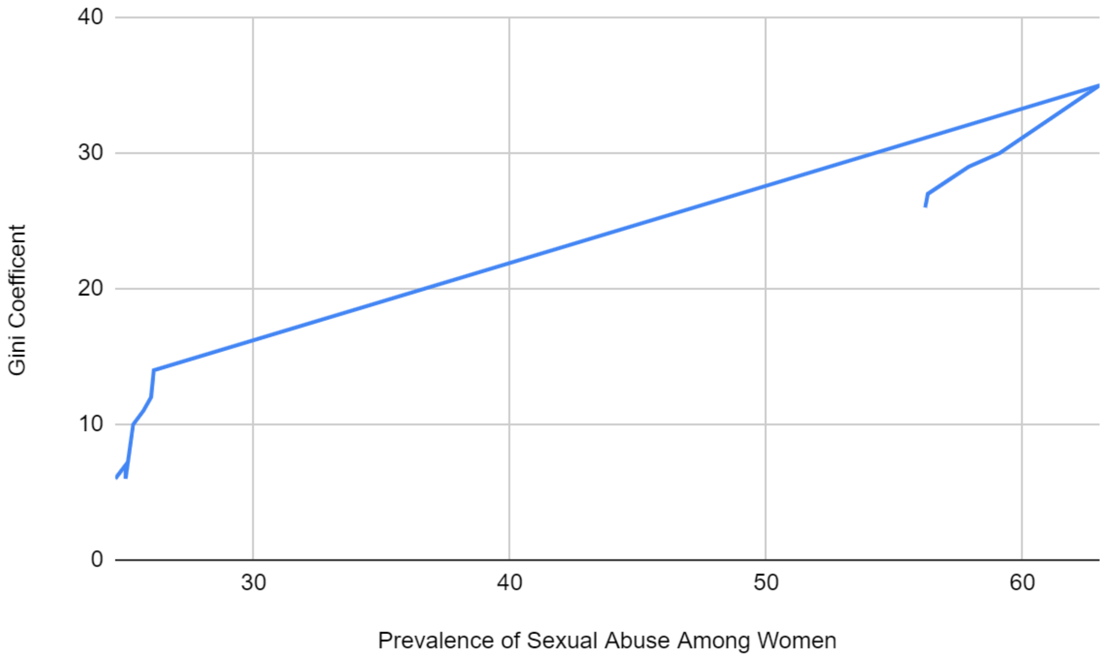


Upon analyzing the data presented in Figures 1 and 2, the research team noticed a significant difference in the prevalence of sexual abuse in central African countries, such as Congo, Chad, and the Central African Republic, between the years 2021 and 2023. This observation raises concerns about a potential shift in the incidence of sexual abuse in these regions. It is essential to note that machine learning models are not perfect and can have limitations. As such, the researchers believe that the observed difference could be a fault of the model used in this study.

However, the data presented in Figures 1 and 2 revealed a clear trend between lower-income countries and higher rates of sexual violence, with countries in the region of Central and West Africa experiencing particularly high levels of prevalence. However, the researchers also observed that some countries with middle or high-income classifications also demonstrated high rates of sexual violence, indicating that factors beyond income may be contributing to the incidence of sexual violence in certain regions.

To investigate this further, the team analyzed the relationship between income inequality and sexual violence rates using the Gini coefficient. The Gini coefficient measures income inequality on a scale from 0 to 1, with 0 representing perfect income equality and 1 representing very low levels of income equality. In Figure 1.3, the analysis shows a positive correlation between income inequality and the incidence of sexual abuse, suggesting that countries with lower income inequality tend to have higher rates of sexual abuse.

Figure 1.3



Despite having the highest levels of income inequality and Gini coefficient, South Africa stands out as an outlier in the data as it displays lower prevalence of sexual abuse among women compared to other countries in the dataset.

Conclusions:

Drawing on the findings of this research, several conclusions can be drawn regarding the relationship between income inequality, colonization, and sexual abuse against women. It is evident that income inequality is a significant factor in the prevalence of sexual abuse, as it can create systemic discrimination and marginalization of certain groups. Women in countries with high Gini coefficients may face significant barriers to economic empowerment, perpetuating harmful gender stereotypes and contributing to the normalization of sexual violence ("Sexual Harassment and the Gender Wage Gap").

Similarly, regions that have been colonized may experience unique challenges related to sexual abuse. Colonization can create unequal power dynamics and marginalize certain groups, perpetuating cultural attitudes and beliefs that contribute to the normalization of violence. Legacies of colonization can continue to impact societies long after colonialism has ended, including ongoing economic and social inequalities that exacerbate the issue (Mannell).

However, the researchers also acknowledge the limitations of their methodology, particularly the reliance on surveys and focus on past-year percentages. Underreporting of sexual abuse cases may be more prevalent in countries with high Gini coefficients due to associated stigma. Social stigma plays a significant role in the issue of sexual abuse. Stigmatization of survivors of sexual abuse can deter reporting and seeking justice, particularly in communities with high economic inequality. Cultural stigmas surrounding premarital sex and extra-marital relationships can also make individuals more vulnerable to sexual abuse, as the perceived impropriety of their actions can be leveraged against them by perpetrators.

Next Steps:

As outlined earlier, the researchers aim to utilize the results of this study to aid women worldwide. One of the potential applications of this research is to inform the United Nations (UN) initiatives to combat sexual violence globally. The UN has long acknowledged sexual violence as a significant human rights issue and has implemented several initiatives and programs to address it. In 2021, the UN pledged $25 million to support sexual assault response services in conflict-affected countries ("UN Action against Sexual Violence in Conflict – United Nations Office of the Special Representative of the Secretary-General on Sexual Violence in Conflict"). The machine learning models developed in this research could be utilized by the UN to estimate the prevalence of sexual violence against women in different countries and regions. By doing so, the UN could identify the areas with the highest rates of sexual violence and allocate resources accordingly to support local organizations and initiatives working to prevent and respond to sexual violence. Additionally, the findings of this research could be used to inform UN advocacy and policy efforts. The data and insights generated from this research could be utilized to highlight the urgent need for action to address sexual violence against women and to promote evidence-based policies and interventions.

Works Cited

"Devastatingly pervasive: 1 in 3 women globally experience violence." *World Health Organization (WHO)*, 9 March 2021, https://www.who.int/news/item/09-03-2021-devastatingly-pervasive-1-in-3-women-globally-experience-violence. Accessed 1 May 2023.

"The impact of sexual abuse on female development: Lessons from a multigenerational, longitudinal research study." *NCBI*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3693773/. Accessed 1 May 2023.

Kimani, Mary. "Taking on violence against women in Africa." *the United Nations*, https://www.un.org/africarenewal/magazine/july-2007/taking-violence-against-women-africa. Accessed 1 May 2023.

Mannell, Jenevieve. "How colonialism is a major cause of domestic abuse against women around the world." *The Conversation*, 25 April 2022, https://theconversation.com/how-colonialism-is-a-major-cause-of-domestic-abuse-against-women-around-the-world-179257. Accessed 1 May 2023.

"Sexual Abuse and Lifetime Diagnosis of Psychiatric Disorders: Systematic Review and Meta-analysis." *NCBI*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2894717/. Accessed 1 May 2023.

"Sexual Harassment and the Gender Wage Gap." *National Partnership for Women & Families*, https://www.nationalpartnership.org/our-work/resources/economic-justice/fair-pay/sexual-harassment-and-the-gender-wage-gap.pdf. Accessed 1 May 2023.

"Sexually Transmitted Diseases Among Adults Who Had Been Abused and Neglected as Children:A 30-Year Prospective Study." *NCBI*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2724945/. Accessed 1 May 2023.

"The stigma and blame attached to rape survivors in India." *Human Rights Watch*, 8 January 2013, https://www.hrw.org/news/2013/01/08/stigma-and-blame-attached-rape-survivors-india. Accessed 1 May 2023.

Tull, Matthew. "Sexual Assault PTSD: Symptoms, Other Effects, and Treatments." *Verywell Mind*, 7 February 2022, https://www.verywellmind.com/symptoms-of-ptsd-after-a-rape-2797203. Accessed 1 May 2023.

"UN Action against Sexual Violence in Conflict – United Nations Office of the Special Representative of the Secretary-General on Sexual Violence in Conflict." *the United Nations*, 7 March 2023, https://www.un.org/sexualviolenceinconflict/about-us/un-action/. Accessed 1 May 2023.