PRELIMINARY PREPRINT VERSION: DO NOT CITE
The AAAI Digital Library will contain the published
version some time after the conference.

Participatory Machine Learning Models in Feminicide News Alert Detection

Amelia Lee Dogan
Data + Feminism Lab
Massachusetts Institute of Technology
dogan@mit.edu

Abstract

After criminal recidivism or hiring machine learning models have inflicted harm, participatory machine learning methods are often used as a corrective positioning. However, little guidance exists on how to develop participatory machine learning models throughout stages of the machine learning development life-cycle. Here we demonstrate how to co-design and partner with community groups, in the specific case of feminicide data activism. We co-designed and piloted a machine learning model for the detection of media articles about feminicide. This provides a feminist perspective on practicing participatory methods in a co-creation mindset for the real-world scenario of monitoring violence against women.

Introduction

Gender-related violence against women and its lethal outcome, feminicide, is a serious problem across the Americas (Fregoso and Bejarano 2010). Although governments have passed legislation criminalizing feminicide, these laws have not been accompanied by relevant policy nor by robust data collection that measures the scope and scale of the problem (Simonovic 2017). To fill this data gap, many activists collect *counterdata* to track and record feminicides in their own communities and context. Collecting counterdata begins with communities understanding the failure of the state and requires them "to realize and subsequently express their situated knowledge and capacity to collect data themselves" (Meng and DiSalvo 2018). While collecting this counterdata, feminicide data activists face challenges like lack of time and financial resources, difficulties in accessing official data, and the mental health burden of reading about violent deaths of women (D'Ignazio et al. 2020).

The Data Against Feminicide project in the Data + Feminism Lab is an international participatory action research project designed to help sustain activist efforts to collect feminicide data by partially automating detection using machine learning. Through the partnership, the team has developed two tools: a browser extension to be used on news articles to highlight relevant key terms (names, places, age/-dates, and custom words) and an email alert system using a machine learning model to suggest articles for activists to read. The initial pilot of these tools in Spring 2021 involved four groups in the United States tracking all feminicides (ie: all feminicides in a region) and racialized feminicides (ie: feminicides of women of a specific race) and also three Spanish language groups in Latin America tracking all feminicides.

Methodology

The initial research to understand the context for machine learning tools involved interviews with ten feminicide data activist groups across the Americas. These interviews ranged between 77 and 115 minutes. Through qualitative analysis of the interviews, we found many organizations rely on online news media sources as a primary method of discovering and recording cases (D'Ignazio et al. 2020).

After collaborative brainstorming with two groups that served as our primary co-design partners, the development of a Chrome browser extension and email alert system started. Using a sample search query provided by a co-design partner, the lab and a co-design partner identified positive and negative news articles for model training. Other interview participants also provided additional positive and negative sample media articles. Previously, the team experimented with a multinomial naive Bayes model (D'Ignazio et al. 2020), but for the pilot, to detect articles we used a logistic regression model to predict the probability of feminicide from the article text. The model achieved 81% accuracy and 93% recall on a held out test set of 74 articles.

For the Spring 2021 pilot, four groups in the United States and three groups in Latin America participated in using the tools for eight weeks. The teams were set up with tools during a group introductory orientation, followed up halfway through the pilot with a focus group, asked to fill out weekly surveys, and debriefed at the end with a meeting with the research team. All of these components used participatory computing methodologies that go beyond community consultation but centered community expertise and an understanding that designing a new system is a continuous process (Caselli et al. 2021). After the end of the Spring 2021 pilot, iterative development started with the English language models to create specialized models to detect racialized feminicides for Black women killed in officer-involved killings and Indigenous women.

Findings

The groups in Latin America tracked all feminicides occurring in their geographic regions, while three out of four groups in the United States tracked racialized feminicides (Black women, Black women killed in officer-involved situations, and missing and murdered Indigenous women and girls) and one group tracked all feminicides in the United States. Through focus group data and weekly survey data to evaluate the model usefulness to partners, it became apparent that despite language differences, groups tracking all feminicides in a geographic area demonstrated success with the email alert system. Groups tracking all feminicides found multiple cases a week using the email alert system, found cases they would not have found otherwise, and the email alerts saved them time and made collection easier. However, groups tracking racialized feminicides at times found no articles using the email alert system or reported spending more time on the email alert system because of an overwhelming number of irrelevant alerts. For example, in the weekly survey, in response to the statement "The tools have made finding and recording cases this week easier and/or less taxing" (1=Strong Disagree and 5=Strongly Agree), the group tracking missing and murdered Indigenous women and girls had an average response of 2.55, the group tracking officer involved deaths of Black women had an average response of 2.75, the group tracking feminicides of Black women had an average response of 2.83, and all other groups had an average response of 3.86. Across both Latin American and United States-based groups the data highlighter browser extension proved helpful in scanning articles and assisting in extracting information.

Possible explanations for the failure of the model for racialized feminicide data activists can occur at several places in the machine learning life cycle (Suresh and Guttag 2021). One explanation could be that the initial English language model article examples did not include a representative number of articles about Black women's and Indigenous women's deaths. The lack of articles could also be a limitation of the general query used to search for articles. This explanation could be exacerbated by the different ways journalists write about Black and Indigenous victims of feminicide (Neely 2015; Gilchrist 2010). Another factor highlighted by activist groups themselves could include the lack of media coverage in general about racialized victims of feminicide meaning no article would be found despite us expanding the number of media sources drawn upon. This limitation at times leads groups to rely heavily on social media for feminicide cataloging (D'Ignazio et al. 2020).

The failure of the email alert tool for racialized groups required a recommitment of model development for groups tracking racialized feminicides. The team decided retraining a new set of models was in line with the data feminism principle of *rethinking binaries and hierarchies* (D'Ignazio and Klein 2020). Currently, another model reiteration process is underway with two American racialized feminicide data activist groups. During this round, a variety of new model techniques such as combining multiple models to target specific intersectional cases of interest are being trained.

Conclusion

Failure of machine learning models and AI systems for intersectional populations has been broadly researched (Buolamwini and Gebru 2018; Noble 2018). However, through our example of feminicide machine learning model development, I show that participatory methods cannot fully mitigate these failures. Yet through a commitment to research partners and collaboration in non-extractive relationships employing data feminist, and participatory methods co-created technology can be developed. The need for co-created tools in the counterdata activist space is growing, as already there are over 150 groups globally involved in the feminicide counterdata movement alone.

References

Buolamwini, J.; and Gebru, T. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability, and Transparency in Proceedings of Machine Learning Research*, 1–15. New York: Association for Computing Machinery.

Caselli, T.; Cibin, R.; Conforti, C.; Encinas, E.; and Maurizo, T. 2015. Guiding Principles for Participatory Design-Inspired Natural Language Processing. In *Proceedings of the 1st Workshop on NLP for Positive Impact*, 27–35. Online: Association for Computational Linguistics.

D'Ignazio, C.; and Klein, L. 2021. *Data Feminism*. Cambridge, MA: The MIT Press.

D'Ignazio, C.; Suarez Val, H.; Fumega, S.; Suresh, H.; Cruxen, I.; So, W.; Martinez, A.; and Garcia-Montes, M. 2020. Feminicide Machine Learning: Detecting Gender-based Violence to Strengthen Civil Sector Activism. In *Mechanism Design for Social Good Workshop*. Online: Association for Computing Machinery.

Fregoso, R.; and Bejarano, C., eds. 2010. *Terrorizing Women: Feminicide in the Americas*. Durham, NC: Duke University Press.

Gilchrist, K. 2010. "Newsworthy" Victims?: Exploring Differences in Canadian Local Press Coverage of Missing/Murdered Aboriginal and White Women. *Feminist Media Studies*, 10(4): 373–90.

Meng, A.; and DiSalvo, C. 2018. Grassroots resource mobilization through counter-data action. *Big Data Society*, 5(2): 1–12.

Neely, C. 2015. *You're Dead—So What? Media, Police, and the Invisibility of Black Women as Victims of Homicide*. East Lansing, MI: Michigan State University Press.

Simonovic, D. 2017. Why We Need a Femicide Watch. In Femicide Volume VII: Establishing a Femicide Watch in Every Country. Technical report, Academic Council on the United Nations System (ACUNS) Vienna Liaison Office.

Suresh, H.; and Guttag, J. 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. arXiv:1901.10002.