Uncovering Invisible Demographic Patterns of Feminicide in Ceara, Brazil, Through Machine Learning and Cartographic Analysis

Queiroz, Yanna (School: Escola de Ensino Medio Joaquim de Figueiredo Correia)

Feminicide remains one of the most severe expressions of gender-based violence worldwide, reflecting structural inequalities of race, territory, age, and social context. In Brazil, underreporting and data fragmentation limit a comprehensive understanding of this phenomenon. To address this issue, this research examined feminicides in the state of Ceara between January 2022 and June 2025, integrating artificial intelligence, statistics, and cartography to analyze the cases. A Natural Language Processing system was developed to collect, extract, and classify demographic information from news articles. The model was trained using 10,000 reports to identify variables such as date, age, race/ethnicity, location, and crime circumstances. Among the tested algorithms, Random Forest achieved the best performance, with 98.8% accuracy, 98.7% sensitivity, and 99.4% AUC-ROC, outperforming Support Vector Machine and rule-based approaches.The extracted data were integrated with official Brazilian databases and analyzed using the Feminicide Index, socioeconomic correlations, and thematic mapping. The results identified 174 victims, with higher prevalence among Black women (49%), individuals aged 31 to 50 (36%), and cases involving intimate partners (48.7%). Spatial analysis showed a concentration in Greater Fortaleza, followed by Cariri, Sertao de Sobral, and Vale do Jaguaribe. Regression analysis indicated strong correlations with population factors, and a peak Feminicide Index of 1.25 per 100,000 women was observed , exceeding official records. This study demonstrates that the proposed system improves detection of underreported cases and reveals hidden demographic and territorial patterns, supporting more effective public policy strategies to prevent feminicide.

Awards Won:
Fourth Award of $600