

Analysing and Classifying Legal Texts on Mexican Cases of Domestic Violence: A Natural Language Processing and CNN-Based Approach

Abstract

Domestic violence remains a persistent issue in Mexico, leading many victims to seek protection through the courts. The legal texts generated from these proceedings constitute a valuable yet understudied source of information. Beyond providing detailed descriptions of individual cases, they offer insight into the contextual factors and recurring patterns through which violence is manifested. Analysing such texts contributes to a deeper understanding of the structural dynamics underpinning gender-based violence, helping to provide better institutional responses. This paper examines the potential of Natural Language Processing (NLP) and Convolutional Neural Networks (CNN) for classifying legal texts of cases of domestic violence. A sample of 443 documents on protection trials concerning domestic violence in Mexico (2004-2024) is analysed to identify prevailing topics and assess their content. A two-stage analysis is followed. In the initial phase, topic modelling is conducted using Non-negative Matrix Factorisation combined with word-window analysis based on BERT embeddings. In the second stage, a Convolutional Neural Network classifier is implemented leveraging the semantic features of the cases and their severity. The findings reveal twelve distinct topics related to domestic violence in Mexico. Among the identified topics, gender equality and divorce stand out as the most severe. The analysis also shows that all the identified topics reveal the prevalence of different types of abuse, such as psychological, physical, and economic. The interplay of these three forms of abuse creates a reinforcing cycle in which each element supports and intensifies the others. The classification model is evaluated using accuracy, precision, recall, and F1 scores and compared to four competing baseline frameworks, namely, Random Forest, Support Vector Machine, Naïve Bayes, and k -Nearest Neighbours. The proposed classifier achieves a performance of 84% accuracy demonstrating its effectiveness. By identifying different topics and levels of severity in domestic violence cases, the model offers valuable input for improving institutional responses. This enables the court to assign specialized judges more effectively, while also supporting the development of more comprehensive victim support by identifying risk factors and prioritizing care for those in more vulnerable situations.

Key words: domestic violence; legal texts; text classification; topic modelling; convolutional neural networks.