

International Trends in Gender Studies Through Scientific Text Mining with Natural Language Processing Techniques

Nora Gavira-Durón¹, Ana Lorena Jiménez-Preciado², Angélica Alonso-Rivera³

^{1,3} *Universidad de las Américas Puebla, México*

² *Instituto Politécnico Nacional, México*

Abstract

This study explores the most discussed topics in scientific discourse on gender through advanced Natural Language Processing (NLP) techniques. The research analyses 168 peer-reviewed articles published by Emerald Publishing Limited, primarily from 2023 and 2024, to identify the most prevalent unigrams, bigrams, and trigrams related to gender in various contexts, such as country, work area, research, religion, and others. The application of NLP improves the quality and depth of gender-related topic analysis in scientific literature, allowing for automated and efficient processing of large volumes of text. Furthermore, it facilitates the identification of trends and patterns within the texts. The detection of prevalent unigrams, bigrams, and trigrams enables the identification of recurrent keywords and phrases, providing a clear view of predominant gender-related themes across diverse contexts. The analysis exposes an emerging interest in improving women's circumstances across all sectors, the recognition of gender disparities, and a substantial commitment from the academic community to redress these imbalances. The topic modeling results in spotlight clusters of research focus, such as workplace equity, educational access, and gender-based violence, offering a wide snapshot of the current state of gender studies. Through Named Entity Recognition (NER), the study pinpoints crucial institutions, policymakers, and researchers spearheading the gender equality agenda and geographical hotspots of gender-focused research. This information is a goldmine for understanding the global landscape of gender studies and identifying potential areas for international collaboration, fostering a sense of hope and optimism. The sentiment analysis evaluates the tone and emotions in the articles, revealing the challenges and frustrations expressed regarding persistent inequalities but also the optimism and determination evident in discussions of progress and proposed solutions. Moreover, approaches and methodologies used to study gender issues are diversifying, reflecting women's challenges and opportunities in the international workplace and educational orbs.

Keywords: Gender equality, Access to education, Gender violence, Natural Language Processing